



**Rol de la compartimentación
nuclear en la regulación de
la expresión génica en
*Trypanosoma cruzi***

Mag. Lucía Pastro

Tesis presentada para acceder al Doctorado en Ciencias Biológicas

Lucia Pastro Cardoso

Orientador: Dra. Beatriz Garat

Co-orientador: Dr. José Sotelo-Silveira

Tribunal:

Dr. Jose A. Chabalgoity

Dra. Elena Fabiano

Dr. Javier De Gaudenzi

Agosto 2017

1. Introducción.....	9
1.1. La enfermedad de Chagas	10
Generalidades.....	10
Distribución.....	10
Síntomas y tratamiento	12
1.2. El parásito.....	13
Ciclo de vida	13
Biología celular de T. cruzi.....	15
1.3. El genoma de T. cruzi	18
Familias multigénicas de T. cruzi.....	20
1.4. Expresión génica en tripanosomátidos.....	22
Regulación postranscripcional de la expresión génica en los tripanosomátidos	24
Actores moleculares en la regulación de la expresión génica en los tripanosomátidos.....	26
1.5. El rol del núcleo en la regulación de la expresión génica	30
2. Hipótesis de trabajo y Objetivos.....	35
2.1. Hipótesis.....	36
2.2. Objetivo general.....	36
2.3. Objetivos específicos	36
3. Estrategia experimental.....	37
4. Transcritos diferencialmente localizados en núcleo y citoplasma en epimastigotas de T. cruzi.....	40
Material Suplementario.....	58
5. Búsqueda de señales involucradas en la compartimentación núcleo-citoplasma de transcritos en epimastigotas de T. cruzi	67
5.1. Proteínas de T. cruzi que reconocerían los motivos identificados en los transcritos enriquecidos en el citoplasma.....	71
5.2. Proteínas de T. cruzi que reconocerían los motivos identificados en los transcritos enriquecidos en el núcleo.	78
Materiales y Métodos.....	88
6. Análisis comparativo de transcritos poliadenilados y totales de epimastigotas de T. cruzi	90
7. Comprendiendo el cis-splicing en T. cruzi: análisis del gen codificante de la poliA polimerasa	123
7.1. Análisis comparativo in silico del gen que codifica para la proteína PAP en cepas de T. cruzi.....	124
7.2. Análisis de la expresión del gen que codifica para la proteína PAP en los transcriptomas citoplásmico, nuclear y total de epimastigotas de T. cruzi.....	129
Materiales y Métodos.....	138
Análisis de secuencias	138

Ensamblado de novo, CLC Genomics Workbench	138
PCR cuantitativo	138
Anexo	138
8. Conclusiones y Perspectivas	152
9. Bibliografía	159

Agradecimientos

A mi familia, a mis padres, por apoyarme y alentarme a hacer lo que me gusta, a mis hermanos y sobrinos porque siempre están, a mis abuelos que siempre nos acompañan. A mis amigos del LIM que me hacen sentir en casa. A Beatriz por enseñarme y comprenderme, a Coya por sus aportes, a María Ana por todo su apoyo, a Leti por hacer los días más divertidos, a Pablo por estar siempre dispuesto a darme una mano, a Lorena por su compañía y ayuda, a Chavez por la buena onda, a Rafael por su consideración, a Martín por "desasnarne", a Caro, Ceci y Santi por las discusiones enriquecedoras y a Meche, Juan, Fabricio y Lucía por escucharme.

A mis compañeros de Genética por la comprensión y las discusiones sobre este tema.

A los compañeros del 3er piso

A Pao, Vivi, Naty, Ile, Ale, Jhony, Noe, Titi, Gaby, Colacha, Marcel y alguno q a lo mejor me olvido, por escucharme y alegrarme los días.

A los miembros del tribunal por los aportes y la comprensión.

A la gente del Laboratorio de desarrollo tecnológico por los *real time*.

A la gente del Departamento de Genómica del IIBCE.

A Alfredo por lo mejor del mundo, Josefina!

Resumen

Trypanosoma cruzi es el protozoario parásito causante de la enfermedad de Chagas, transmitida por diversas especies de insectos de la familia *Reduviidae*. Esta enfermedad constituye un problema importante para la población rural latinoamericana.

Los tripanosomátidos presentan varias peculiaridades en la regulación de la expresión génica que los convierten en un modelo sumamente interesante para el estudio molecular básico de este proceso. El inicio de la transcripción de genes que codifican proteínas no está controlado. Los genes están organizados en largas unidades policistrónicas separadas por regiones intergénicas que contienen las señales para el procesamiento de los ARNs y la regulación de la expresión opera principalmente a nivel postranscripcional. Las etapas postranscripcionales de regulación de la expresión génica, incluyendo estabilidad, traducibilidad y localización del mensaje son consideradas esencialmente procesos citoplásmicos. Sin embargo, en eucariotas, existen otras etapas a nivel nuclear que suceden co- o postranscripcionalmente que constituyen también etapas de control de la expresión génica.

Nuestra hipótesis de trabajo es que la compartimentación nuclear de los ARNs interviene en la regulación de la expresión génica de *Trypanosoma cruzi*.

La mayor parte de los estudios realizados para la caracterización de eventos postranscripcionales en *Trypanosoma cruzi* se han dirigido al estudio de componentes moleculares en homogenados totales de parásitos, centrándose principalmente, en la identificación de señales y factores regulatorios operativos en el citoplasma. No existen a la fecha estudios comparativos entre núcleo y citoplasma.

Para demostrar nuestra hipótesis, realizamos un estudio del transcriptoma nuclear y citoplásmico de epimastigotas de *Trypanosoma cruzi* y analizamos comparativamente los transcritos sobre-representados en cada compartimiento.

Los genes cuyos transcritos están enriquecidos en el citoplasma están involucrados en la síntesis de proteínas y en procesos metabólicos, ambos mecanismos funcionales en el estadio epimastigota. Por otro lado, los genes cuyos transcritos están enriquecidos en el núcleo pertenecen a familias multigénicas, mayormente involucradas en los procesos de infección del parásito y de expresión en otros estadios del ciclo de vida. A su vez, los

transcriptos diferencialmente localizados tienen características estructurales y composicionales diferentes. Detectamos la presencia en estos genes de motivos lineales conservados distintivos, que podrían ser responsables de su compartimentación subcelular. A su vez, el parásito expresa varias proteínas de unión al ARN ortólogas a proteínas que presentan antecedentes en la literatura de unión a estos motivos en otros organismos. La comparación de los transcriptomas total, nuclear y citoplásmico con los transcriptos poliadenilados de epimastigotas de *Trypanosoma cruzi*, permitió definir grupos de genes sujetos a una fuerte regulación a nivel de estado estacionario de mensajeros, maduración nuclear de ARN o retención nuclear. Por último, pudimos confirmar que el *cis-splicing* sucede en el núcleo de epimastigotas de la cepa Dm28c de *Trypanosoma cruzi*, aunque, obtuvimos resultados que apoyan que el mensajero sin procesar se localiza también en el citoplasma.

Nuestros datos muestran que la expresión génica en epimastigotas de *Trypanosoma cruzi* es influenciada por la compartimentación de los ARNs mensajeros entre el núcleo y el citoplasma, siendo un aporte novedoso y de utilidad para la comunidad científica del área.

1. Introducción

1.1. La enfermedad de Chagas

Generalidades

Esta enfermedad potencialmente mortal, lleva el nombre del médico e investigador brasileño, Carlos Ribeiro Justiniano Chagas, que en 1909 describió la enfermedad, el parásito que produce la dolencia y su ciclo de vida, así como el vector que la transmite (Chagas, 1909).

El agente etiológico causal de la enfermedad de Chagas o Tripanosomiasis Americana es el protozooario parásito *Trypanosoma cruzi* (*T. cruzi*), perteneciente a la clase Kinetoplastida, orden Trypanosomatida, familia Trypanosomatidae.

La OMS estima que hay entre 6 y 7 millones de personas infectadas por *T. cruzi* en el mundo sobre todo en zonas endémicas de 21 países de América Latina, donde la transmisión vectorial se produce, aunque actualmente la enfermedad se ha propagado a otros continentes.

La enfermedad de Chagas es catalogada como una de las "*neglected diseases*", enfermedades que reciben esta denominación por ser altamente prevalentes en las regiones más pobres del mundo y que, sin embargo, reciben poco apoyo financiero para investigación y programas de salud específicos con el fin de combatirlas (Hotez *et al*, 2007; O'Connell, 2007).

Distribución

La enfermedad de Chagas se encuentra extendida mayoritariamente en la parte continental de América Latina (figura 1). Más de 150 especies de insectos de la subfamilia Triatominae son los vectores portadores y transmisores del parásito, aunque solo 15 de estos la transmiten a los humanos. En la figura 1 se muestra el estado de transmisión vectorial del parásito.

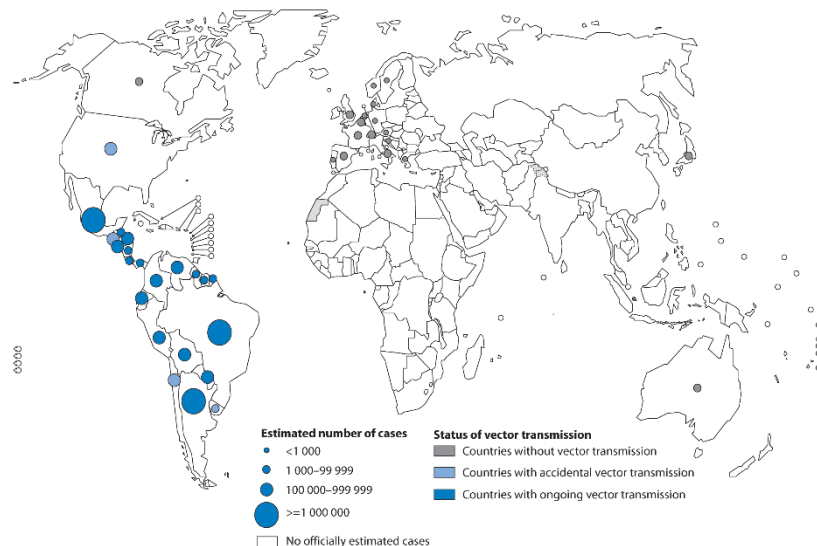


Figura 1. Mapa de distribución mundial de casos de infección por *T. cruzi*. Se indica también el estado de la transmisión vectorial. Extraído de (WHO, 2009).

La prevención de la enfermedad de Chagas en zonas endémicas mediante control vectorial es la forma más útil para su erradicación. Desde hace varias décadas el control del insecto vector, llevado a cabo de forma sistemática en los países del cono sur, ha conducido a un descenso en las cifras de infección y mortalidad (de 50000 muertes al año en el 1991 a 13000 en el año 2001). Este programa de control que fue ampliamente promovido desde el año 1991 en toda la región, consiguió erradicar en Uruguay la transmisión vectorial en el año 1997 (Remme *et al.*, 2006). Sin embargo, la infección se puede adquirir también mediante transfusión y se han detectado casos de infección por órganos donados, aunque estos modos de transmisión son menos frecuentes. En los últimos años, se han detectado casos de enfermedad de Chagas en países no endémicos, principalmente Estados Unidos y Europa (Cantey *et al.*, 2012; Bern *et al.*, 2011). El esparcimiento de la enfermedad se ha dado a causa de las migraciones de individuos infectados a estas regiones (Figura 1). Existe también transmisión vertical, un 5% los recién nacidos de madres portadoras, resultan contagiados. Por estos motivos, es necesario realizar el diagnóstico de la infección en sangre de donantes en todos los casos y en embarazadas en particular en áreas donde la enfermedad es endémica (WHO,

2016). Por otro lado, la infección puede transmitirse por vía oral debido a la ingesta de alimentos contaminados con formas infectivas del parásito (Yoshida, 2008).

Síntomas y tratamiento

La enfermedad de Chagas tiene dos fases bien diferenciadas: una primera, aguda, de unos dos meses de duración, que presenta alta parasitemia, durante la cual, en la mayoría de los casos, los pacientes no presentan síntomas o éstos son leves. Un signo inicial característico, presente en menos de la mitad de las personas picadas por uno de los insectos transmisores, puede ser una lesión cutánea o una hinchazón característica de un párpado (signo de Romana). Además, puede presentarse fiebre, dolor de cabeza, agrandamiento de ganglios linfáticos, palidez, dolores musculares, dificultad para respirar, entre otros síntomas. En la fase crónica de la enfermedad, los parásitos se ocultan del sistema inmune dentro de las células del músculo cardíaco y digestivo. Hasta el 30% de los pacientes padecen trastornos cardíacos y hasta el 10% presentan alteraciones digestivas (agrandamiento del esófago o del colon), neurológicas o una combinación de éstas. Con el tiempo, la infección puede causar muerte súbita o insuficiencia cardíaca por la destrucción progresiva del músculo cardíaco (WHO, 2016).

Los fármacos utilizados para el tratamiento de la enfermedad de Chagas son los mismos desde hace más de 40 años: el nifurtimox, desarrollado en el año 1960 por Bayer y el benznidazol, desarrollado por Roche en el año 1974. A pesar del largo tiempo transcurrido desde su descubrimiento, aún no existe un consenso sobre la acción de estas drogas en el parásito. El nifurtimox, podría inhibir la enzima tripanotona reductasa, que cataliza la eliminación de especies reactivas de oxígeno en la célula (Boiani *et al.*, 2010; Hall *et al.*, 2011). Mientras que la acción del benznidazol podría involucrar la formación de enlaces covalentes u otras interacciones de intermediarios de la nitrorreducción con los componentes del parásito (Polak y Richle, 1978) o su unión al ADN, lípidos o proteínas (Díaz de Toranzo *et al.*, 1988). El benznidazol presenta mayor eficacia y es más seguro, por lo cual es utilizado en los tratamientos de primera línea. Ambos fármacos presentan efectos secundarios adversos que incluyen dermatitis alérgica, neuropatías, leucopenia, granulocitopenia, náuseas, vómitos, anorexia y pueden

aparecer también alteraciones del sistema nervioso central (Rassi *et al.*, 2012). El tratamiento es efectivo solo si es realizado durante la fase aguda de la enfermedad, eliminando los parásitos reduce la duración y magnitud de los síntomas y disminuye la mortalidad. Sin embargo, en la fase crónica, el tratamiento continúa siendo motivo de debate, debido a su dudosa eficacia y a la severidad de los efectos secundarios que producen.

El proyecto BENEFIT (*Benznidazol Evaluation for Interrupting Trypanosomiasis*), fue desarrollado para evaluar la eficacia y la seguridad del tratamiento con Benznidazol en pacientes con cardiomiopatía Chagásica, los resultados no fueron los esperados, ya que mostraron que, a pesar que existe una reducción estadísticamente significativa de la carga parasitaria de los individuos, el deterioro cardíaco no disminuyó de manera significativa en los pacientes tratados respecto a los controles (Morillo *et al.*, 2015).

1.2. El parásito

Ciclo de vida

El parásito tiene un ciclo de vida relativamente complejo, que transcurre en dos hospederos, un insecto hematófago, triatómino o vinchuca y uno vertebrado, dentro de los cuales los animales silvestres son los principales reservorios del parásito (WHO, 2016). Durante el ciclo, el parásito alterna entre formas infectivas no replicativas, y formas no infectivas replicativas, presentando cuatro estadios principales relacionados con el ambiente en que se encuentra. Dos formas se encuentran en el insecto, el epimastigota en la luz del intestino y el tripomastigota metacíclico, que es la forma infectiva, en la ampolla rectal. En el hospedero mamífero, se encuentra la forma amastigota dentro de células y la forma tripomastigota, infectiva, en el torrente sanguíneo (figura 2).

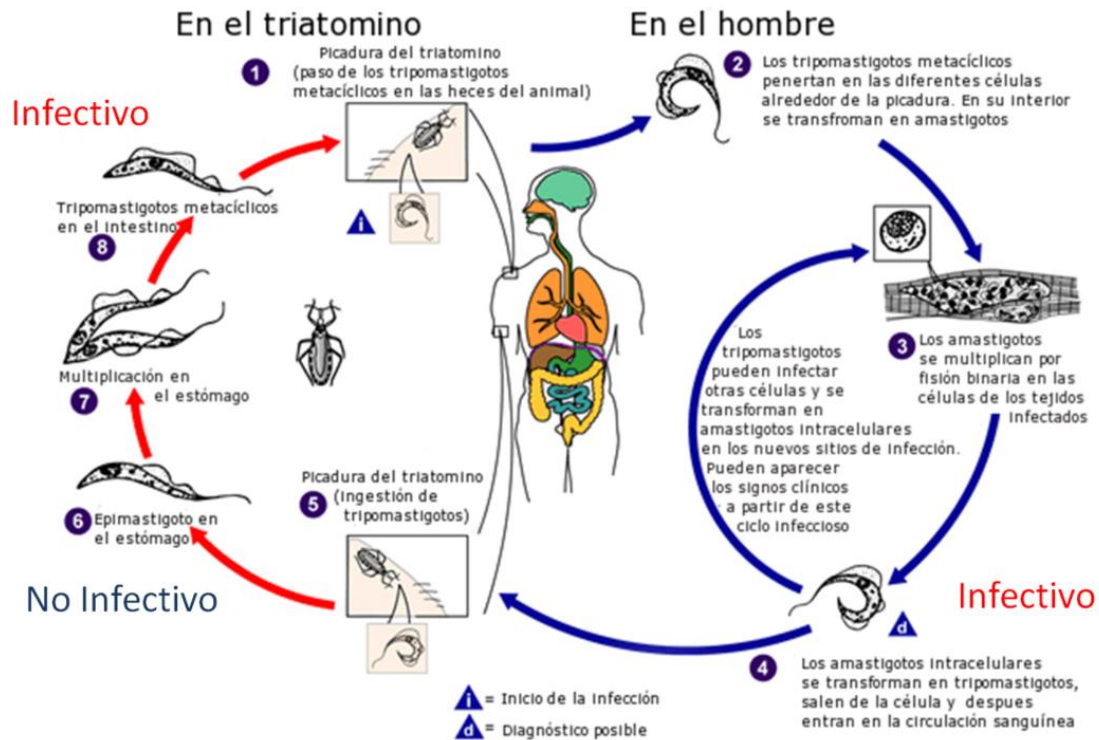


Figura 2. Ciclo de vida del parásito *T. cruzi*. El triatómino infectado al alimentarse de sangre de un mamífero, libera en las heces tripomastigotos infectivos que pueden entrar al hospedero, invadiendo las células cerca del sitio de inoculación. Dentro de las células, los tripomastigotos se diferencian en amastigotos. Estos, proliferan por fisión binaria y se diferencian a tripomastigotos celulares, que lisan las células y salen a la circulación pudiendo infectar células de diferentes tejidos, donde se transforman nuevamente en amastigotos. El insecto vector se infecta al alimentarse de un mamífero infectado que tiene tripomastigotos circulantes. Los tripomastigotos ingeridos se transforman en el intestino del vector en epimastigotos, donde se multiplican. En el intestino bajo del insecto, se diferencian a tripomastigotos metacíclicos infectivos cerrando el ciclo. Extraída y modificada del sitio *web* del "Centers for Disease Control y Prevention". <https://www.cdc.gov/parasites/chagas/biology.html>

Los parásitos en diferentes estadios presentan importantes diferencias morfológicas, metabólicas y de expresión génica. En el microscopio óptico, los estadios se distinguen por la posición relativa del núcleo y del flagelo; además, los estadios extracelulares son alargados y presentan flagelo, mientras que los amastigotos son más pequeños y redondos (de Souza, 2009). Los tripomastigotos, tanto metacíclicos como sanguíneos, no son replicativos, son capaces de infectar las células del hospedero y de evadir el sistema

inmune. Los tripomastigotas entran a las células por la vía fagocítica y la infección celular se produce cuando éstos escapan de las vacuolas fagocíticas. Una vez en el citoplasma de la célula, los parásitos se diferencian en amastigotas y el ciclo celular se reinicia. Durante esta transición, cambia el perfil de proteínas de superficie, comenzando a expresarse las amastinas. Luego de varias divisiones celulares, los amastigotas se diferencian a tripomastigotas y lisan la célula accediendo al torrente sanguíneo o la linfa. Cuando una vinchuca se alimenta de un individuo infectado, los tripomastigotas ingeridos en la sangre se transforman en epimastigotas, los cuales se dividen en el intestino del insecto. Previo a la diferenciación a epimastigotas, las proteínas de superficie son reemplazadas por un conjunto distinto de proteínas (Buscaglia *et al.*, 2004; Campo *et al.*, 2004).

Biología celular de *T. cruzi*

Los tripanosomátidos tienen varias características que los diferencian de los eucariotas modernos. Los TriTryps (tripanosomátidos modelo, *Trypanosoma brucei*, *Leishmania major* y *T. cruzi*) son parásitos que divergieron tempranamente (Haag *et al.*, 1998), esta distancia filogenética es causal de su biología excepcional (Smith y Parsons, 1996). En la figura 3 se observan las diferencias morfológicas más relevantes de los estadios, la posición relativa del núcleo y del flagelo. Así como la forma de los mismos, los estadios extracelulares son comparativamente más largos y flagelados mientras que los amastigotas son más pequeños y redondeados (de Souza *et al.*, 2009).

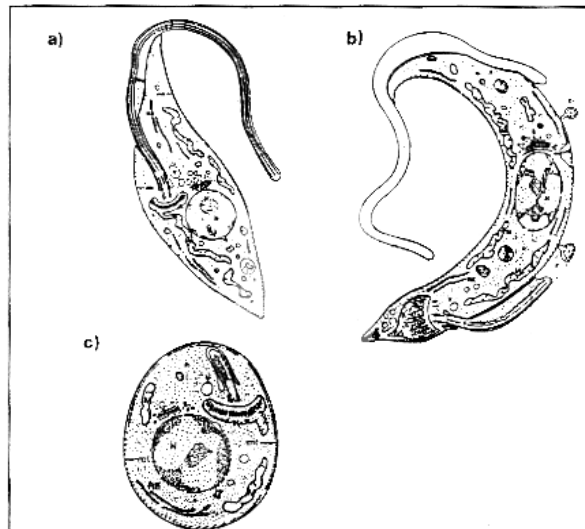


Figura 3. Esquema de tres de los estadios del parásito descritos por Chagas (a) Formas epimastigotas; **(b)** Formas tripomastigotas; **(c)** Formas amastigotas. En los dibujos se distinguen organelos y estructuras de la célula. Dibujos originales de Carlos Chagas (1987-1934).

En la figura 4 se muestra un esquema de una célula de un tripanosomátido, señalándose sus organelos y características estructurales más relevantes. Se indica la única mitocondria ramificada por todo el citoplasma, que contiene el kinetoplasto. Esta estructura se compone de un conjunto de moléculas circulares de ADN topológicamente relajadas conocidas como maxi y mini círculos. Existen varios miles de minicírculos, de entre 0,5 a 2,5 kb y unas pocas decenas de maxicírculos, de entre 20 y 40 kb (Shapiro y Englund, 1995) que contienen los genes que codifican los ARNr y proteínas sintetizadas en la mitocondria. En tripanosomas los ARNs mensajeros mitocondriales sufren la adición o eliminación de uridinas en lugares específicos (*editing*). Los ARNs que guían este fenómeno son codificados por los minicírculos (Shaw *et al.*, 1988; Simpson *et al.*, 2000; Madison-Antenucci *et al.*, 2002; Simpson, 2003).

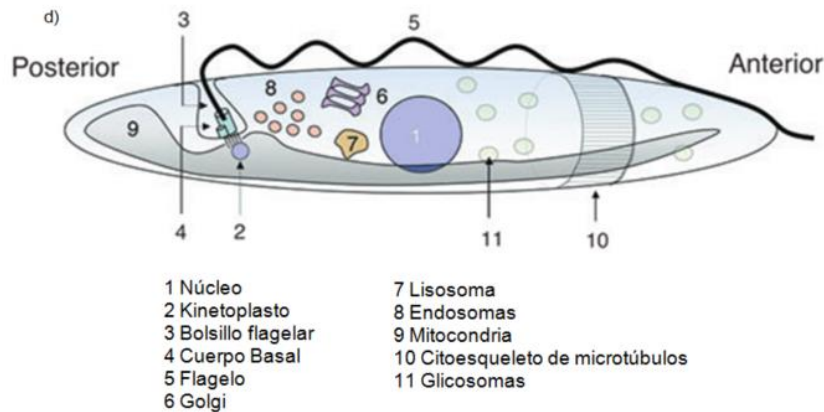


Figura 4. Esquema de una célula de *T. brucei* mostrando sus características estructurales principales La organización general es similar entre los diferentes tripanosomátidos. Extraído y modificado de (Matthews, 2005).

El organelo denominado glicosoma es también característico de estos parásitos (Cazzulo, 1994). Varias etapas de la glucólisis tienen lugar dentro de este organelo. La compartimentación de las enzimas glucolíticas es esencial para el parásito, tanto en estadios que dependen exclusivamente de la glucólisis para la obtención de ATP, como en estadios que obtienen ATP por otras vías (Opperdoes y Borst, 1977; Guerra-Giraldez *et al.*, 2002). Esta compartimentación de algunas enzimas involucradas en la obtención de energía y en el metabolismo de carbohidratos y lípidos incrementa su flexibilidad metabólica y por lo tanto, permite a los parásitos adaptarse rápida y eficientemente a diferentes condiciones ambientales, contribuyendo al establecimiento de los complejos ciclos de vida de los kinetoplastidos (Gualdron-Lopez *et al.*, 2012).

Los acidocalcisomas son organelos capaces de transportar protones y calcio y han sido identificados en todos los miembros de la familia Trypanosomatidae y muchos miembros del *phylum* Apicomplexa (Docampo *et al.*, 2005; Docampo y Moreno, 2011). Los acidocalcisomas están involucrados en varias funciones incluyendo: almacenamiento de calcio, magnesio, sodio, potasio, zinc, hierro, pirofosfato inorgánico, polifosfato; homeostasis del pH y osmorregulación (Linder y Staehelin, 1977). Su función no está del todo clara, pero se postula que podría tener un rol en la regulación del metabolismo de fosfatos (Ulrich *et al.*, 2011).

Otra característica destacable de estos organismos es el proceso de mitosis, ya que en estos parásitos la envoltura nuclear no se desensambla durante la división celular y el huso mitótico se forma dentro del núcleo. Además, la cromatina no se condensa hasta cromosomas discretos aunque sí se producen cambios de compactación del ADN en diferentes etapas del ciclo (Vickerman, 1970; De Souza y Meyer, 1974; Drechsler y McAINSH, 2012). Se ha demostrado también la permanencia del nucléolo durante toda la división en *T. brucei* (De Souza, 2002; Schenkman *et al.*, 2011).

Por otro lado, como en otros eucariotas, la endocitosis es el mecanismo principal de internalización de macromoléculas. Estudios en *T. cruzi* evidenciaron que la actividad endocítica está confinada al bolsillo flagelar y al citosoma (invaginación profunda de la membrana plasmática que puede llegar hasta el núcleo) (de Souza, 2009). Esta vía tiene varias particularidades, la endocitosis ocurre únicamente en la forma epimastigota, además, las moléculas incorporadas por los endosomas son enviadas a los reservosomas, organelos especiales localizados en la región posterior del parásito (Porto-Carreiro *et al.*, 2000).

1.3. El genoma de *T. cruzi*

A mediados de los años 90 comenzó la secuenciación de los genomas de los TriTryps mediante el análisis de grandes fragmentos de ADN clonados en librerías de cósmidos, cromosomas artificiales de bacterias y de levaduras. Estos estudios junto con análisis de electroforesis de campo pulsado y secuenciado de ESTs (*expressed sequence tags*), permitieron construir los primeros mapas físicos de los genomas de los TriTryps (Blackwell y Melville, 1999). En el año 2005 se publicaron los genomas de los tres tripanosomátidos modelo que han servido de base para numerosos estudios incrementando nuestra comprensión de la biología de estos parásitos y han brindado pistas importantes de los mecanismos de interacción hospedero-parásito (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005; El-Sayed *et al.*, 2005b; Ivens *et al.*, 2005).

La cepa CL Brener de *T. cruzi*, que fue la primera en ser secuenciada, resultó ser una cepa híbrida, presentando dos haplotipos (El-Sayed *et al.*, 2005). Actualmente las cepas de *T. cruzi* se asignan a 6 grupos principales que van del I al VI (Zingales *et al.*, 2009). CL Brener pertenece al grupo VI, híbrido formado por los grupos II y III. Este hecho sumado a la gran cantidad de secuencias repetitivas encontradas, generó problemas al momento de ensamblar el genoma (estimado en 55 Mb para el genoma haploide). Para resolver estos problemas, se siguieron varias estrategias complementarias. En primer lugar, el genoma se secuenció a mayor profundidad que la utilizada usualmente para secuenciar un borrador de un genoma. Así fue posible distinguir las variantes alélicas, de los errores de secuenciación. En segundo lugar, para distinguir los haplotipos provenientes de cada cepa original, un miembro de uno de los grupos progenitores (cepa Esmeraldo) fue secuenciado a baja cobertura. A su vez, haciendo uso de los cromosomas ensamblados de *T. brucei*, Weatherly *et al.* lograron ensamblar los *contigs* originales en 41 cromosomas, coincidiendo con el análisis de electroforesis de campo pulsado (Branche *et al.*, 2006). Actualmente se están secuenciando los genomas de varias cepas de *T. cruzi* mediante la tecnología PacBio, que permite la secuenciación de fragmentos de ADN que promedian 12,5 kb pero pueden ser tan largos como 50 kb. Resultado de estos proyectos, la cepa Sylvio X10, está disponible en la base de datos de TriTrypDB aunque no está anotada aún (Talavera-López, 2016). Esta cepa pertenece al grupo I de *T. cruzi*. El tamaño de este genoma, que se considera característico del grupo I, se estimó en 44 Mb. Aparte de esta diferencia en tamaño, los genomas revelan una organización similar. Esta cepa mostró mayor identidad de secuencia con el haplotipo no Esmeraldo de CL Brener, confirmando que *T. cruzi* I está más emparentado con *T. cruzi* III, de donde proviene el haplotipo no Esmeraldo (Franzen *et al.*, 2011). Existen disponibles también otros genomas, el de la cepa Dm28c de (Grisard *et al.*, 2014) que no está bien anotada aún por lo cual es difícil utilizarlo como referencia para estudios transcriptómicos, el de la cepa marinkelleiB7 que no es patógeno humano (Franzén *et al.*, 2012) y los de las cepas JRcl4 y Tulahuen cl2, no anotadas aún.

El secuenciado de los maxicírculos mitocondriales de varias cepas de *T. cruzi* ha permitido realizar inferencias de los mecanismos de generación de los 6 subgrupos que hoy en día se reconocen para este organismo (Zingales *et al.*, 2009; Ruvalcaba-Trejo y Sturm, 2011).

Con estos datos los autores proponen que una cepa tipo I por hibridación con cepas de tipo II, resultó en la generación de las cepas tipo III y IV. Luego una hibridación entre cepas tipo II y III resultó en los tipos V y VI. El genoma, de 25 Kb, de los maxicírculos codifica 18 proteínas mitocondriales y dos genes de ARNr. Estas secuencias mostraron que 15 de los genes codificantes de proteínas tienen cambios en los marcos de lectura, confirmando la necesidad de la edición de los mensajeros para su correcta expresión (Westenberger *et al.*, 2006).

El genoma haploide de *T. cruzi* contiene aproximadamente 12.000 genes que codifican para proteínas, pero solamente se pudo asignar una función putativa a la mitad. Se identificaron también 1.994 genes para ARN no codificante y 3.590 pseudogenes (El-Sayed *et al.*, 2005). Al menos la mitad del genoma está formado por secuencias repetidas, que consisten en retrotrasposones, repetidos subteloméricos y familias multigénicas. El gran número de familias multigénicas encontrado explica en gran medida, el hecho de que *T. cruzi* es el tripanosomátido con más cantidad de genes. Es así que, el 18% de las secuencias codificantes están presentes en más de 14 copias (Arner *et al.*, 2007). Las familias multigénicas más expandidas, consisten en proteínas tipo trans-sialidasas, mucinas, metaloproteasas, DGF-1 (*dispersed gene family protein 1*), proteínas RHS (*retrotransposon hot spot*) y las proteínas de superficie asociadas a mucinas (*mucin associated surface proteins*, MASP) nombradas así por su asociación en el genoma con los genes de las mucinas (De Pablos y Osuna, 2012). Esta familia, que era desconocida hasta ese momento, representa un 6% del genoma diploide del parásito (Bartholomeu *et al.*, 2009; dos Santos *et al.*, 2012).

Familias multigénicas de *T. cruzi*

Las familias multigénicas que codifican para antígenos de superficie son abundantes en *T. cruzi*, y su expresión está regulada a lo largo del ciclo de vida del parásito. Los tripomastigotas de *T. cruzi* no poseen el mecanismo de variación antigénica que posee *T. brucei*, sino que expresan en su superficie varias proteínas diferentes pertenecientes a familias multigénicas (El-Sayed *et al.*, 2005), siendo las más caracterizadas las mucinas (Di Noia *et al.*, 1998; Buscaglia *et al.*, 2006), las trans-sialidasas (Schenkman *et al.*, 1994; Nardy

et al., 2016), las MASP (Bartholomeu *et al.*, 2009; dos Santos *et al.*, 2012), la familia de genes dispersos (DGF) (Kawashita *et al.*, 2009; De Pablos y Osuna, 2012) y la familia de glicoproteínas de superficie gp63 o *Major Surface Protease* (MSP) (Cuevas *et al.*, 2003; Yao, 2010). La función de las mismas se ha asociado frecuentemente con protección y evasión del sistema inmune del hospedero (dos Santos *et al.*, 2012; De Pablos *et al.*, 2016), particularmente importante en el estadio tripomastigota sanguíneo.

En *T. cruzi* existe una familia de proteínas multigénica que contienen un *hot spot* para la inserción de retroelementos (RHS) cuya expresión aumenta ante la exposición a radiación γ (Grynberg *et al.*, 2012), descrita por primera vez en *T. brucei* (Bringaud *et al.*, 2002).

Como fue descrito previamente, las familias multigénicas están altamente representadas en los genomas de los tripanosomátidos (El-Sayed *et al.*, 2005). Recientemente, Jackson *et al.* 2016, plantean que la adquisición de las familias multigénicas de superficie en los tripanosomátidos está relacionada a la adaptación al parasitismo, ya que estas familias no están presentes en el tripanosomátido de vida libre más cercano filogenéticamente, *Bodo saltans* (figura 5). Se sostiene que la expresión de estas proteínas permite a la célula enfrentar el ambiente hostil al que está expuesto dentro del hospedero (Jackson *et al.*, 2016).

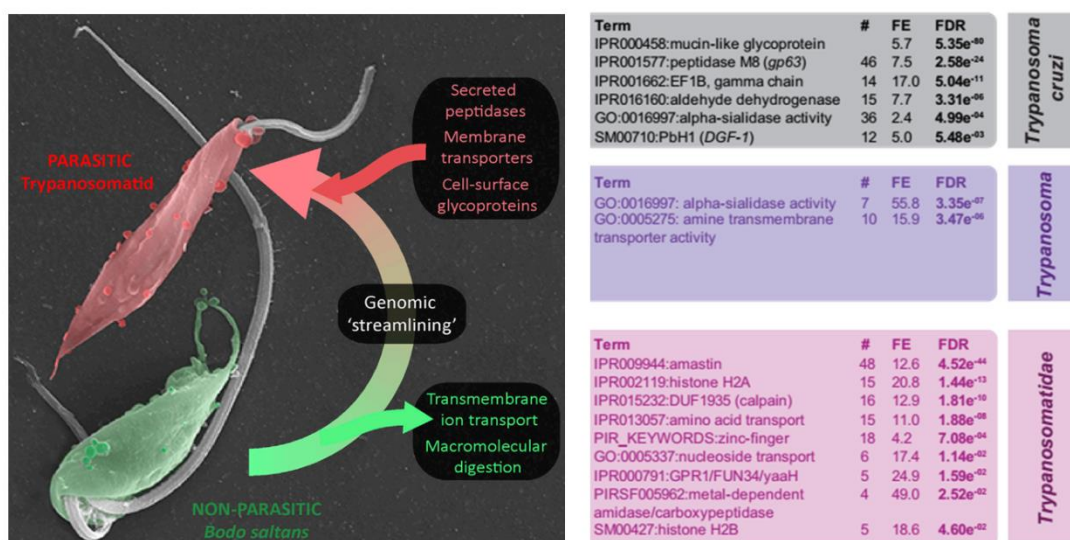


Figura5. Adaptación de tripanosomátidos al parasitismo. A la izquierda, esquema de la pérdida y ganancia de genes de los tripanosomátidos durante su adaptación al parasitismo. A la derecha, análisis de agrupamiento de repertorios génicos en el contexto de la filogenia de los

kinetoplástidos. El sombreado indica los agrupamientos específicos, *T. cruzi* (negro), tripanosomas (azul) y tripanosomátidos (violeta). Se indican también los respectivos enriquecimientos, número involucrado (#), el *fold enrichment* (FE), y el *p value* corregido por *false Discovery rate* (FDR). Tomado y modificado de Jackson *et al.* 2016.

1.4. Expresión génica en tripanosomátidos

En muchos puntos del ciclo de vida, el parásito sufre la reprogramación de la expresión de sus genes para adaptarse a los diferentes ambientes a los que se enfrenta, sin embargo, los mecanismos que dirigen estos procesos aún no se han definido claramente. Probablemente debido a la distancia filogenética respecto a otros eucariotas (los flagelados aparecieron hace aproximadamente 1500 millones de años), los tripanosomátidos presentan características biológicas excepcionales. Los múltiples niveles de regulación del inicio de la transcripción, presentes en la mayoría de los eucariotas, no están presentes en los tripanosomátidos (De Gaudenzi *et al.*, 2011; Kramer, 2012). La mayoría de los genes que codifican para proteínas no poseen promotores propios y se expresan en largas unidades policistrónicas, que a diferencia de lo que sucede en los procariontes, no codifican para productos funcionalmente relacionados (El-Sayed *et al.*, 2005), de hecho, genes adyacentes, pueden codificar para proteínas de expresión diferencial en el ciclo de vida del parásito (Minning *et al.*, 2009; Smircich *et al.*, 2015).

Una vez transcritos, los policistrones son procesados dentro del núcleo a mensajeros individuales mediante la adición por *trans-splicing* de un oligonucleótido de ARN (denominado miniexón) en el extremo 5' (Agabian, 1990) y la poliadenilación del extremo 3' (Ullu *et al.*, 1993). El procesamiento de los mensajeros, sucede de manera acoplada, produciéndose simultáneamente la poliadenilación de un mensajero y el *trans-splicing* del mensajero ubicado inmediatamente río abajo (LeBowitz *et al.*, 1993; Matthews *et al.*, 1994; Liang *et al.*, 2003). Al igual que en otros eucariotas, el proceso de *trans-splicing* es mediado por el spliceosoma, un complejo de ribonucleoproteínas pequeñas nucleares

(*small nuclear ribonucleoproteins*, snRNPs) otras proteínas y ARN pequeños (*small nuclear RNA*, snRNAs). Entre estos, SLRNP, U2, U4/U6 y U5 son factores esenciales para el *trans-splicing* (Liang *et al.*, 2003). En *T. brucei* se han reportado estudios de secuenciación masiva mediante RNA-Seq que permitieron la descripción de más de 30 mil sitios de adición del miniexón en aproximadamente 9.000 genes diferentes, el mapeo de los sitios de procesamiento a nivel global permitió la definición de las respectivas regiones UTRs. (Siegel *et al.*, 2011). Los resultados obtenidos muestran que un 85% de los genes presentan en promedio alrededor de 3 posibles sitios de procesamiento en el extremo 5'. La cantidad de transcripto correspondiente a la variante principal es menos de un 60% del total, indicando que las cantidades de las formas alternativas son fisiológicamente significativas (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010). Respecto a los sitios de poliadenilación, se observó un número aún mayor de transcritos alternativos, describiéndose más de 50 mil sitios de poliadenilación en unos 8.000 genes (Siegel *et al.*, 2010). Este fenómeno, aunque menos pronunciado, fue reportado también para *T. vivax* y *L. major* (Greif *et al.*, 2013; Rastrojo *et al.*, 2013). Aún no está claro si esta gran variedad de transcritos alternativos es un reflejo de una baja fidelidad de la maquinaria de procesamiento, aunque se ha sugerido que podría ser un fenómeno regulado y, por lo tanto, con consecuencias funcionales, al mostrar cada variante diferentes secuencias en las regiones UTRs (Kolev *et al.*, 2010; Nilsson *et al.*, 2010; Siegel *et al.*, 2010).

A su vez, es interesante destacar que en *T. cruzi* se demostró la presencia en el citoplasma de un dicistrón que involucra a las proteínas de unión a ARN TcUBP1 y 2 y su procesamiento citoplasmático por *trans-splicing* (Jager *et al.*, 2007) este fenómeno incrementa los posibles niveles de regulación de la expresión génica en estos organismos.

Otra característica sobresaliente de los kinetoplastídeos, es la ausencia de intrones, aunque existen excepciones a esta regla general. Los genes que codifican para la proteína poliA polimerasa (PAP) de *T. cruzi* y *T. brucei* están interrumpidos por un intrón de 305 pares de bases, estableciéndose que el mecanismo de *cis-splicing* está también presente en estos protozoarios (Mair *et al.*, 2000). Además, el gen que codifica para el ARNt-Tyr también contiene un intrón (Tan *et al.*, 2002; Padilla-Mejía *et al.*, 2009). Con la

secuenciación de los genomas completos se pudo establecer que en *T. brucei*, esta regla podría también romperse en el caso de otros genes como el que codifica para una helicasa de ARN y dos que codifican proteínas hipotéticas (Ivens *et al.*, 2005). Los estudios de RNA-Seq mencionados anteriormente permitieron la confirmación experimental del intrón presente en la PAP (Tb927.3.3160) y en la helicasa (Tb927.8.1510) de *T. brucei* (Kolev *et al.*, 2010; Siegel *et al.*, 2010) no así, los de las dos proteínas hipotéticas.

El proceso de *cis-splicing* requiere el reconocimiento del sitio 5' del intrón del pre-ARNm por el snRNA U1 por apareamiento de bases. Los dos elementos han sido descritos en los tripanosomátidos (Djikeng *et al.*, 2001). Es interesante destacar que, aunque la ausencia del primer exón de la PAP en el transcripto, no produce una proteína funcional, se ha reportado la existencia de competencia entre el *cis-* y el *trans-splicing* del sitio 3' de *splicing* (Mair *et al.*, 2000).

Regulación postranscripcional de la expresión génica en los tripanosomátidos

Debido a la falta de control de inicio de la transcripción, la regulación de la expresión ocurre a nivel de etapas postranscripcionales. Además del procesamiento de los mensajeros, antes mencionado, existen varias etapas a controlar, incluyendo la degradación en el núcleo, el transporte al citoplasma, la estabilidad, la localización subcelular y la traducibilidad, entre otros.

La regulación del nivel de mensajeros disponible para ser traducido es un punto importante en la regulación de la expresión génica, por lo tanto, la tasa de degradación de los ARNs a nivel citoplasmático es un proceso altamente regulado, en el cual están involucradas varias enzimas. En eucariotas, la degradación de los mensajeros comienza por la acción de exonucleasas en ambos extremos luego de la remoción de las estructuras protectoras (o sea el CAP y la cola poliA). El proceso ha sido estudiado profundamente en eucariotas, (Coller *et al.*, 2001; Coller y Parker, 2005; Balagopal y Parker, 2009) y recientemente, se identificó la enzima encargada de la remoción del cap en *T. brucei* (Ignatochkina *et al.*, 2015). La maquinaria de degradación de la vía 5' y otras proteínas del metabolismo del ARN junto con ARNm específicos, se concentran

formando gránulos en el citoplasma que se denominan cuerpos de procesamiento (*P-bodies*), gránulos de estrés o gránulos de ARN, dependiendo de su composición y el momento en el que se forman en la célula (Gerber *et al.*, 2004; Parker y Sheth, 2007). En tripanosomátidos se han encontrado gránulos citoplasmáticos de características análogas y que contienen proteínas con las funciones homólogas a las encontradas en eucariotas superiores (Cassola *et al.*, 2007; Holetz *et al.*, 2007; Kramer *et al.*, 2008, 2010; Kramer, 2014). Este tipo de estructuras funcionarían como reservorio de mensajeros, tanto en condiciones normales como en condiciones de estrés, pudiendo los ARN ser degradados o volver a la población de ARN traducibles (Cassola, 2011; Cassola *et al.*, 2007; Holetz *et al.*, 2007; Kramer *et al.*, 2008). Se ha sugerido un mecanismo regulatorio que puede dar cuenta de la expresión diferencial de genes cuya concentración de ARNm en estado estacionario, no se ve modificada (Saas *et al.*, 2000). Este mecanismo consiste en la movilización diferencial, de determinados mensajeros, desde los gránulos citoplasmáticos hacia la fracción polisomal. Consistentemente con la hipótesis de estas estructuras, reservorios de ARNm no traducibles, se han encontrado evidencias experimentales de regulación traduccional (Gale *et al.*, 1994; Avila *et al.*, 2001; Boucher *et al.*, 2002; Mayho *et al.*, 2006; Nardelli *et al.*, 2007). La regulación de este proceso estaría dada a nivel de la formación del complejo de iniciación de la traducción y del paso subsiguiente de elongación (McCarthy, 1998). Este mecanismo de regulación coexiste con los antes mencionados ya que paralelamente, algunos genes están siendo regulados además a través de la estabilización de sus ARN mensajeros (Avila *et al.*, 2001).

Otro nivel de regulación de la expresión génica es la localización diferencial de los mensajeros en la célula. Este mecanismo garantiza la traducción localizada de los ARNm y, por lo tanto, la ubicación adecuada de las proteínas, en el sitio donde su función es requerida (Jung *et al.*, 2014). Este mecanismo ha sido expuesto en los tripanosomátidos (Cassola *et al.*, 2007; Holetz *et al.*, 2007; Cassola, 2011; Kramer, 2014).

Recientemente, se ha enfatizado en el estudio del rol de la eficiencia traduccional de los mensajeros como forma de regular la abundancia proteica (Schwanhausser *et al.*, 2011), mecanismo que ha sido demostrado también en los tripanosomátidos (Jensen *et al.*, 2014; Vasquez *et al.*, 2014; da Silva Augusto *et al.*, 2015; Smircich *et al.*, 2015).

Aunque se asume que los motivos en *cis* responsables de la regulación de los niveles de mensajero disponibles para ser traducidos se concentran en las regiones 3' UTR, las regiones 5'UTR son importantes en los procesos de control traduccional. Los genes de expresión constitutiva, como los que codifican las proteínas ribosomales, no poseen regiones 5' UTR, estando el codón de inicio inmediatamente después del miniexón. Los autores postulan que de esta forma, se evitaría la presencia de reguladores negativos en esta región (Greif *et al.*, 2013). La tasa traduccional puede estar influida también por la selección de codones sinónimos presentes en cada gen. La hipótesis plantea que los genes de expresión alta estarían optimizados para mejorar la eficiencia de la maquinaria traduccional y/o su fidelidad (Hershberg y Petrov, 2008). En tripanosomátidos, los genes de alta expresión presentan un uso de codones particular (Parsons *et al.*, 1991; Alvarez *et al.*, 1994). La disponibilidad de los genomas completos sumado al conocimiento de los correspondientes niveles de expresión a escala global ha permitido confirmar este hecho. Además, se corroboró que las moléculas de ARNt que reconocen los codones óptimos son las más abundantes, reforzando este hallazgo (Horn, 2008).

Actores moleculares en la regulación de la expresión génica en los tripanosomátidos

Los múltiples niveles de regulación de la expresión génica en estos organismos han sido estudiados con diferentes profundidades, dedicando gran esfuerzo al estudio de factores que actúan en *trans* (principalmente a nivel de proteínas de unión al ARN, RBPs) (Martinez-Calvillo *et al.*, 2010; Kramer y Carrington, 2011; Clayton, 2013; Droll *et al.*, 2013; Perez-Diaz *et al.*, 2013) y su interacción con secuencias en *cis*, principalmente localizadas en las regiones no traducidas de los mensajeros (UTRs). Los elementos en *cis*, son secuencias que modulan o dirigen de alguna manera los diferentes eventos de la dinámica de los ácidos nucleicos, tanto su ubicación como su función están relativamente conservadas entre especies (Wittkopp y Kalay, 2011). Se ha demostrado que las RBPs establecen interacciones con grupos de ARNs que comparten elementos en *cis*, definiendo grupos co-regulados de mensajeros que podrían cumplir funciones relacionadas (Keene, 2007). Estos grupos de genes formarían parte de un operón

postranscripcional o regulón. Este mecanismo está presente también en los tripanosomátidos (Ouellette y Papadopoulou, 2009; Queiroz *et al.*, 2009). En *T. brucei*, la proteína PUF9 formaría parte de un regulón postranscripcional involucrado en la replicación de estructuras subcelulares, estabilizando a sus blancos y generando la acumulación de los mismos inmediatamente después de la replicación del ADN. Los ARNs blancos de PUF9 comparten un motivo regulatorio en *cis* en sus 3' UTR (Archer *et al.*, 2009).

En *T. cruzi* han sido identificados varios elementos en *cis* que modulan la expresión de transcritos (Weston, La Flamme y Van Voorhis, 1999), incluidos genes regulados durante el ciclo de vida (Coughlin *et al.*, 2000; Bayer-Santos *et al.*, 2012; Li *et al.*, 2012; Alves *et al.*, 2014) y genes funcionalmente relacionados (De Gaudenzi *et al.*, 2013). Entre ellos, se han descrito los elementos ricos en AU similares a ARE (*AU rich elements*) en la región 3'UTR de varios mensajeros y se han identificado algunas proteínas estadio específicas que los reconocen (TcUBPs) (Di Noia *et al.*, 2000; D'Orso y Frasch, 2001b). En el 3' UTR de estos mensajeros se identificó también un elemento rico en GU (GRE) en las regiones 3'UTR de los genes tipo mucinas TcSMUG responsables de la regulación estadio específica de los mensajeros que lo contienen (D'Orso y Frasch, 2001a), estos elementos han sido ampliamente estudiados en otros organismos (Vlasova-St Louis y Bohjanen, 2011). Por otro lado, los motivos ricos en AU han sido relacionados a la expresión de los mensajeros que los contienen en el estadio epimastigota, mientras que en otros estadios los desestabilizan (Suganuma *et al.*, 2012). Otro elemento, que parece ser estructural y funcionalmente relacionado al elemento ARE es el UREs (*U rich instability elements*), estos han sido encontrados en varios ARNs, afectando su estabilidad (Haile y Papadopoulou, 2007). Además, un elemento rico en U de 43 nucleótidos controla los niveles de varios mensajeros en el estadio amastigota del parásito (Li *et al.*, 2012). En otros casos, la señal regulatoria es menos clara, postulándose que es la combinación de motivos presentes en el UTR la responsable de la regulación estadio específica (Bayer-Santos *et al.*, 2012; Pastro *et al.*, 2013). En este sentido, se han estudiado varios elementos en *cis* presentes en el ADN que podrían estar involucrados en la regulación de la expresión génica (Duhagon *et al.*, 2001, 2011; Pastro *et al.*, 2013) a su vez, para los

repetidos de dinucleótidos TG, se identificó una proteína que los reconoce (Duhagon *et al.*, 2003, 2009).

Por otra parte, se ha demostrado que la proteína TcPUF6 (Dallagiovanna *et al.*, 2005, 2008) se asocia mayoritariamente a ARNm que contienen el motivo UGUR desestabilizándolos. Este elemento, que es blanco de la familia de proteínas PUF en eucariotas, ha sido asociado a la regulación de la estabilidad y traducción de los mensajeros, aumentando la deadenilación y la consecuente degradación de los mismos o suprimiendo el inicio de la traducción (Wickens *et al.*, 2002; Gerber *et al.*, 2004; Wharton y Aggarwal, 2006; Wang *et al.*, 2009).

Más recientemente fueron identificados varios motivos involucrados en la regulación del procesamiento del dicistrón que contiene los genes codificantes de las proteínas TcUBP1 y 2, estos motivos, ricos en AU o GA y trectos ricos en pirimidinas se localizan en la región inter-cistrónica (ICR, región intergénica) y son blancos de la proteína TcDRBD4/PTB2. Esta proteína regula negativamente el *splicing* del dicistrón, por lo que podría representar otro nivel de regulación de la expresión génica en los tripanosomátidos, enmascarando las señales de *trans-splicing*/poliadenilación (De Gaudenzi *et al.*, 2016).

Si bien se sabe que ambas UTRs pueden influir en los niveles del mensajero, la región 3' es en general entre dos y tres veces más larga que la 5' (posiblemente por cómo se generan durante el procesamiento de los policistrones) y presenta mayor nivel de secuencias regulatorias en *cis* (Campos *et al.*, 2008). Las regiones 5'UTR de los genes en *T. cruzi* tienen entre 10 y 400 nt mientras que las 3'UTR tienen entre 17 y 2800 nt (Brandão y Jiang, 2009).

Consistentemente con la existencia de un gran número de señales diferentes, el análisis de los genomas de los TriTryps revela la existencia de un gran número de proteínas de unión al ARN (Kramer y Carrington, 2011; Houston-Ludlam *et al.*, 2016). Sin embargo, sólo algunas han sido caracterizadas (Clayton, 2002; D'Orso *et al.*, 2003; Dallagiovanna *et al.*, 2005; Perez-Diaz *et al.*, 2007, 2012; Guerra-Slompo *et al.*, 2012; Morking *et al.*, 2012; Alves *et al.*, 2015; Pérez-Díaz *et al.*, 2017). Varios integrantes de la familia de proteínas con dominios RRM (*RNA Recognition Motif*) han sido estudiados (D'Orso y Frasch, 2001b, 2002; Manger y Boothroyd, 1998; Hartmann *et al.*, 2007; Hartmann y Clayton, 2008;

Perez-Diaz *et al.*, 2007; Guerra-Slompo *et al.*, 2012). El dominio RRM está presente, en una o varias copias, en muchas proteínas eucariotas que reconocen ARN simple hebra. Está compuesto por aproximadamente 90 aminoácidos y posee dos regiones altamente conservadas: un dominio hidrofóbico de seis residuos (llamado motivo RNP-2) y un octapéptido (llamado RNP-1 o RNP-CS) (Bandziulis *et al.*, 1989). Varias proteínas de la familia de proteínas con dominios de tipo dedos de zinc han sido también estudiadas (Hendriks *et al.*, 2001; Morking *et al.*, 2004, 2012; Hendriks y Matthews, 2005; Paterou *et al.*, 2006; Alves *et al.*, 2014). También han sido descritas proteínas de la familia Pumilio como fue expuesto previamente (Dallagiovanna *et al.*, 2005, 2008).

Estas proteínas están vinculadas a la regulación de la expresión génica a nivel de mensajeros interviniendo en los procesos de maduración (Matthews *et al.*, 1994; Xu *et al.*, 2001; Portal *et al.*, 2003; Vazquez *et al.*, 2003), edición (Madison-Antenucci *et al.*, 2002), recambio o traducibilidad (D'Orso *et al.*, 2003; Haile *et al.*, 2003; Haile y Papadopoulou, 2007; Perez-Diaz *et al.*, 2013).

Existen también otros factores en *trans* capaces de afectar la expresión de los genes, son moléculas de pequeños ARNs, que interaccionan con los mensajeros y generalmente regulan negativamente la expresión del gen (Jonas e Izaurralde, 2015). En los tripanosomátidos, la mayor parte de la información proviene de *T. brucei* y, aunque se han predicho moléculas de micro ARNs, hasta el momento se han encontrado experimentalmente únicamente moléculas de ARNs pequeños interferentes (siRNAs) (Mallick *et al.*, 2008). La maquinaria sería responsable principalmente del control de las secuencias transponibles, dando estabilidad al genoma (Atayde *et al.*, 2011). Sin embargo, la existencia de siRNAs derivados de pseudogenes abre la posibilidad de otras funciones reguladoras (Wen *et al.*, 2012). El mecanismo de interferencia no está presente en *T. cruzi*, y la población de ARNs pequeños identificada, deriva de secuencias de ARNs de transferencia no estando clara aún la función de los mismos (García-Silva *et al.*, 2010; Franzen *et al.*, 2011) aunque se ha propuesto que tendrían un rol diferencial durante el ciclo de vida del parásito (Reifur *et al.*, 2012) siendo secretados en vesículas en el estadio epimastigota, por lo que se propone un rol en la interacción entre parásitos y el hospedero (García-Silva *et al.*, 2014).

Aunque el estudio de la regulación de la expresión génica en tripanosomátidos ha permitido identificar diferentes procesos así como varias señales y factores regulatorios (Martinez-Calvillo *et al.*, 2010; Kramer y Carrington, 2011), no hay estudios que se dediquen a discernir entre la contribución de los procesos nucleares y citoplásmicos.

1.5. El rol del núcleo en la regulación de la expresión génica

En los tripanosomátidos, la regulación de la expresión génica ocurre mayormente a nivel postranscripcional y la regulación de la disponibilidad de los mensajeros en el citoplasma, donde tienen lugar los procesos de degradación, localización y traducibilidad del mensaje (Halbeisen *et al.*, 2008), puede tener gran importancia. Por lo tanto, el proceso de transporte hacia el citoplasma debe jugar un rol particularmente relevante en estos organismos.

La arquitectura central de la célula eucariótica se estableció hace más de mil millones de años y se conserva en gran medida en todos los linajes existentes (Cavalier-Smith, 2010). Sin embargo, las células eucarióticas también poseen características específicas de linaje, frecuentemente identificadas con requisitos funcionales específicos. El complejo de poros nucleares (NPC), es la estructura eucariótica central responsable de regular el intercambio de macromoléculas entre el núcleo y el citoplasma, está compuesto mayormente por nucleoporinas (Rout *et al.*, 2000; Cronshaw *et al.*, 2002). El andamio estructural central del NPC está conservado en tripanosomátidos, pero los elementos periféricos presentan pérdidas, duplicaciones u otras alteraciones en sus componentes (DeGrasse y Devos, 2010). Los tripanosomas carecen de los componentes principales de la plataforma de exportación de mensajeros que están asimétricamente localizados dentro de los NPCs en levaduras y vertebrados, reflejando funcionalidad divergente. De hecho, en los tripanosomátidos, el proceso de exportación de ARNm en la cara citoplasmática de NPC es dependiente de un gradiente Ran-GTP/Ran-GDP entre el núcleo y el citoplasma, similar al transporte de proteínas y otros ARNs, mientras que en organismos superiores el transporte de mensajeros es dependiente de ATP.

Concomitantemente, el NPC de tripanosoma es completamente simétrico, siendo la cesta nuclear la única fuente importante de asimetría. Estas características apuntan hacia una evolución escalonada del NPC en el que un andamio de revestimiento estabilizó por primera vez el poro después del cual la entrada selectiva emergió y se expandió, llevando a la adición de maquinarias de remodelación periférica en los lados nucleoplásmicos y citoplásmicos (Wente y Rout, 2010; Obado *et al.*, 2016; Obado *et al.*, 2017). El NPC, media el transporte de las moléculas interaccionando transitoriamente con proteínas de la familia karyoferinas β (llamadas también importinas y exportinas) (Strambio-De-Castillia *et al.*, 2010) las cuales son receptores nucleares conservados (Strom y Weis, 2001) y otras proteínas que no pertenecen a esta familia (Herold *et al.*, 2003). El proceso de exportación se inicia durante la transcripción al asociarse, las proteínas específicas (complejo THO) a los mensajeros nacientes formando un complejo ribonucleoproteico (mRNP) competente para ser exportado al citoplasma (Jimeno y Aguilera, 2010). Este mRNPs recluta otros factores de procesamiento y de exportación, resultando en la formación del complejo de transcripción y exportación (TREX). Este complejo interacciona con el spliceosoma y finalmente, los mensajeros procesados son exportados a través del NPC. Una vez en el citoplasma, el complejo es desensamblado por la helicasa dependiente de ATP, Dbp5 (Kohler y Hurt, 2007).

La identidad del ARNm a ser exportado es rigurosamente evaluada por la maquinaria de exportación e involucra señales en el ARNm en vertebrados (Palazzo y Akef, 2012). En metazoarios, los intrones también pueden actuar como señales de exportación ya que, la maquinaria de *splicing* se acopla al complejo TREX (Luo y Reed, 1999; Valencia *et al.*, 2008). En levaduras, probablemente debido a que la mayor parte de los genes no poseen intrones, el complejo TREX se carga en los mensajeros co-transcripcionalmente por acción de ARN polimerasa II (Chávez *et al.*, 2000; Jimeno *et al.*, 2002). Recientemente fue demostrado que el ARNm de la β -globina es retenido en el núcleo mediante una señal ubicada en el intrón, que al ser eliminada del mensajero por *splicing*, permite su exportación del núcleo (Akef *et al.*, 2017). A su vez, han sido identificadas señales en ARNs largos no codificantes responsables de su retención nuclear (Zhang *et al.*, 2014).

En los tripanosómátidos, los complejos TREX y THO no están presentes (Serpeloni *et al.*, 2011; Erben *et al.*, 2014) pero el complejo exportina 1 es funcional (Zeiner *et al.*, 2003;

Cuevas *et al.*, 2005). En *T. cruzi*, se demostró que la proteína DEAD/H RNA helicasa, Hel45, está involucrada en el transporte de los mensajeros hacia el citoplasma de forma dependiente de la proteína receptora Mex67 (Inoue *et al.*, 2014). La proteína TcUBP1 ha sido vinculada también al transporte de mensajeros desde el núcleo al citoplasma (Cassola y Frasch, 2009). Además, en *T. cruzi*, ante un detenimiento de la traducción, la biogénesis de los ARN ribosomales se detiene, en un proceso regulado por la adenilato kinasa (TcADKn), que funciona como proteína transportadora de ARN entre núcleo y citoplasma (Camara *et al.*, 2013).

Las aproximaciones para el estudio de la expresión génica global en *T. cruzi* se han basado en el uso de homogenados de diferentes estadios del ciclo de vida del parásito mediante análisis de microarreglos (Minning *et al.*, 2009) y más recientemente mediante RNA-Seq (Smircich *et al.*, 2015; Houston-Ludlam *et al.*, 2016; Li *et al.*, 2016; Berná *et al.*, 2017). En todos los casos, el ARN analizado es extraído a partir de extractos de células completas y por lo tanto se incluyen en los análisis ARNs nucleares que probablemente lleven a la identificación errónea de ARNs que realmente no se expresan. Una forma habitual de sortear este problema se basa en la selección de ARNs poliadenilados para la secuenciación. Sin embargo, la existencia de diferentes proporciones de ARN poliadenilado en el núcleo ha conducido a la identificación de falsos positivos (Trask *et al.*, 2009). La retención nuclear de ARNm procesado en mamíferos podría ser un mecanismo para disminuir el ruido debido a eventos de transcripción no regulada (Bahar Halpern *et al.*, 2015). Estas situaciones están particularmente representadas en *T. cruzi* debido, en primer lugar, la ausencia de regulación del inicio de la transcripción (De Gaudenzi *et al.*, 2011) y, en segundo término, a la existencia de un proceso generalizado de maduración de transcritos (Smircich *et al.*, 2015). Existen varios trabajos publicados que demuestran que la compartimentación núcleo citoplasma es un mecanismo funcional en mamíferos y la compartimentación diferencial de ARNs pequeños ha sido demostrada (Liao *et al.*, 2010). A su vez, un estudio en el que se compara el ARN total con el citoplasmático de tres líneas celulares humanas, y se infiere la contribución nuclear a partir del mismo, demuestra que los mensajeros retenidos a nivel nuclear son más largos y poseen más sitios blancos de miRNAs (micro ARNs) que los enriquecidos en citoplasma y los no diferenciales. Los autores proponen que la estructura de los

mensajeros, así como la regulación por miRNAs puede ser responsable de la compartimentación diferencial (Solnestam *et al.*, 2012) sin embargo, los mensajeros citoplasmáticos poseen a su vez, UTRs más cortas, por lo que la regulación por miRNAs puede no ser la causa del fenómeno observado, sino simplemente una consecuencia del largo del mensajero.

Varios trabajos aportan datos relevantes acerca del rol de la compartimentación nuclear en la regulación de los niveles de estado estacionario en los tripanosomátidos. A su vez, ha sido demostrado que el nucléolo, además de su función en la biosíntesis de ribosoma, podría estar involucrado en la regulación del metabolismo del ARNm y funcionar como sensor y coordinador de la respuesta al estrés (Boulon *et al.*, 2010). En efecto, tanto proteínas de unión al ARN como el ARN poliadenilado se acumulan en el nucléolo ante diferentes situaciones de estrés (actinomicina D, calor). La respuesta a estrés es observada en *T. cruzi* y *Leishmania* pero no en *T. brucei*, sugiriendo que el nucléolo tiene roles regulatorios diferentes en estos organismos (Nazer y Sanchez, 2011; Nazer *et al.*, 2011).

Por otro lado, la degradación co-transcripcional de mensajeros juega un rol importante en la determinación de los niveles de mensajeros en estado estacionario, sugiriendo un rol importante del *trans-splicing* y la poliadenilación de los mensajeros en la estabilización de los mismos (Fadda *et al.*, 2014). En este sentido, Kramer *et al.* demostraron que varias de las proteínas del exosoma se localizan en el núcleo y que los ARNs policistrónicos son blancos para la degradación por exosoma, lo que sugiere un rol nuclear en el control de calidad de los ARNs mensajeros, ARNs ribosomales y ARNs nucleolares pequeños (Kramer *et al.*, 2016). A su vez, ha sido demostrado en *T. brucei* que los niveles de ARN son determinados principalmente por el largo y su tasa de decaimiento en las formas procíclicas del parásito, sin embargo en el estadio sanguíneo los eventos nucleares juegan un rol importante en la regulación del nivel de mensajero, lo que sugiere que estos mecanismos habrían sido adquiridos durante la adaptación al parasitismo en mamíferos (Antwi *et al.*, 2016).

El rol de la compartimentación nuclear en la regulación de la expresión génica en eucariotas es de gran relevancia (Kohler y Hurt, 2007; Palazzo y Akef, 2012), no hay hasta

el momento, estudios globales del impacto de la compartimentación nuclear en tripanosomátidos.

2. Hipótesis de trabajo y Objetivos

2.1. Hipótesis

Como hipótesis de trabajo nos planteamos que la compartimentación de los ARNs a nivel nuclear interviene en la regulación de la expresión génica de *T. cruzi*.

2.2. Objetivo general

Contribuir al conocimiento de los mecanismos de regulación de la expresión génica en *T. cruzi* mediante el estudio de la compartimentación nuclear de transcritos.

2.3. Objetivos específicos

1. Identificación de los transcritos que posean distribución diferencial entre el núcleo y el citoplasma en epimastigotas de *T. cruzi*
2. Búsqueda de señales conservadas distintivas en los transcritos de los grupos con localización diferencial
3. Estudio comparativo de transcritos poliadenilados y transcriptomas de núcleo, citoplasma y extractos totales de epimastigotas de *T. cruzi*
4. Estudio del *cis-splicing* del transcritos de la poliA polimerasa (PAP) en *T. cruzi*

3. Estrategia experimental

La regulación postranscripcional de la expresión génica, juega un rol fundamental en los tripanosomátidos. En este trabajo nos propusimos estudiar el rol de la compartimentación núcleo-citoplasma en la definición de la naturaleza y nivel de los transcritos maduros en el citoplasma. Entendemos que la dilucidación de este fenómeno tiene relevancia no solo básica sino también sanitaria, ya que podría estar revelando un blanco adecuado para el desarrollo de drogas específicas.

Para ello, definimos la naturaleza de los transcritos en el núcleo y citoplasma mediante una aproximación ómica, basándonos en el fraccionamiento subcelular y la obtención de ARN de ambos compartimientos de epimastigotas de *T. cruzi* (figura 6). Esto requirió el aislamiento de fracciones enriquecidas en núcleos y citoplasma de epimastigotas en alto grado de pureza. Para realizar los fraccionamientos utilizamos la cepa Dm28c con la cual hemos trabajado en el laboratorio desde hace más de quince años.

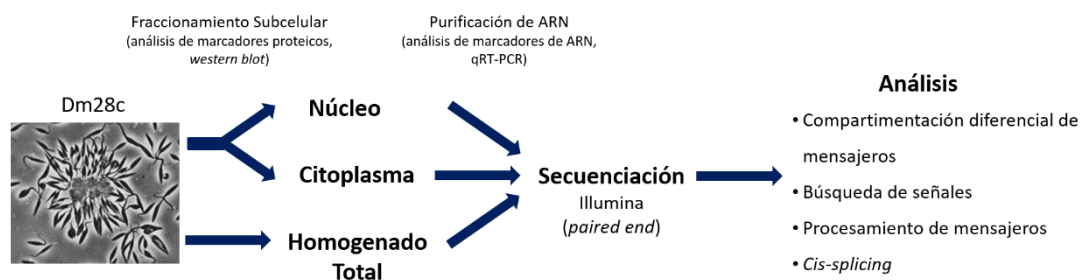


Figura 6. Estrategia de trabajo.

Realizamos el secuenciado masivo de los ARNs extraídos de las fracciones nucleares y citoplásmicas de epimastigotas de *T. cruzi*. Debido a los altos costos de secuenciación masiva, recurrimos a la combinación de cantidades equimolares de ARNs de réplicas independientes para dar lugar a una muestra única a secuenciar. Esta estrategia ha sido utilizada previamente con éxito (Allison *et al.*, 2006; Walrad *et al.*, 2012; Choi *et al.*, 2014; Araújo *et al.*, 2016). Mediante un análisis comparativo de las regiones codificantes establecimos la relación de contenido núcleo-citoplasma de los mensajeros. Los

resultados obtenidos fueron verificados en las réplicas independientes mediante qRT-PCR y/o análisis de FISH. Para comprender las diferencias de estos grupos, categorizamos los genes por ontología y analizamos las características estructurales de estos transcritos.

Analizamos también la presencia de señales conservadas distintivas en los transcritos de los grupos diferenciales.

A partir de los datos de secuenciación masiva de transcritos poliadenilados publicados por Smircich *et al.* (Smircich *et al.*, 2015) y de los obtenidos en este trabajo con ARN total realizamos un análisis comparativo para establecer grupos de transcritos con procesamiento distintivo, ya sea a nivel de maduración, exportación o estabilidad. En el grupo de genes caracterizados por una marcada regulación de la expresión por estabilidad, realizamos también un estudio comparativo con los datos masivos de estabilidad de mensajeros publicados para *T. brucei* (Fadda *et al.*, 2014).

Por último, los datos de los transcriptomas nuclear y citoplasmático fueron utilizados para evaluar el proceso de *cis-splicing* en epimastigotas de *T. cruzi* mediante el análisis de uno de los pocos genes reportados que sufren este procesamiento.

4. Transcritos diferencialmente localizados en núcleo y citoplasma en epimastigotas de *T. cruzi*

Artículo publicado en la revista *Frontiers in Cell and Developmental Biology – Cellular Biochemistry*, DOI: 10.3389/fcell.2017.00008.

En este trabajo, planteamos estudiar si el fenómeno de la compartimentación de ARN entre núcleo y citoplasma es relevante en la regulación de la expresión génica en epimastigotas de *T. cruzi*. Para cumplir con este objetivo, debíamos obtener fracciones enriquecidas en núcleo y citoplasma. En primer término, debíamos realizar una lisis adecuada de los parásitos, que conservara la estructura nuclear. A su vez, la mitocondria del parásito, que ocupa gran parte del citoplasma celular, no debía contaminar la fracción nuclear. Como estrategia, decidimos realizar el protocolo publicado por Rout basado en centrifugación diferencial en gradientes discontinuos de sacarosa (Rout y Field, 2001). Este protocolo implica la lisis de los parásitos en una solución que contiene PVP 8% (polivinilpirrolidona). Las soluciones con esta concentración de PVP son sumamente viscosas y debido a que no contábamos con un sonicador, la lisis de los parásitos y su disgregación no funcionaron adecuadamente. Probamos entonces, otros métodos de lisis celular, uno basado en lisis mediante congelado y descongelado de los parásitos (Gomez *et al.*, 1989) y otro basado lisis hipotónica en presencia de detergentes y homogenización (Roberts *et al.*, 1998). El último resultó más adecuado para nuestro objetivo, ya que lográbamos un mayor porcentaje de lisis de los parásitos (aproximadamente el 90%) mientras que para el protocolo de Gómez *et al.* obteníamos aproximadamente el 70% de los parásitos lisados. Finalmente, realizamos el método de lisis publicado por Roberts *et al.*, 1998 pero modificamos la velocidad de centrifugación a 500g en lugar de 1100g ya que, al realizar este paso tal como lo propone Gómez *et al.*, la contaminación de la fracción nuclear con kinetoplastos disminuyó hasta un 90%. El protocolo de Roberts implica un paso de centrifugación de la muestra resuspendida en un *buffer* conteniendo 250 mM de sacarosa, sobre un colchón de sacarosa 350 mM. Los núcleos de *T. cruzi* tienen una densidad mayor que los kinetoplastos quedando en el *pellet*, mientras que los kinetoplastos se localizan en el sobrenadante. Agregando un lavado del *pellet* que contiene los núcleos en *buffer* con 350 mM de sacarosa, obtuvimos una fracción nuclear con pureza y en cantidades suficientes para nuestro objetivo (figura 7) y una fracción citoplasmática que contiene básicamente el resto del parásito. En las

figuras 1 y 2 del artículo publicado, que se adjunta en este capítulo, se muestra el estudio de marcadores proteicos y de ARN realizados para comprobar la naturaleza de las fracciones obtenidas. Estos análisis demuestran que el fraccionamiento subcelular realizado fue adecuado y que las fracciones presentan una pureza adecuada para su utilización para secuenciación masiva de ARN.

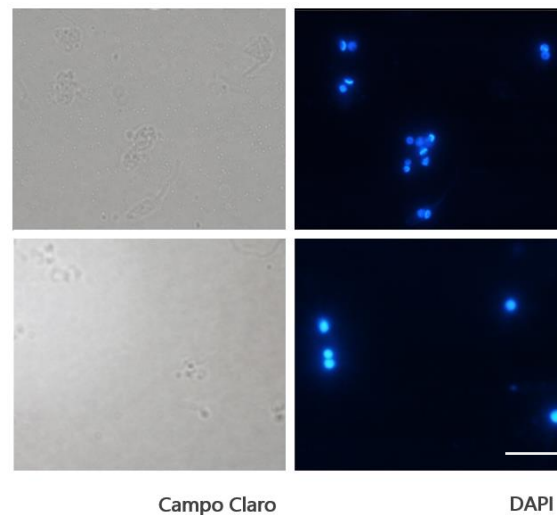


Figura 7. Microscopia de fluorescencia de la fracción nuclear de epimastigotas de *T. cruzi*. Arriba, parásitos enteros; abajo, fracción enriquecida en núcleos obtenida a partir de extractos de epimastigotas de *T. cruzi*. En base a modificaciones del protocolo de (Roberts *et al.*, 1998). A la izquierda se muestran las imágenes en Campo Claro y a la derecha las muestras teñidas con DAPI. La barra indica 10 μ m.

Resumiendo, pudimos establecer un protocolo que usa el paso de lisis hipotónica y homogenización en presencia de detergente (Roberts *et al.*, 1998) y las velocidades de centrifugación publicadas por (Gomez *et al.*, 1989) que nos permitió obtener fracciones nuclear y citoplasmática con pureza y cantidad adecuadas para realizar los estudios mediante RNA-Seq propuestos. A partir de las fracciones obtenidas realizamos extracciones de ARN siguiendo el protocolo de purificación de ARN total de mirVana (Ambion) que permite recuperar también los ARNs pequeños. Posteriormente, combinamos las fracciones nucleares, citoplasmáticas y totales de cuatro experimentos independientes, y procedimos a la secuenciación de las muestras combinadas.

El análisis comparativo de los transcriptomas muestra que el transcriptoma citoplásmico no se diferencia significativamente del transcriptoma total (figura 3, panel superior), y, por lo tanto, el fraccionamiento sub-celular no mejora la representación del proteoma celular en el transcriptoma. Sin embargo, sí existen transcritos diferencialmente localizados entre el núcleo y el citoplasma (figura 3, panel inferior), los transcritos enriquecidos en el núcleo están vinculados al proceso de infección, mientras que los enriquecidos en citoplasma están involucrados en el proceso de traducción y en el metabolismo celular (figura 7). La distribución diferencial de los transcritos fue validada por qRT-PCR y FISH (figuras 4 y 9). A su vez, los transcritos diferencialmente localizados son, estructural y composicionalmente diferentes (figura 5), lo que podría justificar en parte su compartimentación.

Los transcritos enriquecidos en el citoplasma son altamente traducidos en el estadio epimastigota de *T. cruzi* (figura 6), mientras que los enriquecidos en el núcleo lo son pobremente, pero aumentan su traducibilidad en el estadio metacíclico (figura 8).

Los resultados obtenidos demuestran que la compartimentación nuclear juega un rol importante en la regulación de la expresión génica en epimastigotas de *T. cruzi*.

Mi contribución en este artículo fue, la obtención de las muestras para su secuenciación, el análisis de los datos obtenidos y la identificación de los genes diferencialmente localizados, su categorización y la validación de los resultados obtenidos mediante qRT-PCR. Además, realicé los análisis composicionales y estructurales de los transcritos (el análisis de GC3 fue realizado por el Dr. Pablo Smircich y el de energía libre fue realizado por la Mag. Lorena Becco). Hice los ensayos de FISH y la microscopía confocal (junto con el Lic. Andrés Di Paolo). Participé en la escritura del artículo y en el análisis y la discusión de los resultados.



Nuclear Compartmentalization Contributes to Stage-Specific Gene Expression Control in *Trypanosoma cruzi*

Lucía Pastro^{1,2}, Pablo Smircich^{1,2}, Andrés Di Paolo³, Lorena Becco¹,
María A. Duhagon^{1,2}, José Sotelo-Silveira³ and Beatriz Garat^{1*}

¹ Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay,

² Departamento de Genética, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay, ³ Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

OPEN ACCESS

Edited by:

Catalina Dirney Alba Soto,
Instituto de Investigaciones en
Microbiología y Parasitología Médica,
UBA-CONICET, Argentina

Reviewed by:

Santuza Maria Ribeiro Teixeira,
Universidade Federal de Minas Gerais,

Brazil

Pedro José Alcolea,
CSIC, Spain

*Correspondence:

Beatriz Garat
bgarat@fcien.edu.uy

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 21 October 2016

Accepted: 25 January 2017

Published: 13 February 2017

Citation:

Pastro L, Smircich P, Di Paolo A,
Becco L, Duhagon MA,
Sotelo-Silveira J and Garat B (2017)
Nuclear Compartmentalization
Contributes to Stage-Specific Gene
Expression Control in *Trypanosoma
cruzi*. *Front. Cell Dev. Biol.* 5:8.
doi: 10.3389/fcell.2017.00008

In the protozoan parasite *Trypanosoma cruzi*, as in other trypanosomatids, transcription of protein coding genes occurs in a constitutive fashion, producing large polycistronic transcription units. These units are composed of non-functionally related genes which are pervasively processed to yield each mRNA. Therefore, post-transcriptional processes are crucial to regulate gene expression. Considering that nuclear compartmentalization could contribute to gene expression regulation, we comparatively studied the nuclear, cytoplasmic and whole cell transcriptomes of the non-infective epimastigote stage of *T. cruzi*, using RNA-Seq. We found that the cytoplasmic transcriptome tightly correlates with the whole cell transcriptome and both equally correlate with the proteome. Nonetheless, 1,200 transcripts showed differential abundance between the nuclear and cytoplasmic fractions. For the genes with transcript content augmented in the nucleus, significant structural and compositional differences were found. The analysis of the reported epimastigote transcriptome and proteome, revealed scarce ribosome footprints and encoded proteins for them. Ontology analyses unveiled that many of these genes are distinctive of other parasite life-cycle stages. Finally, the relocalization of transcript abundance in the metacyclic trypomastigote infective stage was confirmed for specific genes. While gene expression is strongly dependent on transcript steady-state level, we here highlight the importance of the distribution of transcripts abundance between compartments in *T. cruzi*. Particularly, we show that nuclear compartmentation is playing an active role in the developmental stage determination preventing off-stage expression.

Keywords: gene expression regulation, RNA-Seq, subcellular fractionation, trypanosomes, differentiation, life cycle

INTRODUCTION

Chagas' disease is caused by the protozoan parasite *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae), which is transmitted to humans by diverse species of the Reduviidae bug family (Chagas, 1909). The infected insect releases the parasite metacyclic trypomastigote forms with its feces while feeding host blood, entering the host through the wound area. Within the host, the metacyclic trypomastigotes can invade nearby cells and therein differentiate into the

intracellular amastigote forms. The latter multiply intracellularly and differentiate into bloodstream trypomastigotes, which are released into the circulation and infect cells from a variety of tissues, differentiating inside them into amastigotes. The bloodstream trypomastigotes can be ingested by bugs and transform in the vector's midgut into the replicative epimastigote and afterward, in the hindgut, into the infective metacyclic trypomastigotes.

Currently, there are approximately 6 to 8 millions of people suffering this disease and several millions are at risk of infection (WHO, 2015). Vector-borne transmission occurs in Central and South America but, since the parasite can also be transmitted by contaminated food, from mother to child and through contaminated blood or organ donations, Chagas' disease has spread to other continents (WHO, 2015).

Trypanosomatids show exceptional mechanisms for the expression of protein coding genes as the constitutive and polycistronic production of primary transcripts and mRNA maturation through the coordinate mechanisms of *trans*-splicing and polyadenylation. Since the mature messengers from the polycistronic transcriptional units (PTUs), are present in different levels, it is considered that post-transcriptional mechanisms should be actively controlling the gene expression in these organisms (Haile and Papadopoulou, 2007; Kramer and Carrington, 2011). Indeed, in *T. cruzi*, transcriptome studies have proved the existence of differences in mRNA abundance between genes which are consecutive in polycistronic transcriptional units (PTUs) (Minning et al., 2009; Smircich et al., 2015). These differences have been mainly attributed to mRNA stability and several *cis*-acting motifs and *trans*-acting factors have been identified (De Gaudenzi et al., 2005; Araujo and Teixeira, 2011; Bayer-Santos et al., 2012; Guerra-Slompo et al., 2012; Li et al., 2012). Besides, other post-transcriptional mechanisms occurring at the cytoplasm, such as those affecting mRNA localization (Cassola et al., 2007; Holetz et al., 2007; Cassola, 2011; Kramer, 2014) and translatability (da Silva Augusto et al., 2015; Smircich et al., 2015) provide additional steps for the control of mRNA availability for protein production.

The role of the nuclear compartment in the control of the cytoplasmic steady-state transcript levels has been recognized as crucial in eukaryotic gene expression regulation (Kohler and Hurt, 2007; Palazzo and Akef, 2012) and is also being studied in trypanosomatids (Nazer et al., 2011, 2012; Fadda et al., 2014; Antwi et al., 2016; De Gaudenzi et al., 2016; Kramer et al., 2016), for a very recent review see (Clayton, 2016). Although specific molecular actors involved in the nucleus-cytoplasmic transport in *T. cruzi* are being unraveled (Cassola and Frasch, 2009; Serpeloni et al., 2011a,b; Camara Mde et al., 2013; Inoue et al., 2014), no global analysis of the impact of nuclear compartmentation has been performed yet.

Transcriptome approaches in *T. cruzi*, have focused on the parasite life cycle, firstly studied by microarray analysis (Minning et al., 2009) and more recently, by RNA-Seq (Smircich et al., 2015; Houston-Ludlam et al., 2016; Li et al., 2016). In these cases, as for the majority of the transcriptome data from other organisms, the analyzed RNA is extracted from the whole cell disregarding the nuclear contribution to the total RNA population. This premise

is supported by the selection of polyA tailed RNA. However, it has been claimed that eukaryotic gene expression analyses using whole cell lysates, inadvertently measure a substantial number of mRNAs that are restrained into the nucleus (Trask et al., 2009). More recently, this assertion has been endorsed by the demonstration of nuclear retention of spliced polyadenylated mRNA (Bahar Halpern et al., 2015).

Since the characterization of nuclear and cytoplasmic RNA may contribute to further understand trypanosomatid gene expression regulation, a comparative *in masse* analysis of nuclear and cytoplasmic RNA in *T. cruzi* epimastigotes was performed and the impact of nuclear-cytoplasmic RNA partitioning in whole cell RNA was evaluated. We found that the use of the cytoplasmic transcriptome does not significantly improve the estimation of protein abundance obtained from whole cell parasite transcriptome. However, we detected that up to 20% of the genes have differential transcript levels between nucleus and cytoplasmic compartments. Transcripts with higher level in the nucleus are significantly longer and have a higher GC content than the ones with higher content in the cytoplasm. Using data from the reported transcriptome and proteome for *T. cruzi* epimastigotes (de Godoy et al., 2012; Smircich et al., 2015), we observed that these transcripts are also characterized by scarce depiction of ribosome footprints and encoded proteins. In addition, we found that the transcripts enriched in the cytoplasmic fraction correspond to genes expressed in the epimastigote stage, while the ones enriched in the nucleus are distinctive of other life cycle stages. For selected specific transcripts, fluorescent *in situ* hybridization (FISH) was used to study the RNA subcellular localization in non-infective epimastigotes and in infective metacyclic trypomastigotes. Altogether, these results support an active role of nuclear compartmentalization in stage-specific gene regulation in *T. cruzi*.

MATERIALS AND METHODS

Parasites Culture

The *T. cruzi* Dm28c clone (Contreras et al., 1988) was used. Epimastigotes were maintained at 28°C in liver infusion tryptose (LIT) medium supplemented with 10% heat inactivated fetal bovine serum (FBS). Metacyclic trypomastigotes were prepared as previously (Duhagon et al., 2009). Briefly, epimastigotes at stationary phase were incubated in TAU medium at 28°C for 2 h (Contreras et al., 1985). The parasites were then washed twice with PBS and immediately used.

Parasite Fractionation

To obtain nucleus and cytoplasmic fractions we used a previous reported methodology (Gomez et al., 1989) with a few modifications. Briefly, approximately 5×10^9 parasites were used for each isolation procedure; the parasites were harvested at 1,700 g for 5 min. Cell pellets were resuspended in phosphate-buffered saline, washed twice, and resuspended in 3 volumes of ice-cold hypotonic buffer (10 mM HEPES, pH 7.9; 1.5 mM MgCl₂; 10 mM KCl; 0.5 mM DTT; 1 mg/mL pepstatin; 0.5 mg/mL leupeptin; 0.5 mM PMSF) and incubated on ice for

10 min. Nonidet P-40 was added to a final concentration of 0.2% and parasites were disrupted by 13 strokes in a 15-mL glass Potter-Elvehjem Dounce homogenizer (Wheaton). Then sucrose was added to a final concentration of 0.35 M. An aliquot was separated and named “Whole cell” (Wc). The lysate was centrifuged at 500 g for 15 min at 4°C. The supernatant was separated and named “Cytoplasm” (C). The pellet was resuspended in 5 volumes of sucrose buffer (0.35 M sucrose; 10 mM HEPES, pH 7.9; 3.3 mM MgCl₂; 10 mM KCl; 0.5 mM DTT; 1 mg/mL pepstatin; 0.5 mg/mL leupeptin; 0.5 mM PMSF) and centrifuged in a swinging bucket rotor at 1,100 g for 15 min at 4°C. The pellet was resuspended in the same sucrose solution and named “Nucleus” (N). Four independent replicates of the fractioning procedure were done.

Western Blot Assay

Each Wc, C, and N protein fractions (corresponding to 5×10^6 parasites per lane) were separated by electrophoresis in 10 or 12% SDS-PAGE and electro-transferred onto Hybond C Extra membranes (GE Healthcare) following standard procedures. Membranes were blocked by incubation in 5% skim milk powder in PBS-0.1% Tween and were then incubated for 1 h at room temperature with polyclonal, anti-TcTXN1 (Pineyro et al., 2011), anti-TcH2A, FioCruz, Brazil), anti-Tc38 polyclonal antibody (Duhagon et al., 2009), and anti-TcRBP40 (Guerra-Slompo et al., 2012), were used as purification controls. Bound antibodies were detected using an IRDye 800CW and 680CW Goat anti-Rabbit or anti-Mouse IgG (H + L) (Li-Cor), diluted 1:2,500 and analyzed in a G-Box (Syngene).

RNA Preparation and Sequencing

RNA was isolated from the Wc, C, and N fractions using MirVana kit (Ambion), according to Total RNA purification manufacturer's protocol. The obtained RNA was treated with DNase according to manufacturer's protocol (DNA-free, Ambion), quantified by Nanodrop (Thermo Scientific, USA) and Qubit 2.0 Fluorometer (Invitrogen) and its integrity was checked by Bioanalyzer (Agilent, USA). Genomic DNA control, in the cytoplasmic fraction, was done by qPCR against *gapdh* housekeeping gene. Then, RNA was treated with riboMinus (Invitrogen), the purified RNA was quantified by Nanodrop (Thermo Scientific, USA) and Qubit 2.0 Fluorometer (Invitrogen) and their integrity as well as rRNA depletion were analyzed by Bioanalyzer (Agilent, USA). Identical quantities of RNA derived from each of the four independent biological replicates were pooled to obtain the Whole cell, Nucleus and Cytoplasmic pooled samples. Pooled samples were sequenced by Illumina HiSeq2000 platform with a paired-end read running type and 100 bp cycle as a running condition at Macrogen (Korea).

Data Analysis

The obtained data were analyzed by CLC Genomics Workbench. In first place a 3' quality trimming was done to the three sequence sets. To determine RNA transcript levels, the *T. cruzi* CLBrener Esmeraldo-like, genome release 4.2 from the TriTrypDB were used for RNA-Seq analysis included in the package. The reads

per kilobase of transcript per million mapped reads (RPKM) were obtained for each gene in each condition. Only genes with 10 reads or more per transcript were used for quantification purposes.

Functional annotation of the NET and CET genes was performed using the Blast2GO tool (Conesa and Gotz, 2008) using a gene cut off of 10. Overrepresentation analyses were calculated by DAVID (Database for Annotation, Visualization and Integrated Discovery) (Dennis et al., 2003).

The intergenic RNA contribution for each fraction was calculated through the differences of reads that map to genome minus those that map to transcriptome (intergenic reads) as percentages.

For the fine location of the *trans*-splicing sites, we used SLAP mapper (Fiebig et al., 2014) *Trypanosoma cruzi* CLBrener Esmeraldo-like V6.0 was set as genome reference. Reads containing at least 8 nucleotides of the splice leader sequence were retrieved and the highest length upstream to the CDS or the one with more read counts was selected for each gene. Meanwhile, for the fine location of the polyadenylation sites, reads containing a track of at least 8 A were listed and the highest length downstream to the CDS or the one with more read counts was selected for each gene. To avoid inconsistencies of UTR delimitation due to partial maturation, only the data derived from the cytoplasmic transcriptome were used for these purposes.

GC-content (GC%) and GC-content in the third codon position (GC3) were calculated using in-house python scripts.

The RNAfold Vienna RNA Package 2.0 algorithm was used to calculate the minimum free energy for the 5' and 3' UTR of transcripts. Default parameters were used and the temperature was set at 37°C (Lorenz et al., 2011).

Quantitative RT-PCR

cDNA was synthesized from 1 µg of DNase treated RNA (DNA-free, Ambion) using Superscript III kit first strand synthesis (Invitrogen) and random hexamer primers. Quantification of specific products was done by qRT-PCR, using QuantiTect SYBR Green PCR Master Mix (Qiagen). Double stranded products were amplified using specific primers (Table S1) in a real time rotary analyzer RotorGene 6000 (Corbett). Relative amounts of target genes were calculated by normalization with the *gapdh* housekeeping gene that displayed similar values of absolute Ct in the quadruplicates: Wc (16.1 ± 0.2), N (16.1 ± 0.3) and C (16.5 ± 0.1) and of rpkm in the respective transcriptomes (216.5 for Wc, 263.6 for N and 266.5 for C). PCR reaction mixture containing 0.9 µM of each primer was carried out in a final volume of 10 µL. RNA levels were compared using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Fluorescence *In situ* Hybridization (FISH)

FISH analysis were done as previously described (Garcia-Silva et al., 2010) with some modifications. Briefly, epimastigotes were cultured, allowed to adhere to polylysine-coated microscope slides for 20 min at room temperature and, after washing twice in PBS, parasites were fixed with 4% paraformaldehyde in PBS for 10 min at room temperature, washed twice with PBS and further incubated in 25 mM NH₄Cl for 10 min. Parasites were

permeabilized with 0.2% Triton X-100 in PBS for 5 min. Slides were then blocked and prehybridized for 2 h at room temperature in bovine serum albumin 2%, $5 \times$ Denhardt, $4 \times$ SSC and 35% deionized formamide (hybridization solution). Assays were performed under denaturing conditions by heating slides at 75°C for 3 min just prior to the hybridization step (see Table S1 for used probes). Hybridization was performed overnight at 45°C in a humid chamber in the presence of 1 ng/ml of the indicated oligonucleotide conjugated either to fluorescein amidite (6-FAM) or to the cyanine 3 dye (Cy3) (Table S1). After hybridization, the slides were washed once in $2 \times$ SSC plus 50% deionized formamide, once in $2 \times$ SSC, once in $1 \times$ SSC, and twice in $0.5 \times$ SSC. Slides were mounted with Prolong (Molecular Probes) containing DAPI. Confocal images were acquired at room temperature using an OLYMPUS FV 300 with lasers Ar 488 and He-Ne 633 (Melles Griot). Merged images were obtained by superimposing the indicated images files in ImageJ software (Schneider et al., 2012).

Availability of Supporting Data

The data sets supporting the results of this article are available in the Sequence Read Archive repository, BioProject ID: PRJNA342509.

RESULTS

Transcriptomics of Nuclear and Cytoplasmic Fractions of *T. cruzi* Epimastigotes

In order to study the distinctive RNA contribution of the nuclear and cytoplasmic compartments to the total transcriptome of *T. cruzi* epimastigotes, a subcellular fractionation was performed. Proteins known to be specifically localized in: nucleus (histone TcH2A), cytoplasm (tryparedoxin TcTXN1), mitochondria (Tc38) and reservosomes (TcRBP40) were used in western blots to evaluate the purity of the fractions (Duhagon et al., 2009; Pineyro et al., 2011; Guerra-Slompo et al., 2012). The latter two markers were used to evaluate the frequent mitochondrial and reservosomal contamination of the nuclear fraction in this organism. As shown in **Figure 1**, the four independent replicates displayed very similar patterns. All the markers were present in the whole cell extract (Wc fraction), whereas the nuclear (N) fraction exhibited a conspicuous signal for the histone protein TcH2A and a faint to not detectable signals for the other three markers. On the contrary, the cytoplasmic (C) fraction showed strong bands for TcTXN1, Tc38, and TcRBP40, and a faint band for the nuclear marker.

Total RNA from each isolated fraction was extracted and qualitative and quantitatively analyzed (Figure S1). The four biological replicates yielded 0.6 ± 0.1 pg (mean \pm S.E.) of total RNA per epimastigote cell. The nuclear RNA contribution represented $10 \pm 1\%$ of the RNA in the cytoplasm fraction per epimastigote cell. The differential distribution of well-established nuclear confined RNAs, such as unprocessed rRNA and snoRNA, was analyzed by quantitative RT-PCR (qRT-PCR). A remarkable enrichment of snoRNA and intergenic rRNA in the nuclear fractions was revealed (**Figure 2**). The small standard error

among the four independent experiments argues in favor of the similarity of the quadruplicates (also supported by additional analyses presented in the following section). Besides, no genomic DNA could be detected by qPCR in the non-retrotranscribed RNA from the cytoplasmic fraction. Overall, the protein and RNA markers profile supported the adequacy of the nucleus and cytoplasmic fractions from the four independent replicates for a further RNA-Seq analysis.

Given the reproducibility of the replicates, the RNA extractions of the four biological replicates of each fraction were then pooled and the electrophoretic profile was analyzed. As observed for each individual experiment, the whole cell and the cytoplasmic pooled fractions displayed similar profiles whereas the nuclear fraction was more dissimilar (Figure S2). The pooled fractions were then sequenced as indicated in Materials and Methods section and the raw statistic data are shown in Table S2. RNA-Seq analysis using the *T. cruzi* CLBrener transcriptome yielded roughly 17 million mapped reads for Wc, N, and C. When the reads were mapped to the genomic sequences, we observed a higher proportion of mapped reads in the nuclear than in cytoplasmic fraction. This result was expected because of the presence of intergenic regions in the polycistronic RNA which are lost during maturation in the nucleus previous to the export to cytoplasm (Clayton, 2016). When the reads were mapped to the transcriptome sequences (see Data analysis in Materials and Methods section), we found that most of the 10,600 annotated transcripts present at least one mapped read: 9,079 in the Wc, 9,225 in the N and 9,011 in the C RNA fractions.

Transcripts Showing Differential Distribution between the Nuclear and Cytoplasmic Fractions Are Found in *T. cruzi* Epimastigotes

A great proportion of the annotated transcripts were present in the cytoplasmic and nuclear transcriptomes, indicating that the majority of the transcripts are exported to the cytoplasm. This finding is in agreement with the recent hypothesis of pervasive RNA maturation in *T. cruzi* (Smircich et al., 2015). However, since there might be differences between transcript abundance in the cell compartments, we quantitatively compared the transcriptomes. We found a strong correlation between C and Wc transcriptomes (Pearson correlation coefficient $r = 0.98$ $p < 0.0001$, **Figure 3** upper panel). This indicates that the Wc transcriptome adequately represents the transcripts present in C, and therefore it is appropriate to describe the global mRNA steady-state level in this organism using whole cell transcriptome approaches. On the other hand, the nuclear transcriptome also correlated with the whole cell transcriptome, although with a smaller correlation coefficient (Pearson correlation coefficient $r = 0.43$, **Figure 3** middle panel). Similarly, when we analyzed the correlation between the N and C transcriptomes we determined a Pearson correlation coefficient $r = 0.53$ (**Figure 3** lower panel). These findings pointed us to the existence of transcript abundance differences between compartments. Consistent with these results on transcriptomes' correlations, similar correlations for Wc and C transcriptome with the epimastigote proteome (de Godoy et al., 2012) were deduced. As expected, a minor

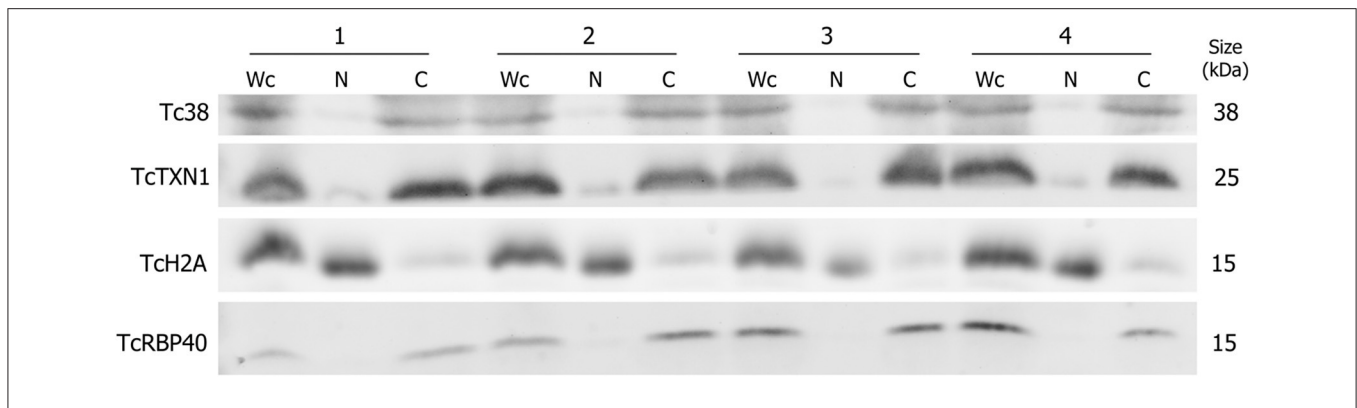


FIGURE 1 | Protein control of *T. cruzi* epimastigote subcellular fractions. Western blot analysis of whole cell (Wc), nucleus (N) and cytoplasm (C) fractions using specific antibodies for selected protein markers: Tc38 (mitochondria), TcTXN1 (cytoplasm), TcH2A (nucleus), TcRBP40 (reservosome) are shown for four independent biological replicates (1 to 4). The molecular weights (kDa) of the proteins are indicated on the right.

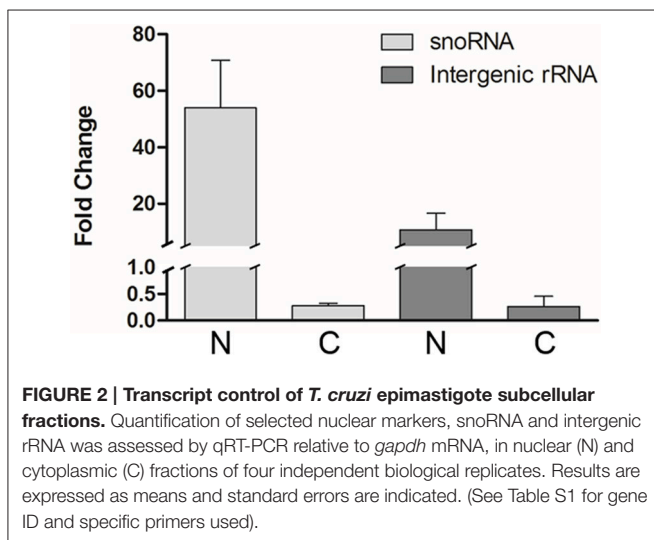


FIGURE 2 | Transcript control of *T. cruzi* epimastigote subcellular fractions. Quantification of selected nuclear markers, snoRNA and intergenic rRNA was assessed by qRT-PCR relative to *gaph* mRNA, in nuclear (N) and cytoplasmic (C) fractions of four independent biological replicates. Results are expressed as means and standard errors are indicated. (See Table S1 for gene ID and specific primers used).

correlation coefficient with proteome was obtained for the nucleus transcriptome (Table S3).

In order to identify the sets of genes with differential transcript abundance between N and C, we compared their transcriptomes. A total of 1,182 genes (20% of the 6,039 genes with more than 10 reads in the Wc transcriptome) were found to have at least a 2-fold change between the two compartments ($FC \geq 2$, gray colored in Figure 3 lower panel and Table S4 for the list of gene IDs). Among them, 444 were genes with cytoplasmic enriched transcript abundance (CET, dark gray in Figure 3 lower panel) and 738 had nuclear enriched transcript abundance (NET, light gray in Figure 3 lower panel), while 4,857 were not differentially distributed (NDT, black in Figure 3 lower panel). These results indicate that even when the transcriptomes from Wc and C are highly correlated, a set of genes show significant differences of transcript abundance between the nuclear and cytoplasmic compartments in *T. cruzi* epimastigotes.

Robustness of the compartment distribution differences found by RNA-Seq was confirmed for eight genes by qRT-PCR in the

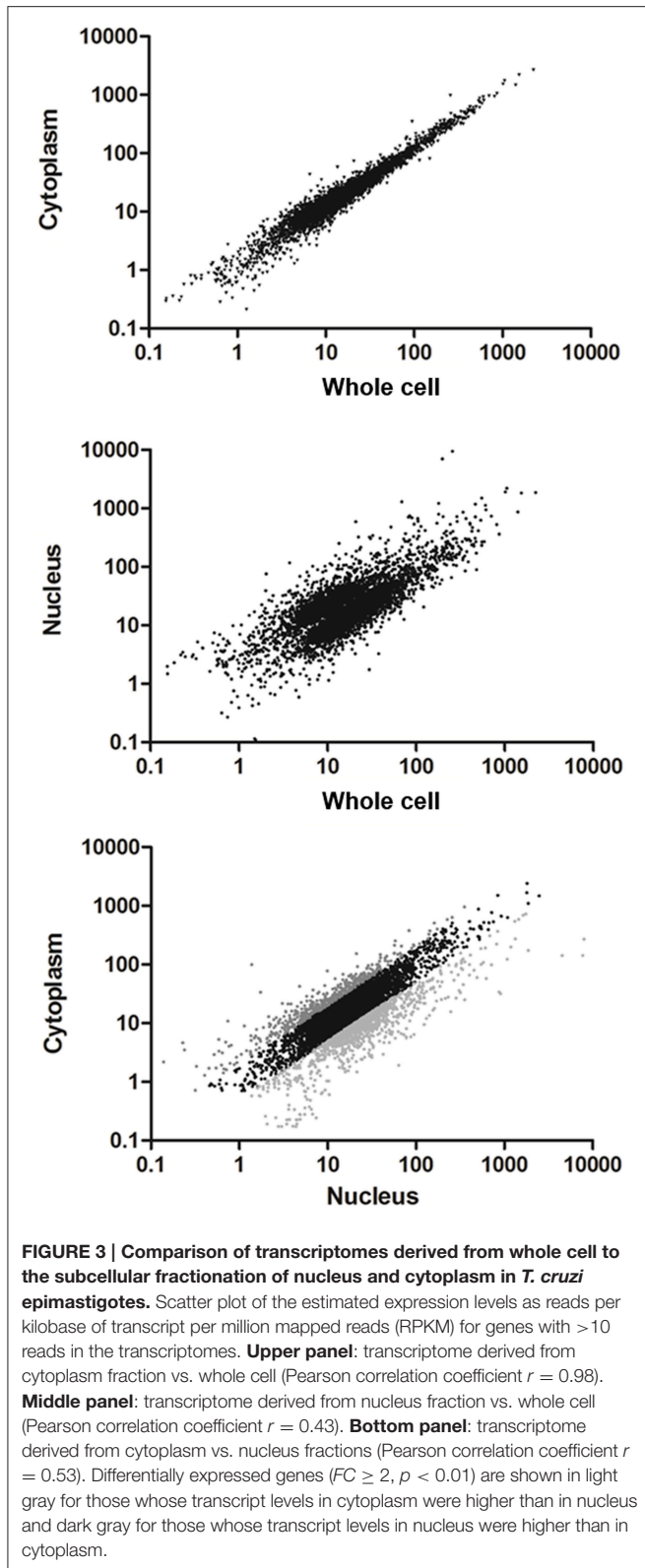
four independent biological replicates (Figure 4). Furthermore, the high correlation of transcript quantification by qRT-PCR and by RNA-Seq (Pearson $r = 0.86$) strongly support the RNA-Seq results (Figure S3). These data together with the RNA markers (Figure 2) were used to compare the individual samples within each cellular fraction. The quadruplicates showed average correlations of $r \geq 0.96$ with standard deviation ≤ 0.02 (Figure S4A). In addition, we found that the largest component of the variance of the samples, identified by PCA, comes from the identity of the cellular fraction (Figure S4B). These analyses validate the use of a pool of replicates of each subcellular fraction for the identification of differentially expressed genes.

The Genes with Enriched Transcript Levels in the Nucleus of *T. cruzi* Epimastigotes Have Distinctive Compositional and Structural Characteristics

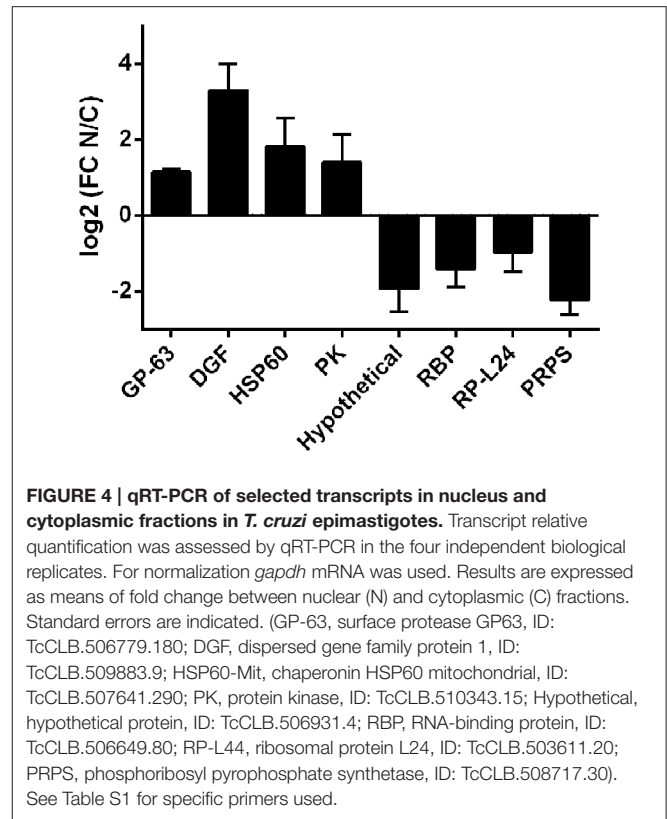
We then investigated the association of properties such as mRNA length and base composition with the nuclear-cytoplasmic differential transcript distribution. Firstly, we had to identify the 5' and 3' ends for the UTRs, since in *T. cruzi* only CDS are annotated. We sought out the UTRs boundaries using the strategy outlined in Materials and Methods section which is based on the detection of at least 8 nucleotides of the splice leader for the 5' UTRs and a track of at least 8 A for the 3' UTRs. For the 1,182 genes exhibiting differentially distributed transcripts, the 5' UTR and 3' UTR could be assigned to 505 and 162 genes respectively. Nonetheless, complete transcripts could only be determined for 81 genes.

The analysis of transcript length revealed that genes whose transcripts were enriched at the cytoplasm were significantly shorter than the non-differentially distributed transcripts (NDT) ($p < 0.0001$) while those whose transcripts were enriched at the nucleus were slightly longer than the ones in the cytoplasm ($p < 0.0001$) (Figure 5A). Although this difference mostly relied on CDS size a similar trend was observed for both UTRs.

Concerning the RNA composition, we observed that genes whose transcripts were enriched in the nucleus had a significantly



higher GC% content than NDT genes (Figure 5B). This difference was seen in the GC content at CDS and a similar trend was observed at the UTRs. In addition, the CET GC content at

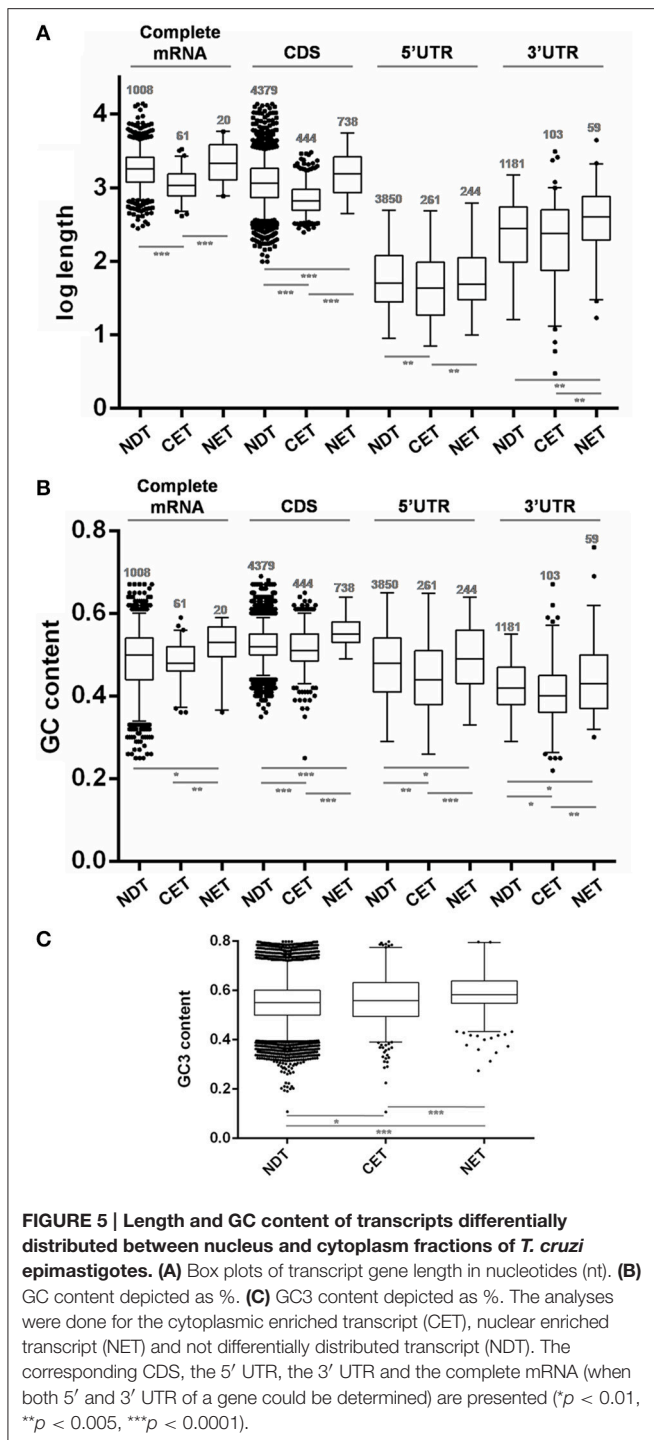


the CDS and also at the determined UTRs was significantly lower than the NET and NDT. For the complete mRNA sequences, a similar tendency, with restricted significances probably due to the low number of determined sequences, was observed. A good agreement between the CDS GC content and GC3 usage was also found (Figure 5C). In addition, since the GC content could affect the RNA conformation, we also investigated the existence of significant thermodynamic differences at the determined UTRs of these sets of transcripts. A significant difference in the free energy was only found for the 5' UTRs of the NET genes when compared to the NDT genes ($p < 0.01$). The lower free energy of the 5' UTRs of the NET genes may be predicting the presence of more stable structures at these regions (Figure S4).

These findings support that gene compositional and structural characteristics may underlay the differential transcript partitioning between nucleus and cytoplasm.

The Genes with Enriched Transcript Levels in the Nucleus of *T. cruzi* Epimastigotes Are Poorly Translated

Taking advantage of the availability of data derived from ribosome profiling of *T. cruzi* epimastigote (Smircich et al., 2015), we determined the ribosome occupancy of the nucleus-cytoplasm differentially distributed transcripts. A high number of ribosome footprints were observed for CET, while NET had significantly fewer ribosome footprints than either CET or NDT (Figure 6A). Using the quantitative proteomic data for *T. cruzi* epimastigotes (de Godoy et al., 2012) a similar result



was obtained (Figure 6B). Interestingly, virtually all the CET genes were present in the proteome (99%, 440 out of the 444 - dark gray in Figure 3 lower panel), on the contrary, only a few NET genes are detected (6%, 44 out of the 738 - light gray in Figure 3 lower panel). In addition, while CET levels, measured in Wc transcriptome, significantly correlate with their proteomic expression, no correlation is seen for NET genes (Table S5). Consistently, the presence of transcripts of pseudogenes in the

NET set was higher than expected by chance (122 out of 1,302 annotated pseudogenes, 36% increase, Fisher exact test $p < 0.01$). Even more, pseudogenes were underrepresented in the CET genes (21 pseudogenes, 62% decrease, Fisher exact test $p < 0.01$).

Taking together, these findings further support nuclear partitioning impact on protein content modulating transcript availability for translation.

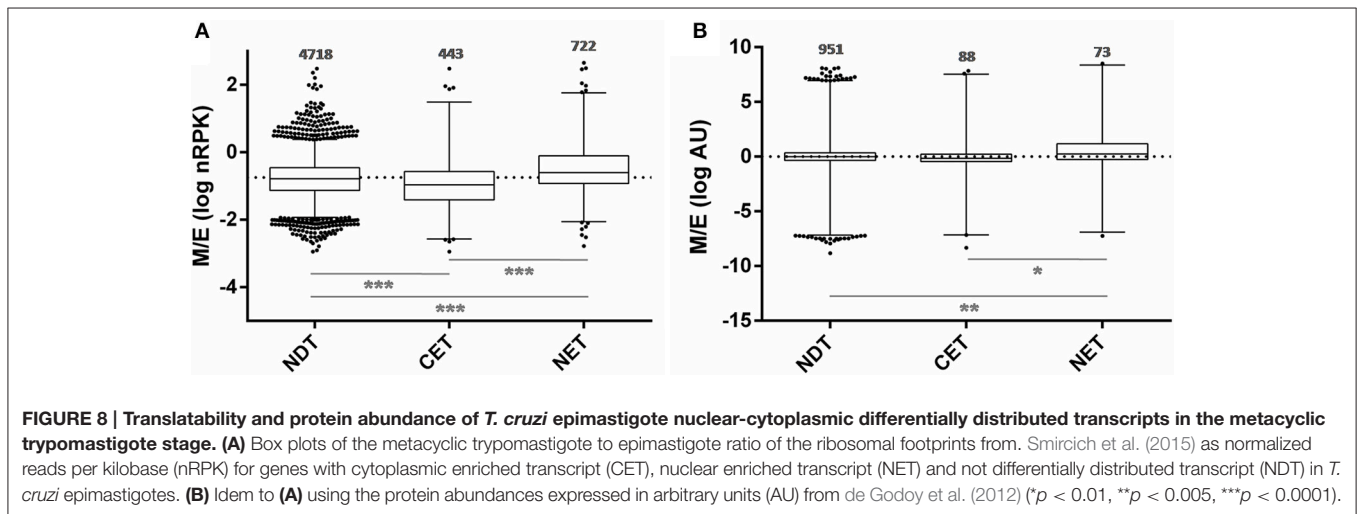
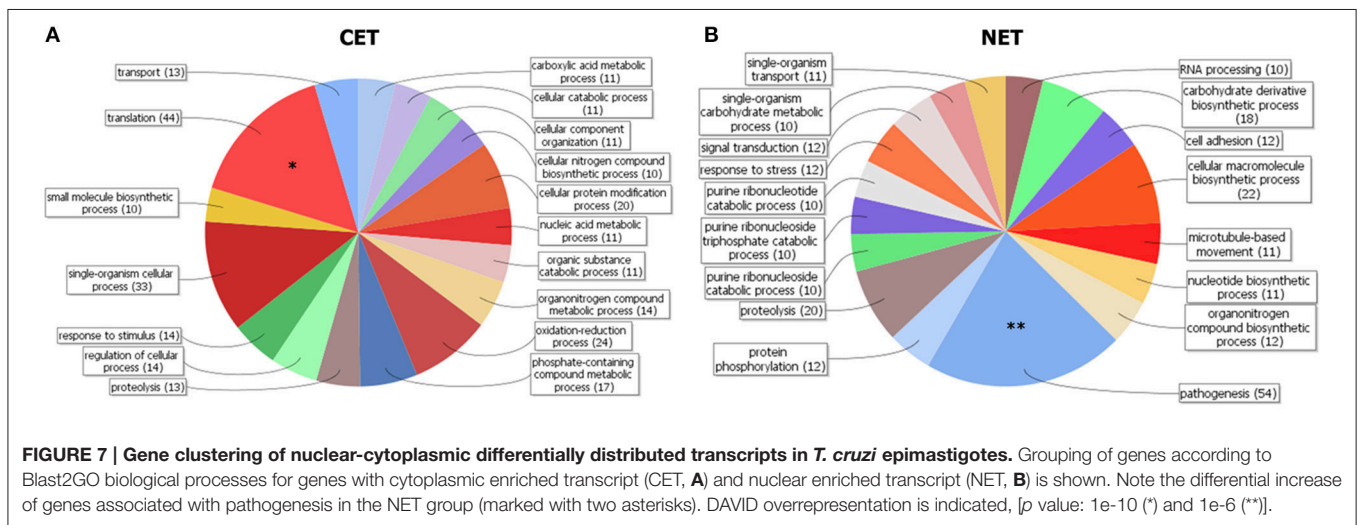
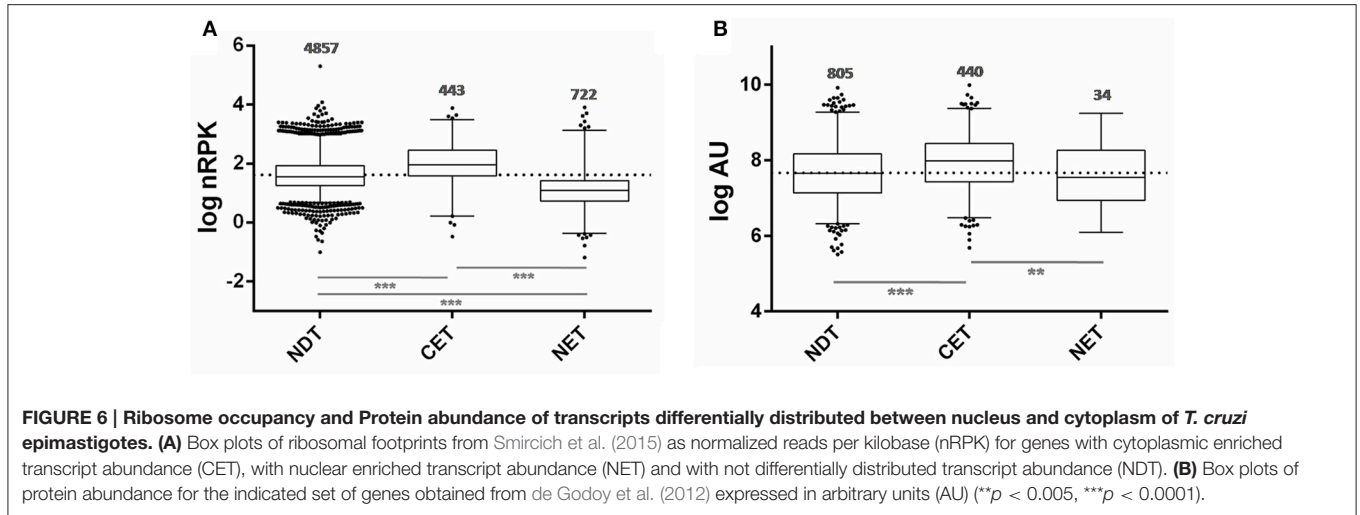
Off-Stage Transcripts Are Enriched in the Nucleus Compartment of the *T. cruzi* Epimastigote Stage

In order to study the biological characteristics of the NET and CET genes, we performed an analysis of overrepresentation of gene ontology terms (Figure 7). Using the Blast2GO tool (Conesa and Gotz, 2008) we found that the CET genes are involved in biological processes of the replicative and non-infective epimastigote stage such as translation and metabolic pathways (Figure 7A). Conversely, genes encoding factors with active roles in life cycle stages different from the epimastigote stage, such as pathogenesis, have higher transcript abundance in the nucleus (Figure 7B). Concurrently, a detailed inspection of the NET genes exposed an overrepresentation of *trans*-sialidase, DGE, RHS, GP63 and MASP gene family's members (χ^2 , $p < 0.001$) and an underrepresentation of ribosomal proteins. The former families of genes have been extensively studied and their expression has been related to the infective metacyclic stage of the parasite (Cuevas et al., 2003; Bartholomeu et al., 2009; Kawashita et al., 2009; De Pablos and Osuna, 2012; Grynberg et al., 2012). Indeed, we found that the ribosome footprints of the epimastigote NET genes increase in the metacyclic trypomastigote stage while those of the epimastigote CET genes are diminished (Figure 8A). A similar pattern was obtained when comparing the metacyclic trypomastigote and epimastigote proteomic data from (de Godoy et al., 2012) (Figure 8B).

These findings support an active role of the nuclear compartment limiting the off-stage translation of genes from stages different from the actual that could occur due to the constitutive and massive characteristics of transcription in this parasite.

Epimastigote Nuclear-Cytoplasmic Transcript Distribution Changes after the Transition to the Metacyclic Trypomastigote Stage

Taking into account all the above results, we hypothesized that transcript subcellular distribution is developmentally regulated. To test this hypothesis, we compared the localization of mRNAs belonging to CET (L44) and NET (GP63 and adenylate cyclase) in the non-infective epimastigote and the infective metacyclic trypomastigote, using fluorescent *in situ* hybridization (FISH). In contrast with the global mRNA localization, as seen by the detection of polyadenylated transcripts, these mRNAs showed developmental dependent localization (Figures 9, 10). The infection related gene, gp63, is faintly expressed in the epimastigote but increases its expression in the metacyclic trypomastigotes and amastigote stages (Cuevas et al., 2003). As



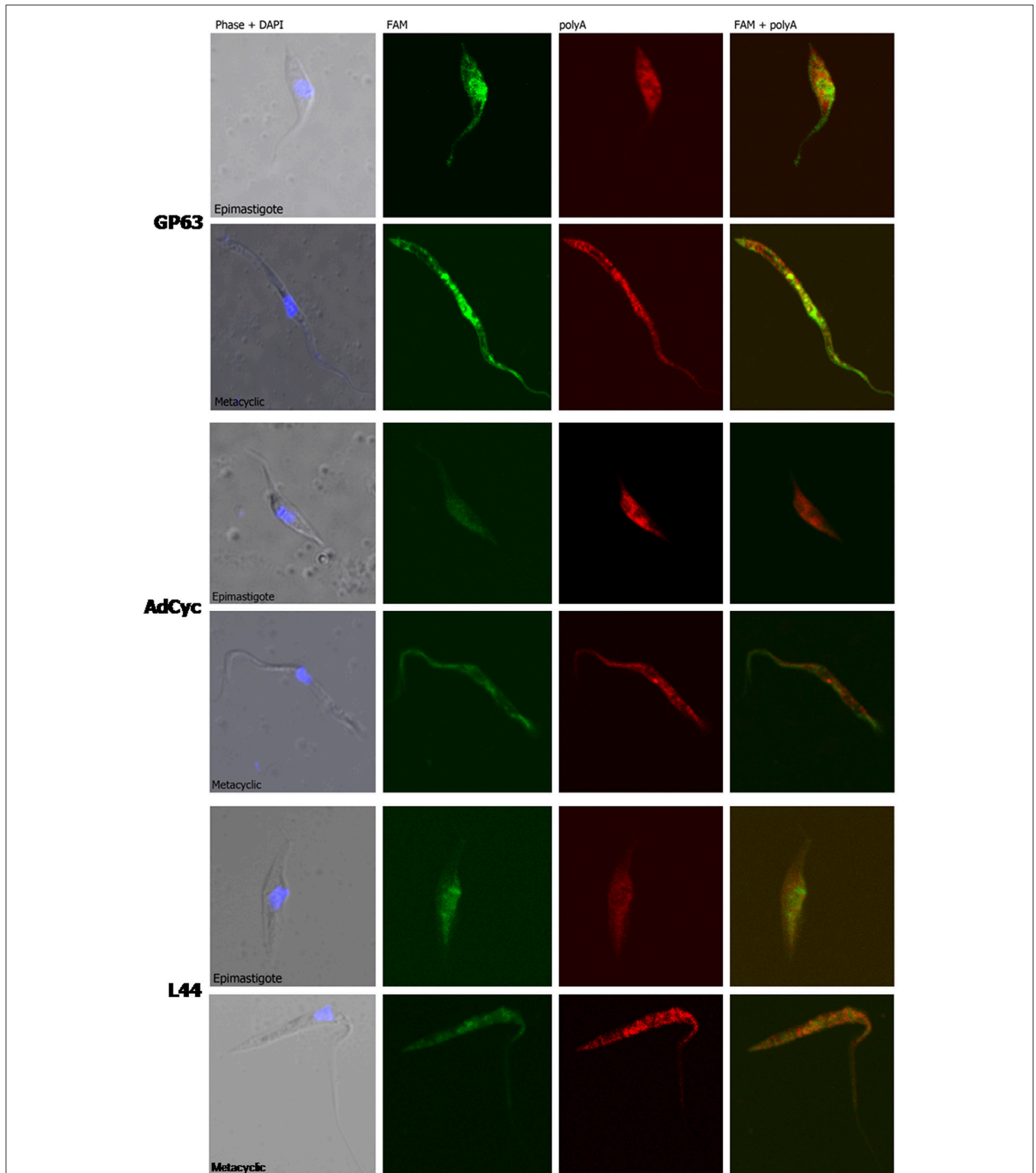
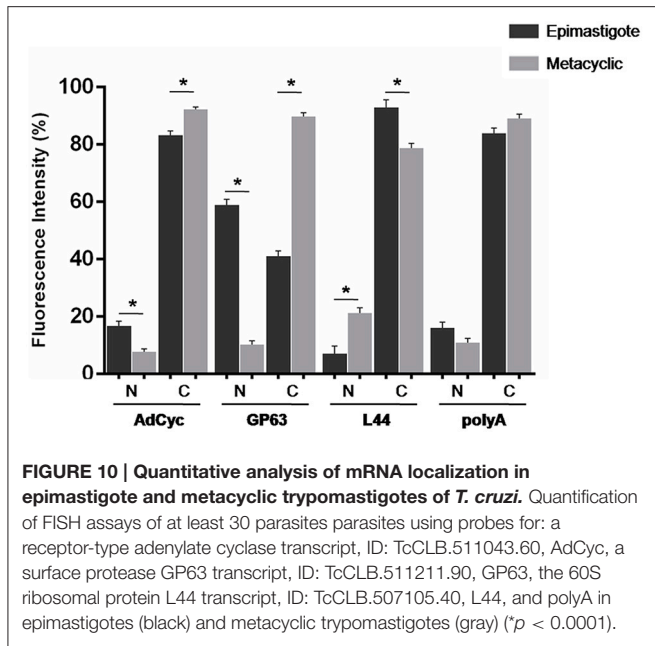


FIGURE 9 | mRNA localization in epimastigote and metacyclic trypomastigotes of *T. cruzi*. Confocal images of FISH assays of representative parasites using probes for: a surface protease GP63 transcript, ID: TcCLB.511211.90, GP63, upper panel. A receptor-type adenylate cyclase transcript, ID: TcCLB.511043.60, AdCyc, middle panel; and the 60S ribosomal protein L44 transcript, ID: TcCLB.507105.40, L44 bottom. At least 30 parasites were analyzed for each probe in each condition. Staining of epimastigotes (upper panels) and metacyclic trypomastigotes (bottom panels) are presented. These probes (See Table S2) were 5' labeled with 6-FAM (green). For the detection of polyA tailed transcripts a 5' Cy5 labeled polyT probe was used (red). DNA was visualized by DAPI staining (blue). Phase contrast images were obtained. Magnification 120x.



shown in **Figure 9**, the gp63 transcripts were mainly localized in the nucleus in the epimastigote stage, while they acquired a cytoplasmic localization in the metacyclic trypomastigote stage. In addition, the superposition of polyA and the specific gp63 transcripts suggests an active translation in the metacyclic trypomastigote. A similar behavior was displayed by the adenylate cyclase transcript (also belonging to the NET gene set), being clear a depletion of the nuclear signal in the metacyclic trypomastigotes. For, the ribosomal protein L44 transcripts (belonging to the CET genes) a decrease of the colocalization with the polyA probe was observed in the transition from the epimastigote to the metacyclic trypomastigote stage. The quantitative analysis of the images for at least 30 parasites for each probe at the epimastigote or metacyclic trypomastigote stages, confirmed the displayed pattern changes (**Figure 10**).

These results support the existence of changes in the subcellular localization of transcript abundances accompanying the metacyclogenesis developmental process in *T. cruzi*.

DISCUSSION

The steady-state level of proteins is the result of successive, alternative and combinatorial steps starting at transcription initiation. In *T. cruzi*, as well as in other trypanosomatids, the control of this step seems to be mostly absent. Therefore, trypanosomatids have been considered unique models for the study of post-transcriptional processes and their regulation. Indeed, processes like editing and *trans*-splicing were firstly described in trypanosomatids and later revealed to also occur, in other eukaryotes. However, while in eukaryotes the co-transcriptional RNA processing, surveillance and transport from nucleus to cytoplasm have been widely recognized as regulatory steps in gene expression (Kohler and Hurt, 2007), their survey in trypanosomes is very recent (Nazer et al., 2011, 2012; Fadda

et al., 2014; Antwi et al., 2016; De Gaudenzi et al., 2016; Kramer et al., 2016). Likewise, genome-wide approaches have been used to identify the RNA nature and abundance distribution in the nucleus and the cytoplasm in many eukaryotes (Liao et al., 2010; Solnestam et al., 2012; Tilgner et al., 2012; Zaghlool et al., 2013; Bai et al., 2014), but such analyses have not yet been reported in trypanosomatids.

Seeking to study the role of the nuclear compartment in gene expression control in trypanosomes, we undertook cellular fractionation to obtain RNA from purified nucleus and cytoplasm fractions from *T. cruzi* epimastigotes to perform RNA-Seq and find whether transcripts were differentially distributed between these two compartments.

The fractionation reproducibility and the quality of the fractions were confirmed by protein and RNA compartment specific markers. An expected profile of protein markers was obtained for each of the independent experiment. The whole cell and cytoplasm extracts exhibited similar content of the different markers except for the nuclear marker TcH2A, whose almost disappearance in the later fractions supports the appropriateness of the selected fractionation method. Analogously, the mitochondrion, the reservosome as well as the cytoplasm markers exhibited weak to null signals in the nuclear fractions. The observed pattern may be pointing out to a negligible cross contamination of the fractions, however, the faint signal of TcH2A in the cytoplasm may also be attributed to the protein synthesis occurring at this compartment.

The content of total RNA per cell obtained from *T. cruzi* epimastigotes (0.6 ± 0.1 pg) is in good accordance with reported measurements in *T. brucei* bloodstream (0.5 pg, Haanstra et al., 2008, and procyclic forms 0.8 pg, Pays et al., 1993, and 1.1 pg, Haanstra et al., 2008). Variations of total RNA per cell content in the same range (0.5 to 1.6 pg) have also been reported for *Saccharomyces cerevisiae*, depending on the growth conditions (von der Haar, 2008). While the amount of total RNA in the nuclear fraction may vary with the physiological state of the cell, a rough estimation of 5 to 10% is commonly accepted based on the contribution of rRNA, mRNA and tRNA plus pre-mRNAs of 85, 2, and 13% respectively (Finka et al., 2015). Thus, we considered that the distribution of total RNA between nucleus and cytoplasm that we obtained (10 and 90% respectively) was consistent with the literature.

Thereafter, the RNA derived from the fractions of four independent experiments were pooled, sequenced and analyzed. Despite the fact that the genome of CLBrenner Esmeraldo-like *T. cruzi* strain was used to map the RNA transcripts, the use of a paired end strategy strengthens the reliability of the assignments, and more than 9,000 genes (out of 10,600) were detected in each of the transcriptomes analyzed. Thus, in *T. cruzi* epimastigotes most of the genes are detected in the RNA extracted from whole cells or the nuclear or cytoplasmic fractions. While the good correlation between cytoplasmic and whole cell transcriptome (*r* = 0.98 *p* < 0.0001) was expected due to the reduced total RNA content of the nuclear fraction, the wide representation of transcribed genome in the cytoplasmic fraction (85%) further supports pervasive mechanisms for transcription and maturation as previously proposed (Smirich et al., 2015). However, though

almost all the genes are shared by nuclear and cytoplasmic transcriptomes (84%), the poor correlation obtained between nucleus and cytoplasm transcripts levels ($r = 0.53$, $p < 0.001$) points out to the existence of an active compartment role. Indeed, 20% of the transcripts detected (NET+CET) showed at least a 2-fold change in expression levels between the two compartments analyzed. The reliability of these changes was confirmed for eight genes by qRT-PCR in each of the four independent biological replicates. A high correlation between qRT-PCR and RNA-Seq data was found. In addition, the low dispersion of the obtained qRT-PCR results provides evidence for the similarity of the used replicates. Altogether these findings support the existence of differentially distributed transcripts between nucleus and cytoplasm.

In agreement with previous reports for three human cell lines (Solnestam et al., 2012), where longer transcripts in the nucleus were inferred, we found that NET genes are significantly longer either than the NDT or CET genes. While this profile is mostly due to the size of the CDSs, the size of the UTRs also accompanies this trend. Concordantly, human genome-wide studies have revealed that highly expressed genes are significantly smaller and produce shorter mRNAs with shorter 3' UTRs (Chiaromonte et al., 2003; Urrutia and Hurst, 2003). Similar results have been reported for *Arabidopsis thaliana* (Caldwell et al., 2010) and yeasts (Lackner et al., 2007; Lu et al., 2007). Recently, the length of mRNA has been related to developmental expression in *T. brucei* (Antwi et al., 2016). This finding is consistent with the role of nuclear compartmentalization in gene expression regulation during differentiation that we have here unraveled. Nonetheless, other features of gene sequences, such as G+C contents, are also associated with gene expression levels (Konu and Li, 2002). We found that NET genes have a significantly higher GC% content than NDT genes either at CDSs or UTRs. Since this trend is also accompanied by GC3 content at the CDS, codon usage may be contributing to further regulate gene expression of these transcripts. Codon usage has been identified as a major determinant of mRNA stability (Presnyak et al., 2015). In addition, high GC contents have been related to thermostability in different studies in bacteria (Nishio et al., 2003). Indeed, we found a significantly higher predicted thermodynamic stability for the 5' UTR of NET genes which may be further influencing their low translation rate. Therefore, although the subcellular fractionation to obtain a cytoplasmic transcriptome may not introduce a significant advantage comparing to the whole cell transcriptome, we here provide evidence for a putative bias estimation due to length and composition of the transcripts in *T. cruzi* epimastigotes.

Considering that the transcriptome derived from a purified cytoplasmic fraction, following separation of the nuclear contribution, could better reflect the translationally active transcripts than the whole cell transcriptome, we analyzed the correlations between the transcriptome and the proteome. Again, we found that cytoplasmic transcriptome does not significantly improve the estimation of protein abundance obtained from whole cell parasite transcriptome. However, for the set of genes which have differential transcript abundance between the nucleus and cytoplasmic compartments, a functional bias was observed. The transcripts enriched in the cytoplasm correspond to genes

which are: expressed in the epimastigote following ontology analyses; translationally active according to the ribosome footprint analysis (Smircich et al., 2015), and are widely represented in the epimastigote proteome (de Godoy et al., 2012). Meanwhile, an important number of the transcripts enriched in the nucleus correspond to pseudogenes. This is not surprising since several studies have demonstrated the functional roles of the pseudogene transcripts in post-transcription regulation mainly in the nucleus, either as competing endogenous RNAs or *trans*-acting RNA (Johnsson et al., 2014; Sen et al., 2014). A deeper analysis of the specific pseudogene transcript nature, sequence and structure could shed light to better understand their mechanism of action in this organism. In addition to pseudogenes, several *bona fide* gene transcripts showed a nuclear enrichment. From reported ribosome footprint studies, we found that these genes are less translated than the average in epimastigote. Likewise, very few of them are present in the epimastigote proteome. Ontology analyses revealed that the genes with nucleus enriched transcripts in the epimastigote are distinctive of other life cycle stages. Indeed, several genes belong to the extended multigenic families that codify for surface antigens such as the trans-sialidase (Schenkman et al., 1994; Nardy et al., 2016), the mucin associated surface proteins (Bartholomeu et al., 2009; dos Santos et al., 2012), the dispersed gene family (Kawashita et al., 2009; De Pablos and Osuna, 2012) and the glycoprotein family gp63 (Cuevas et al., 2003; Yao, 2010). Their function has been frequently associated with parasite protection and evasion of the host immune system (dos Santos et al., 2012; De Pablos et al., 2016; Houston-Ludlam et al., 2016). Since these gene families are absent in the phylogenetically closest free-living trypanosomatid *Bodo saltans*, it has been recently suggested that their acquisition could be related to the parasite adaptation to the hostile environments of the hosts (Jackson et al., 2016). Besides, the family of genes with hot spot for the insertion of retroelements, which also has nucleus enriched transcripts, is known to increase its expression upon exposure to γ radiation (Grynberg et al., 2012). Thus, the nuclear compartmentation may contribute to maintain the availability of transcripts necessary to readily respond to environmental challenges.

Using the available data for ribosome footprinting and proteome of *T. cruzi* metacyclic trypomastigote, we analyzed the fate of the NET and CET epimastigote genes in the infective non-replicative parasite stage. We could demonstrate a significant change of the ribosome occupancy of transcripts and encoded proteins from these two transcript sets during metacyclogenesis, that is in good agreement with a role of the nuclear compartment in determining which mRNAs will be retained or released to the translationally active cytoplasmic compartment.

We further studied whether this developmental relocalization could be also proved through direct visualization of selected transcripts. In epimastigotes, the ribosomal protein L44 mRNA is found in the cytoplasm, mainly at the nucleus boundaries, and colocalizing with the polyadenylated mRNAs. Meanwhile, in addition to its reported decreased expression in the metacyclic trypomastigote, a wider spread pattern along the parasite, with no nuclear delimitation and a lesser colocalization with polyadenylated mRNAs was observed. On the other hand, the adenylate cyclase, known to be expressed in the metacyclic

trypomastigote stage showed a diminished nuclear transcript content in this stage in comparison to the epimastigote stage. The change of location was even more evident for a gp63 gene involved in the infectivity process, whose transcript was markedly restricted to the nuclear translational inactive compartment in the non-infective epimastigote stage, moving to the translationally active cytoplasm, with a good colocalization with polyadenylated mRNAs, in the infective metacyclic trypomastigote stage.

These findings support an active role of the nuclear compartment confining the transcripts that are not immediately needed to produce proteins but may be rapidly required upon differentiation. The analysis of the actual contribution of each transcript species, i.e., unmaturing, partial or complete processed transcript to the nuclear content would shed light about the implied regulatory mechanisms. In addition, whether this differential RNA distribution is achieved by nuclear retention or differential nuclear-cytoplasm RNA decay should be investigated. Altogether these results provide strong evidence indicating that the nucleus–cytoplasm partitioning constitutes a control step that contributes to the differential regulation of a life cycle specific set of genes in *T. cruzi*.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: MD, JS, and BG; Performed the experiments: LP; Analyzed the data: LP, PS, AD,

LB, MD, JS, and BG; Contributed reagents/materials/analysis tools: PS, AD, LB, JS, and BG; Wrote the paper: LP, PS, MD, JS, and BG.

FUNDING

This work was supported by CSIC and CAP UdelaR, ANII and PEDECIBA, Uruguay.

ACKNOWLEDGMENTS

We wish to thank Gisele Picchi and Eloise Guerra-Slomp (Instituto Carlos Chagas, FioCruz Paraná, Brasil) and Dolores Piñeyro (Institut Pasteur Montevideo, Uruguay) for kindly providing TcH2A, TcRBP40, and TcTXN1 specific antibodies, respectively. We acknowledge all members of the Laboratorio de Interacciones Moleculares at UDELAR and the Department of Genomics at IIBCE for constant discussion and technical support. We also thank the colleagues that have provided critical insight into this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fcell.2017.00008/full#supplementary-material>

REFERENCES

- Antwi, E. B., Haanstra, J. R., Ramasamy, G., Jensen, B., Droll, D., Rojas, F., et al. (2016). Integrative analysis of the *Trypanosoma brucei* gene expression cascade predicts differential regulation of mRNA processing and unusual control of ribosomal protein expression. *BMC Genomics* 17:306. doi: 10.1186/s12864-016-2624-3
- Araujo, P. R., and Teixeira, S. M. (2011). Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. *Mem. Inst. Oswaldo Cruz* 106, 257–266. doi: 10.1590/S0074-02762011000300002
- Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., et al. (2015). Nuclear retention of mRNA in mammalian tissues. *Cell Rep.* 13, 2653–2662. doi: 10.1016/j.celrep.2015.11.036
- Bai, B., Yegnasubramanian, S., Wheelan, S. J., and Laiho, M. (2014). RNA-Seq of the nucleolus reveals abundant SNORD44-derived small RNAs. *PLoS ONE* 9: e107519. doi: 10.1371/journal.pone.0107519
- Bartholomeu, D. C., Cerqueira, G. C., Leao, A. C., daRocha, W. D., Pais, F. S., Macedo, C., et al. (2009). Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Res.* 37, 3407–3417. doi: 10.1093/nar/gkp172
- Bayer-Santos, E., Gentil, L. G., Cordero, E. M., Correa, P. R., and da Silveira, J. F. (2012). Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes. *Acta Trop.* 123, 230–233. doi: 10.1016/j.actatropica.2012.03.014
- Caldwell, R., Lin, Y. X., and Zhang, R. (2010). Assessment of length distributions between non-coding and coding sequences amongst two model organisms. *Int. J. Data Min. Bioinform.* 4, 535–552. doi: 10.1504/IJDMB.2010.035899
- Camara Mde, L., Bouvier, L. A., Canepa, G. E., Miranda, M. R., and Pereira, C. A. (2013). Molecular and functional characterization of a *Trypanosoma cruzi* nuclear adenylate kinase isoform. *PLoS Negl. Trop. Dis.* 7:e2044. doi: 10.1371/journal.pntd.0002044
- Cassola, A. (2011). RNA Granules Living a Post-transcriptional Life: the Trypanosomes' Case. *Curr. Chem. Biol.* 5, 108–117. doi: 10.2174/2212796811105020108
- Cassola, A., De Gaudenzi, J. G., and Frasch, A. C. (2007). Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes. *Mol. Microbiol.* 65, 655–670. doi: 10.1111/j.1365-2958.2007.05833.x
- Cassola, A., and Frasch, A. C. (2009). An RNA recognition motif mediates the nucleocytoplasmic transport of a trypanosome RNA-binding protein. *J. Biol. Chem.* 284, 35015–35028. doi: 10.1074/jbc.M109.031633
- Chagas, C. (1909). Nova tripanozomíaze humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi*, agente etiologico da nova entidade mórbida do homem. *Mem. Inst. Oswaldo Cruz* 1, 159–219. doi: 10.1590/S0074-02761909000200008
- Chiaromonte, F., Miller, W., and Bouhassira, E. E. (2003). Gene length and proximity to neighbors affect genome-wide expression levels. *Genome Res.* 13, 2602–2608. doi: 10.1101/gr.1169203
- Clayton, C. E. (2016). Gene expression in Kinetoplastids. *Curr. Opin. Microbiol.* 32, 46–51. doi: 10.1016/j.mib.2016.04.018
- Conesa, A., and Gotz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832
- Contreras, V. T., Araujo-Jorge, T. C., Bonaldo, M. C., Thomaz, N., Barbosa, H. S., Meirelles Mde, N., et al. (1988). Biological aspects of the Dm 28c clone of *Trypanosoma cruzi* after metacyclogenesis in chemically defined media. *Mem. Inst. Oswaldo Cruz* 83, 123–133. doi: 10.1590/S0074-02761988000100016
- Contreras, V. T., Morel, C. M., and Goldenberg, S. (1985). Stage specific gene expression precedes morphological changes during *Trypanosoma cruzi* metacyclogenesis. *Mol. Biochem. Parasitol.* 14, 83–96. doi: 10.1016/0166-6851(85)90108-2

- Cuevas, I. C., Cazzulo, J. J., and Sanchez, D. O. (2003). gp63 homologues in *Trypanosoma cruzi*: surface antigens with metalloprotease activity and a possible role in host cell infection. *Infect. Immun.* 71, 5739–5749. doi: 10.1128/IAI.71.10.5739-5749.2003
- da Silva Augusto, L., Moretti, N. S., Ramos, T. C., de Jesus, T. C., Zhang, M., Castilho, B. A., et al. (2015). A membrane-bound eIF2 alpha kinase located in endosomes is regulated by heme and controls differentiation and ROS levels in *Trypanosoma cruzi*. *PLoS Pathog.* 11:e1004618. doi: 10.1371/journal.ppat.1004618
- De Gaudenzi, J., Frasch, A. C., and Clayton, C. (2005). RNA-binding domain proteins in Kinetoplastids: a comparative analysis. *Eukaryot. Cell* 4, 2106–2114. doi: 10.1128/EC.4.12.2106-2114.2005
- De Gaudenzi, J. G., Jager, A. V., Izcovich, R., and Campo, V. A. (2016). Insights into the regulation of mRNA processing of polycistronic transcripts mediated by DRBD4/PTB2, a trypanosome homolog of the polypyrimidine tract-binding protein. *J. Eukaryot. Microbiol.* 63, 440–452. doi: 10.1111/jeu.12288
- de Godoy, L. M., Marchini, F. K., Pavoni, D. P., Rampazzo Rde, C., Probst, C. M., Goldenberg, S., et al. (2012). Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis. *Proteomics* 12, 2694–2703. doi: 10.1002/pmic.201200078
- Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4:P3. doi: 10.1186/gb-2003-4-5-p3
- De Pablos, L. M., Diaz Lozano, I. M., Jercic, M. I., Quinzada, M., Gimenez, M. J., Calabuig, E., et al. (2016). The C-terminal region of *Trypanosoma cruzi* MASP5 is antigenic and secreted via exovesicles. *Sci. Rep.* 6:27293. doi: 10.1038/srep27293
- De Pablos, L. M., and Osuna, A. (2012). Multigene families in *Trypanosoma cruzi* and their role in infectivity. *Infect. Immun.* 80, 2258–2264. doi: 10.1128/iai.06225-11
- dos Santos, S. L., Freitas, L. M., Lobo, F. P., Rodrigues-Luiz, G. F., Mendes, T. A., Oliveira, A. C., et al. (2012). The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLoS Negl. Trop. Dis.* 6:e1779. doi: 10.1371/journal.pntd.0001779
- Duhagon, M. A., Pastro, L., Sotelo-Silveira, J. R., Perez-Diaz, L., Maugeri, D., Nardelli, S. C., et al. (2009). The *Trypanosoma cruzi* nucleic acid binding protein Tc38 presents changes in the intramitochondrial distribution during the cell cycle. *BMC Microbiol.* 9:34. doi: 10.1186/1471-2180-9-34
- Fadda, A., Ryten, M., Droll, D., Rojas, F., Farber, V., Haanstra, J. R., et al. (2014). Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol. Microbiol.* 94, 307–326. doi: 10.1111/mmi.12764
- Fiebig, M., Gluenz, E., Carrington, M., and Kelly, S. (2014). SLaP mapper: a webserver for identifying and quantifying spliced-leader addition and polyadenylation site usage in kinetoplastid genomes. *Mol. Biochem. Parasitol.* 196, 71–74. doi: 10.1016/j.molbiopara.2014.07.012
- Finka, A., Sood, V., Quadroni, M., Rios Pde, L., and Goloubinoff, P. (2015). Quantitative proteomics of heat-treated human cells show an across-the-board mild depletion of housekeeping proteins to massively accumulate few HSPs. *Cell Stress Chaperones* 20, 605–620. doi: 10.1007/s12192-015-0583-2
- Garcia-Silva, M. R., Frugier, M., Tosar, J. P., Correa-Dominguez, A., Ronaltes-Alves, L., Parodi-Talice, A., et al. (2010). A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules. *Mol. Biochem. Parasitol.* 171, 64–73. doi: 10.1016/j.molbiopara.2010.02.003
- Gomez, M. L., Erijman, L., Arauzo, S., Torres, H. N., and Tellez-Inon, M. T. (1989). Protein kinase C in *Trypanosoma cruzi* epimastigote forms: partial purification and characterization. *Mol. Biochem. Parasitol.* 36, 101–108. doi: 10.1016/0166-6851(89)90182-5
- Grynberg, P., Passos-Silva, D. G., Mourao Mde, M., Hirata, R. Jr., Macedo, A. M., Machado, C. R., et al. (2012). *Trypanosoma cruzi* gene expression in response to gamma radiation. *PLoS ONE* 7:e29596. doi: 10.1371/journal.pone.0029596
- Guerra-Slomp, E. P., Probst, C. M., Pavoni, D. P., Goldenberg, S., Krieger, M. A., and Dallagiovanna, B. (2012). Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs. *Biochem. Biophys. Res. Commun.* 420, 302–307. doi: 10.1016/j.bbrc.2012.02.154
- Haanstra, J. R., Stewart, M., Luu, V. D., van Tuijl, A., Westerhoff, H. V., Clayton, C., et al. (2008). Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form *Trypanosoma brucei*. *J. Biol. Chem.* 283, 2495–2507. doi: 10.1074/jbc.M705782200
- Haile, S., and Papadopolou, B. (2007). Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Curr. Opin. Microbiol.* 10, 569–577. doi: 10.1016/j.mib.2007.10.001
- Holetz, F. B., Correa, A., Avila, A. R., Nakamura, C. V., Krieger, M. A., and Goldenberg, S. (2007). Evidence of P-body-like structures in *Trypanosoma cruzi*. *Biochem. Biophys. Res. Commun.* 356, 1062–1067. doi: 10.1016/j.bbrc.2007.03.104
- Houston-Ludlam, G. A., Belew, A. T., and El-Sayed, N. M. (2016). Comparative Transcriptome Profiling of Human Foreskin Fibroblasts Infected with the Sylvio and Y Strains of *Trypanosoma cruzi*. *PLoS ONE* 11:e0159197. doi: 10.1371/journal.pone.0159197
- Inoue, A. H., Serpeloni, M., Hiraiwa, P. M., Yamada-Ogatta, S. F., Muniz, J. R., Motta, M. C., et al. (2014). Identification of a novel nucleocytoplasmic shuttling RNA helicase of trypanosomes. *PLoS ONE* 9:e109521. doi: 10.1371/journal.pone.0109521
- Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., et al. (2016). Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr. Biol.* 26, 161–172. doi: 10.1016/j.cub.2015.11.055
- Johnsson, P., Morris, K. V., and Grander, D. (2014). Pseudogenes: a novel source of trans-acting antisense RNAs. *Methods Mol. Biol.* 1167, 213–226. doi: 10.1007/978-1-4939-0835-6_14
- Kawashita, S. Y., da Silva, C. V., Mortara, R. A., Burleigh, B. A., and Briones, M. R. (2009). Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins. *Mol. Biochem. Parasitol.* 165, 19–31. doi: 10.1016/j.molbiopara.2008.12.010
- Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nat. Rev. Mol. Cell Biol.* 8, 761–773. doi: 10.1038/nrm2255
- Konu, O., and Li, M. D. (2002). Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. *J. Mol. Evol.* 54, 35–41. doi: 10.1007/s00239-001-0015-z
- Kramer, S. (2014). RNA in development: how ribonucleoprotein granules regulate the life cycles of pathogenic protozoa. *Wiley Interdiscip. Rev. RNA* 5, 263–284. doi: 10.1002/wrna.1207
- Kramer, S., and Carrington, M. (2011). Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends Parasitol.* 27, 23–30. doi: 10.1016/j.pt.2010.06.011
- Kramer, S., Piper, S., Estevez, A., and Carrington, M. (2016). Polycistronic trypanosome mRNAs are a target for the exosome. *Mol. Biochem. Parasitol.* 205, 1–5. doi: 10.1016/j.molbiopara.2016.02.009
- Lackner, D. H., Beilharz, T. H., Marguerat, S., Mata, J., Watt, S., Schubert, F., et al. (2007). A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell* 26, 145–155. doi: 10.1016/j.molcel.2007.03.002
- Li, Y., Shah-Simpson, S., Okrah, K., Belew, A. T., Choi, J., Caradonna, K. L., et al. (2016). Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection. *PLoS Pathog.* 12:e1005511. doi: 10.1371/journal.ppat.1005511
- Li, Z. H., De Gaudenzi, J. G., Alvarez, V. E., Mendiondo, N., Wang, H., Kissinger, J. C., et al. (2012). A 43-nucleotide U-rich element in 3'-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes. *J. Biol. Chem.* 287, 19058–19069. doi: 10.1074/jbc.M111.338699
- Liao, J. Y., Ma, L. M., Guo, Y. H., Zhang, Y. C., Zhou, H., Shao, P., et al. (2010). Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS ONE* 5:e10563. doi: 10.1371/journal.pone.0010563
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26. doi: 10.1186/1748-7188-6-26

- Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 25, 117–124. doi: 10.1038/nbt1270
- Minning, T. A., Weatherly, D. B., Atwood, J. III, Orlando, R., and Tarleton, R. L. (2009). The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. *BMC Genomics* 10:370. doi: 10.1186/1471-2164-10-370
- Nardy, A. F., Freire-de-Lima, C. G., Perez, A. R., and Morrot, A. (2016). Role of *Trypanosoma cruzi* Trans-sialidase on the escape from host immune surveillance. *Front. Microbiol.* 7:348. doi: 10.3389/fmicb.2016.00348
- Nazer, E., Verdun, R. E., and Sanchez, D. O. (2011). Nucleolar localization of RNA binding proteins induced by actinomycin D and heat shock in *Trypanosoma cruzi*. *PLoS ONE* 6:e19920. doi: 10.1371/journal.pone.0019920
- Nazer, E., Verdun, R. E., and Sanchez, D. O. (2012). Severe heat shock induces nucleolar accumulation of mRNAs in *Trypanosoma cruzi*. *PLoS ONE* 7:e43715. doi: 10.1371/journal.pone.0043715
- Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., Sugimoto, S., et al. (2003). Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res.* 13, 1572–1579. doi: 10.1101/gr.1285603
- Palazzo, A. F., and Akef, A. (2012). Nuclear export as a key arbiter of “mRNA identity” in eukaryotes. *Biochim. Biophys. Acta* 1819, 566–577. doi: 10.1016/j.bbagr.2011.12.012
- Pays, E., Hanocq-Quertier, J., Hanocq, F., Van Assel, S., Nolan, D., and Rolin, S. (1993). Abrupt RNA changes precede the first cell division during the differentiation of *Trypanosoma brucei* bloodstream forms into procyclic forms *in vitro*. *Mol. Biochem. Parasitol.* 61, 107–114. doi: 10.1016/0166-6851(93)90163-R
- Pineyro, M. D., Parodi-Talice, A., Portela, M., Arias, D. G., Guerrero, S. A., and Robello, C. (2011). Molecular characterization and interactome analysis of *Trypanosoma cruzi* tryparedoxin 1. *J. Proteomics* 74, 1683–1692. doi: 10.1016/j.jprot.2011.04.006
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111–1124. doi: 10.1016/j.cell.2015.02.029
- Schenkman, S., Eichinger, D., Pereira, M. E., and Nussenzweig, V. (1994). Structural and functional properties of *Trypanosoma* trans-sialidase. *Annu. Rev. Microbiol.* 48, 499–523. doi: 10.1146/annurev.mi.48.100194.002435
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Sen, R., Ghosal, S., Das, S., Balti, S., and Chakrabarti, J. (2014). Competing endogenous RNA: the key to posttranscriptional regulation. *Sci. World J.* 2014:896206. doi: 10.1155/2014/896206
- Serpeloni, M., Moraes, C. B., Muniz, J. R., Motta, M. C., Ramos, A. S., Kessler, R. L., et al. (2011a). An essential nuclear protein in trypanosomes is a component of mRNA transcription/export pathway. *PLoS ONE* 6:e20730. doi: 10.1371/journal.pone.0020730
- Serpeloni, M., Vidal, N. M., Goldenberg, S., Avila, A. R., and Hoffmann, F. G. (2011b). Comparative genomics of proteins involved in RNA nucleocytoplasmic export. *BMC Evol. Biol.* 11:7. doi: 10.1186/1471-2148-11-7
- Smircich, P., Eastman, G., Bispo, S., Duhagon, M. A., Guerra-Slompo, E. P., Garat, B., et al. (2015). Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genomics* 16:443. doi: 10.1186/s12864-015-1563-8
- Solnestam, B. W., Stranneheim, H., Hallman, J., Kaller, M., Lundberg, E., Lundeberg, J., et al. (2012). Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. *BMC Genomics* 13:574. doi: 10.1186/1471-2164-13-574
- Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakraborty, S., Djebali, S., et al. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625. doi: 10.1101/gr.134445.111
- Trask, H. W., Cowper-Sal-lari, R., Sartor, M. A., Gui, J., Heath, C. V., Renuka, J., et al. (2009). Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. *RNA* 15, 1917–1928. doi: 10.1261/rna.1677409
- Urrutia, A. O., and Hurst, L. D. (2003). The signature of selection mediated by expression on human genes. *Genome Res.* 13, 2260–2264. doi: 10.1101/gr.641103
- von der Haar, T. (2008). A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC Syst. Biol.* 2:87. doi: 10.1186/1752-0509-2-87
- World Health Organization (WHO) (2015). *Chagas Disease (American trypanosomiasis)*. Fact sheet N°340. Geneva: WHO. Available online at: <http://www.who.int/mediacentre/factsheets/fs340/en/>
- Yao, C. (2010). Major surface protease of trypanosomatids: one size fits all? *Infect. Immun.* 78, 22–31. doi: 10.1128/IAI.00776-09
- Zaghlool, A., Ameer, A., Nyberg, L., Halvardson, J., Grabherr, M., Cavelier, L., et al. (2013). Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues. *BMC Biotechnol.* 13:99. doi: 10.1186/1472-6750-13-99

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pastro, Smircich, Di Paolo, Becco, Duhagon, Sotelo-Silveira and Garat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Material Suplementario

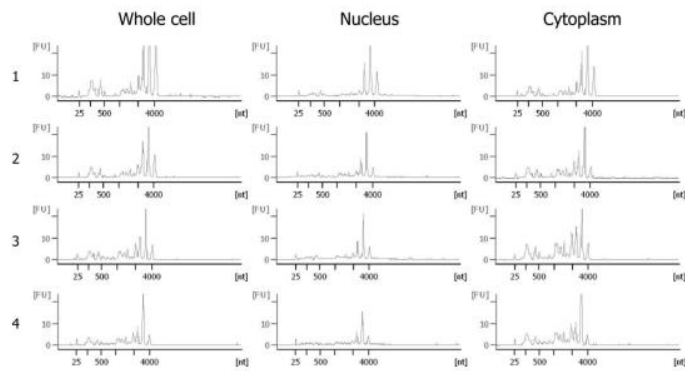


Figure S1. RNA electrophoretic profile of subcellular fractions of *T. cruzi* epimastigotes

The RNA electrophoretic profiles of whole cell, nucleus and cytoplasm fractions of four independent biological replicates are shown (1 to 4).

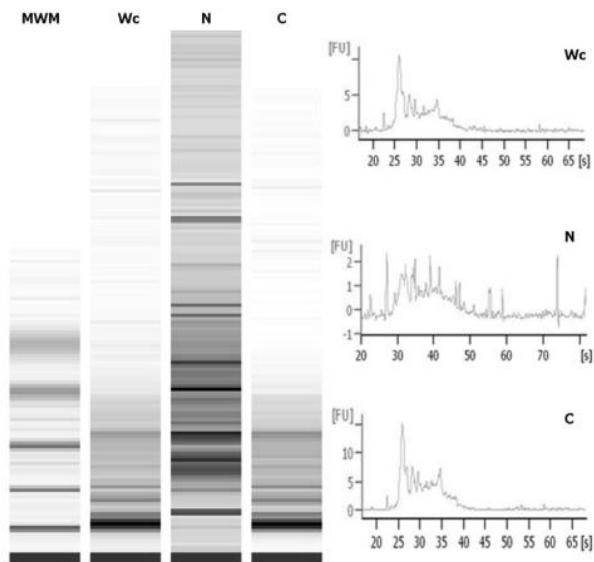


Figure S2. RNA electrophoretic profile of pooled subcellular fractions of *T. cruzi* epimastigotes

The RNA electrophoretic profile for the pooled whole cell (Wc), nucleus (N) and cytoplasm (C) fractions of four independent biological replicates is shown. MWM: molecular weight marker.

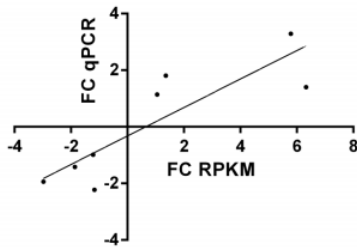


Figure S3. Correlations between qRT-PCR and RNA-Seq results
FC of RPKM in N / C vs FC of qRT-PCR in N / C.

A

	Wc1	Wc2	Wc3	Wc4				
Wc1	1.00	0.95	0.98	0.95	Mean	0.97		
Wc2		1.00	0.96	0.98			SD	0.02
Wc3			1.00	0.98				
Wc4				1.00				

	N1	N2	N3	N4				
N1	1.00	0.94	0.97	0.96	Mean	0.96		
N2		1.00	0.95	0.98			SD	0.02
N3			1.00	0.96				
N4				1.00				

	C1	C2	C3	C4				
C1	1.00	0.96	0.98	0.96	Mean	0.97		
C2		1.00	0.96	0.98			SD	0.01
C3			1.00	0.96				
C4				1.00				

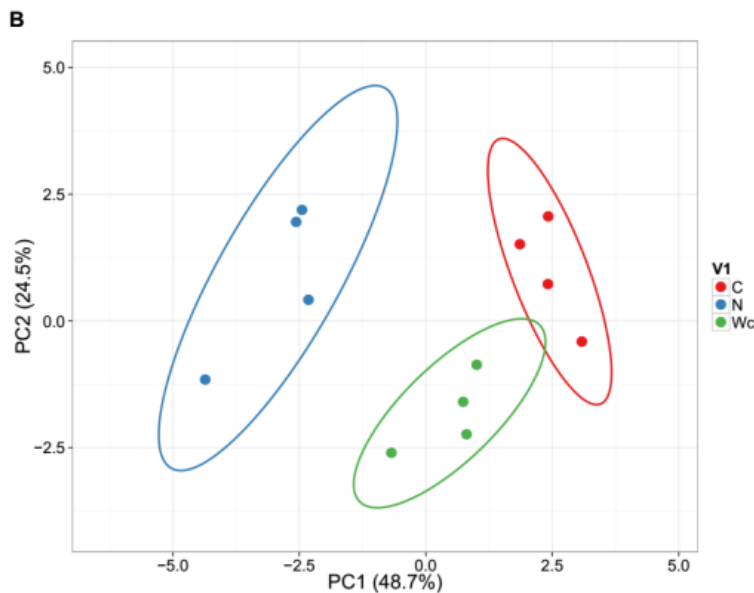


Figure S4. Analysis of the expression patterns on the independent quadruplicate fractions that were pooled for RNA-Seq
The quantification of transcripts of 12 genes (see Table S1 for IDs and primers) was done by qRT-PCR on each sample: whole cell (Wc), nucleus (N) and cytoplasm (C) fractions of four independent biological replicates. For normalization *gapdh* mRNA was used. **(A)** Pearson correlation analysis **(B)** Principal component analysis plot.

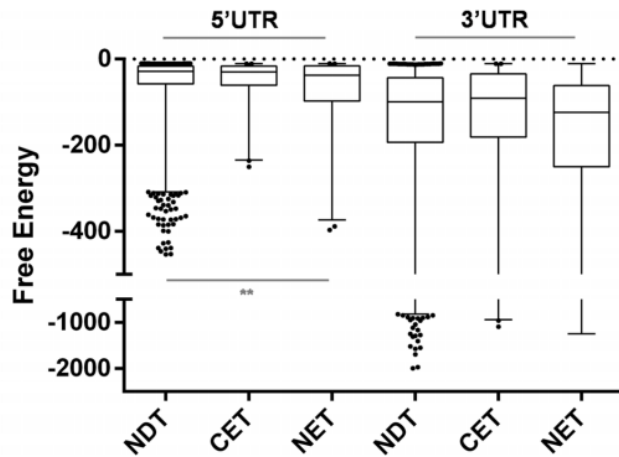


Figure S4. Free energy for the UTR predicted structures of genes with differentially distributed nucleus-cytoplasm abundance transcripts in *T. cruzi* epimastigotes

Box plots of the free energy (kcal/mol) for the predicted structures both at the 5'UTR and 3'UTR of genes with cytoplasmic enriched transcript abundance (CET), with nuclear enriched transcript abundance (NET) and with not differentially distributed transcript abundance (NDT) (** test t $p < 0.01$).

Table S1. Specific primers used

	Target Gene CL Brenner Esmeraldo like	Sequence primer Forward or probe	Sequence primer Reverse
snoRNA	TcCLB.510283.26	5'- ATGATGATTGACGTAACATCACAGAC -3'	5'- TCAGACTGGGGCGACAGTGTTC -3'
Gapdh	TcCLB.506943.50 - TcCLB.506943.60	5'- CGACAACGAGTGGGATACT -3'	5'-CTACAACCTTGCCGAACGAT -3'
rRNA-Lgama-Lalfa		5'- aaggggcaacagagaacctggg -3'	5'- agctcaccgaggaggattg -3'
rRNA-Lgama-Lbeta		5'- gtcattgttgagttctgcgc -3'	5'- ctctatgaccagagcattgatgc -3'
PK	TcCLB.510343.15	GTTGAGGCCTTCCTGCTTGT	CAAATGCGTTTCTGTCAATTGC
GP63	TcCLB.506779.180	CTGCTTGACTGGGAGAACC	CACACTTAGCGTGCCCTCAC
DGF-1	TcCLB.509883.9	5'- GACGGTGTGGACTGTGGTG -3'	5'- ACTCACAGCCGCTGATGCTC -3'
HSP60-Mit	TcCLB.507641.290	5'- GTGCAGGAGGGCATTGTTC -3'	5'- GTGAGTGACGAGTCGCCAAG -3'
Hypothetical	TcCLB.506931.4	5'- TGGGTAGCCTCACGAAAGT -3'	5'- TCGTCCATTGGTAGAATGC -3'
RBP	TcCLB.506649.80	5'- TGTGTTTTTGCCCAAGAATG -3'	5'- GCTGTCAAACGCACCAAGG -3'
RP-L24	TcCLB.503611.20	5'- AAGAAGAACCCCGCTTTGT -3'	5'- GGTGTGCATCTTCGGTTGA -3'
PRPS	TcCLB.508717.30	5'- GGTGGATGCGTTGGTAGAGG -3'	5'- ACCGGGATATGCAACTCC -3'
polyApolM	TcCLB.510317.30	5'- TAGCGATCTCGATATTGTGCTG -3'	5'- GCAGCAGCGGAGGGAATC -3'
polyApol_Inm	TcCLB.510317.34	5'- GGTGCCTCCGTTGCAGATAG -3'	5'- ggggattagagaggggtgg -3'
L44_FAM	TcCLB.507105.40	5'- AATGATGGACTTGACGTTGGAGCATTGGAGCTTCAGGACA -3'	
GP63_FAM	TcCLB.511211.90	5'- AGACGACCATCATCCAGCCGAAACGCACGGACCCGCCA -3'	
AdCyc_FAM	TcCLB.511043.60	5'- GCGACGAGCACGGACAGTAATAATOGAATCATCCGTCGA -3'	
polyT-Cy5		5'- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT -3'	

Table S2. RNAseq data of subcellular fractions of *T. cruzi* epimastigotes

	Whole cell	Nucleus	Cytoplasm
RNA-seq data (nt)	9,727,558,056	7,310,872,476	9,360,517,996
Average length (nt)	100	100	100
Mean library size (nt)	450	409	438
Reads mapped to genome	23,830,921	23,460,632	21,076,033
Reads mapped to transcriptome	18,739,614	17,793,314	16,532,099
Reads mapped to annotated CDSs	6,316,999	8,344,884	4,907,619

Table S3. Correlation analysis between protein abundances and transcriptomes from whole cell, cytoplasmic and nuclear fractions of *T. cruzi* epimastigotes

Fraction / Set	Transcriptome*	Proteome**	Spearman	Significance
Whole cell	6039	1280	0.5494	<0.0001
Cytoplasm	7746	1280	0.5638	<0.0001
Nucleus	8276	1280	0.4231	<0.0001

* Number of genes with more than 10 reads per transcript in the corresponding fraction or set.

** Number of transcripts for each transcriptome set whose encoded proteins were quantitatively determined by de Godoy et al. .

Table S4. Differentially distributed transcripts between nucleus y cytoplasm IDs.

Differentially distributed transcripts					
Cytoplasm					
TcCLB.507677.39	TcCLB.506563.79	TcCLB.510101.30	TcCLB.506695.20	TcCLB.511707.20	TcCLB.508647.110
TcCLB.507603.270	TcCLB.506567.77	TcCLB.511325.40	TcCLB.507625.99	TcCLB.507093.250	TcCLB.504005.54
TcCLB.473111.10	TcCLB.507809.60	TcCLB.508135.10	TcCLB.507809.39	TcCLB.508277.120	TcCLB.506287.209
TcCLB.510877.55	TcCLB.477983.10	TcCLB.506755.260	TcCLB.504165.20	TcCLB.506753.110	TcCLB.432027.9
TcCLB.511067.20	TcCLB.510341.10	TcCLB.511529.210	TcCLB.511039.6	TcCLB.506289.30	TcCLB.503899.119
TcCLB.506655.20	TcCLB.506495.9	TcCLB.504033.80	TcCLB.503873.10	TcCLB.503893.40	TcCLB.508041.10
TcCLB.506777.20	TcCLB.506533.106	TcCLB.504153.110	TcCLB.506563.220	TcCLB.506289.9	TcCLB.508539.20
TcCLB.503899.110	TcCLB.510349.60	TcCLB.508675.29	TcCLB.510259.30	TcCLB.510743.120	TcCLB.508277.300
TcCLB.506563.180	TcCLB.508549.20	TcCLB.510761.14	TcCLB.504057.178	TcCLB.508777.70	TcCLB.503689.20
TcCLB.511179.160	TcCLB.506227.40	TcCLB.503959.78	TcCLB.511217.180	TcCLB.508719.50	TcCLB.503527.10
TcCLB.506937.30	TcCLB.506181.59	TcCLB.508805.80	TcCLB.510999.39	TcCLB.507817.9	TcCLB.503579.80
TcCLB.511001.240	TcCLB.506953.30	TcCLB.509831.50	TcCLB.510609.34	TcCLB.506729.50	TcCLB.511357.9
TcCLB.509007.50	TcCLB.506529.50	TcCLB.509833.10	TcCLB.506753.40	TcCLB.508357.23	TcCLB.510089.210
TcCLB.511369.15	TcCLB.509965.134	TcCLB.507873.34	TcCLB.506529.324	TcCLB.507993.250	TcCLB.506341.10
TcCLB.509197.50	TcCLB.509167.140	TcCLB.509171.80	TcCLB.503599.60	TcCLB.510145.20	TcCLB.510911.10
TcCLB.507105.40	TcCLB.509011.90	TcCLB.507389.60	TcCLB.509979.90	TcCLB.507093.100	TcCLB.484299.10
TcCLB.510879.220	TcCLB.506773.90	TcCLB.506885.30	TcCLB.509393.30	TcCLB.503419.30	TcCLB.508133.19
TcCLB.507993.30	TcCLB.506529.40	TcCLB.503989.10	TcCLB.503689.4	TcCLB.508569.10	TcCLB.507977.9
TcCLB.506529.516	TcCLB.506945.200	TcCLB.508567.30	TcCLB.510877.160	TcCLB.510285.60	TcCLB.504181.10
TcCLB.506931.4	TcCLB.503835.30	TcCLB.508955.10	TcCLB.506811.90	TcCLB.506739.99	TcCLB.507483.20

TcCLB.508137.20	TcCLB.508207.70	TcCLB.509769.60	TcCLB.510603.80	TcCLB.503959.60	TcCLB.511627.110
TcCLB.510431.90	TcCLB.510103.24	TcCLB.504203.40	TcCLB.506657.79	TcCLB.506947.10	TcCLB.503893.70
TcCLB.503881.39	TcCLB.508781.120	TcCLB.504111.20	TcCLB.506869.40	TcCLB.510759.120	TcCLB.508955.20
TcCLB.510769.90	TcCLB.510863.30	TcCLB.507873.30	TcCLB.505843.10	TcCLB.510283.160	TcCLB.509767.230
TcCLB.508975.4	TcCLB.508093.39	TcCLB.508719.40	TcCLB.508717.30	TcCLB.506933.107	TcCLB.509551.70
TcCLB.511145.46	TcCLB.511425.19	TcCLB.508727.50	TcCLB.511733.100	TcCLB.509599.170	TcCLB.504033.130
TcCLB.503909.68	TcCLB.506167.40	TcCLB.506861.30	TcCLB.507625.186	TcCLB.507625.130	TcCLB.504163.110
TcCLB.442297.10	TcCLB.508241.20	TcCLB.508923.10	TcCLB.507143.20	TcCLB.510351.150	TcCLB.506885.410
TcCLB.507521.44	TcCLB.509231.39	TcCLB.510717.40	TcCLB.503531.40	TcCLB.507939.70	TcCLB.506221.110
TcCLB.510429.40	TcCLB.511867.140	TcCLB.507673.20	TcCLB.511417.94	TcCLB.503453.30	TcCLB.506885.340
TcCLB.511437.30	TcCLB.504017.10	TcCLB.506401.304	TcCLB.508989.9	TcCLB.510743.130	TcCLB.507639.30
TcCLB.503425.20	TcCLB.506979.10	TcCLB.511421.170	TcCLB.506885.180	TcCLB.508169.129	TcCLB.503829.10
TcCLB.506943.114	TcCLB.508839.60	TcCLB.508811.40	TcCLB.503697.40	TcCLB.503897.70	TcCLB.510101.430
TcCLB.510743.30	TcCLB.503611.20	TcCLB.509805.230	TcCLB.506943.90	TcCLB.510101.90	TcCLB.503653.60
TcCLB.447255.10	TcCLB.507811.114	TcCLB.510243.40	TcCLB.511621.100	TcCLB.509965.210	TcCLB.510743.70
TcCLB.504929.5	TcCLB.511711.100	TcCLB.511539.18	TcCLB.505843.20	TcCLB.506265.50	TcCLB.507083.40
TcCLB.505025.40	TcCLB.503433.30	TcCLB.506265.150	TcCLB.508041.30	TcCLB.510763.50	TcCLB.503683.11
TcCLB.506773.160	TcCLB.510609.70	TcCLB.504179.20	TcCLB.508949.20	TcCLB.510879.80	TcCLB.511325.50
TcCLB.506755.250	TcCLB.504163.100	TcCLB.510285.74	TcCLB.507677.70	TcCLB.511421.90	TcCLB.506213.50
TcCLB.508389.90	TcCLB.506699.40	TcCLB.511867.40	TcCLB.507837.50	TcCLB.510295.59	TcCLB.510349.20
TcCLB.510101.420	TcCLB.511173.430	TcCLB.503527.4	TcCLB.506399.80	TcCLB.510101.320	TcCLB.511115.30
TcCLB.506519.70	TcCLB.506931.10	TcCLB.510755.129	TcCLB.509807.11	TcCLB.506529.300	TcCLB.509967.140
TcCLB.507927.90	TcCLB.511135.20	TcCLB.503799.4	TcCLB.509141.40	TcCLB.511585.100	TcCLB.507083.70
TcCLB.510855.10	TcCLB.506979.50	TcCLB.411427.30	TcCLB.486465.9	TcCLB.507625.150	TcCLB.507625.120
TcCLB.510029.50	TcCLB.507521.110	TcCLB.506513.190	TcCLB.509279.10	TcCLB.504427.20	TcCLB.503801.30
TcCLB.507105.50	TcCLB.507101.10	TcCLB.506181.50	TcCLB.510879.150	TcCLB.510857.4	TcCLB.507639.90
TcCLB.504421.19	TcCLB.506405.149	TcCLB.507107.40	TcCLB.507929.20	TcCLB.506773.104	TcCLB.506227.210
TcCLB.511145.20	TcCLB.506945.210	TcCLB.510427.10	TcCLB.505931.40	TcCLB.506127.170	TcCLB.506633.90
TcCLB.506205.20	TcCLB.503575.40	TcCLB.511211.70	TcCLB.506401.120	TcCLB.504163.60	TcCLB.508207.64
TcCLB.504167.14	TcCLB.511577.170	TcCLB.508183.10	TcCLB.506925.210	TcCLB.511715.10	TcCLB.511545.120
TcCLB.506287.140	TcCLB.506949.50	TcCLB.510173.130	TcCLB.507099.110	TcCLB.506503.50	TcCLB.510565.70
TcCLB.510181.20	TcCLB.511055.40	TcCLB.511571.20	TcCLB.509965.330	TcCLB.507091.10	TcCLB.510105.40
TcCLB.506863.30	TcCLB.505943.79	TcCLB.511573.40	TcCLB.504005.6	TcCLB.509245.29	TcCLB.506941.150
TcCLB.507611.280	TcCLB.507389.20	TcCLB.509965.190	TcCLB.430605.40	TcCLB.506775.30	TcCLB.503903.60
TcCLB.508163.330	TcCLB.509067.40	TcCLB.508183.20	TcCLB.510029.70	TcCLB.511145.10	TcCLB.506495.6
TcCLB.510523.20	TcCLB.511751.90	TcCLB.504163.50	TcCLB.510737.79	TcCLB.506971.20	TcCLB.506625.90
TcCLB.509233.170	TcCLB.511739.10	TcCLB.506525.104	TcCLB.509499.29	TcCLB.506673.30	TcCLB.506411.39
TcCLB.503527.29	TcCLB.510101.240	TcCLB.407477.40	TcCLB.504163.30	TcCLB.511627.130	TcCLB.511627.90
TcCLB.511353.60	TcCLB.510515.40	TcCLB.503449.10	TcCLB.510751.4	TcCLB.506375.80	TcCLB.511423.160
TcCLB.511123.20	TcCLB.509979.281	TcCLB.509825.40	TcCLB.424195.9	TcCLB.510351.90	TcCLB.509993.31
TcCLB.509233.150	TcCLB.508277.40	TcCLB.511065.36	TcCLB.503833.50	TcCLB.503487.70	TcCLB.509965.429
TcCLB.511419.4	TcCLB.504017.40	TcCLB.506493.100	TcCLB.506567.90	TcCLB.507711.40	TcCLB.506147.130
TcCLB.506925.120	TcCLB.507467.20	TcCLB.503779.70	TcCLB.508089.71	TcCLB.509591.10	TcCLB.506213.80
TcCLB.511407.70	TcCLB.503457.10	TcCLB.509455.80	TcCLB.510293.50	TcCLB.506553.10	TcCLB.510939.20
TcCLB.509165.40	TcCLB.506677.10	TcCLB.511071.171	TcCLB.506529.430	TcCLB.503813.30	TcCLB.509205.60

TcCLB.508389.10	TcCLB.510719.100	TcCLB.511685.50	TcCLB.511911.90	TcCLB.437545.10	TcCLB.511545.80
TcCLB.511685.20	TcCLB.408437.20	TcCLB.504149.124	TcCLB.511531.50	TcCLB.507817.18	TcCLB.507993.380
TcCLB.503779.20	TcCLB.507837.20	TcCLB.505989.110	TcCLB.510409.39	TcCLB.511625.50	TcCLB.510073.70
TcCLB.506797.120	TcCLB.506925.300	TcCLB.510761.18	TcCLB.511145.50	TcCLB.509499.38	TcCLB.506649.80
TcCLB.511911.94	TcCLB.510091.110	TcCLB.508207.110	TcCLB.510755.50	TcCLB.511139.20	TcCLB.511071.190
TcCLB.507165.20	TcCLB.506941.40	TcCLB.504035.70	TcCLB.504089.70	TcCLB.506885.40	TcCLB.508265.10
TcCLB.510535.64	TcCLB.507669.160	TcCLB.507639.10	TcCLB.508943.30	TcCLB.508041.40	TcCLB.508137.30
TcCLB.509147.50	TcCLB.510089.130	TcCLB.504103.10	TcCLB.511407.10	TcCLB.506147.50	TcCLB.506943.69
TcCLB.506563.110	TcCLB.505843.24	TcCLB.506811.120	TcCLB.506857.100	TcCLB.508385.40	TcCLB.509979.14
TcCLB.503779.40					
Nucleus					
TcCLB.506925.60	TcCLB.511211.10	TcCLB.504149.190	TcCLB.506127.120	TcCLB.504017.70	TcCLB.508951.29
TcCLB.419417.19	TcCLB.507039.30	TcCLB.506413.89	TcCLB.506483.60	TcCLB.509595.70	TcCLB.508321.30
TcCLB.508771.55	TcCLB.506961.90	TcCLB.484311.10	TcCLB.508995.10	TcCLB.510431.290	TcCLB.511617.9
TcCLB.506367.50	TcCLB.510513.30	TcCLB.498829.10	TcCLB.509233.220	TcCLB.507477.20	TcCLB.504101.10
TcCLB.509871.70	TcCLB.506529.600	TcCLB.509151.120	TcCLB.507277.30	TcCLB.506605.220	TcCLB.509993.51
TcCLB.503895.10	TcCLB.507527.20	TcCLB.511211.40	TcCLB.507089.220	TcCLB.419703.10	TcCLB.508673.20
TcCLB.511395.70	TcCLB.504277.30	TcCLB.507939.14	TcCLB.510105.130	TcCLB.503579.40	TcCLB.439803.9
TcCLB.506973.20	TcCLB.510759.10	TcCLB.510857.30	TcCLB.510323.120	TcCLB.508557.20	TcCLB.507041.132
TcCLB.506715.20	TcCLB.507093.130	TcCLB.504425.50	TcCLB.506637.20	TcCLB.507559.20	TcCLB.504025.19
TcCLB.506457.149	TcCLB.503697.30	TcCLB.506801.70	TcCLB.506673.50	TcCLB.509817.40	TcCLB.506713.40
TcCLB.506289.140	TcCLB.511423.170	TcCLB.507669.50	TcCLB.511401.80	TcCLB.508681.31	TcCLB.511873.10
TcCLB.506223.30	TcCLB.511533.10	TcCLB.506765.43	TcCLB.504867.110	TcCLB.511733.120	TcCLB.511055.50
TcCLB.506181.120	TcCLB.510613.30	TcCLB.504147.260	TcCLB.508563.20	TcCLB.509011.80	TcCLB.506591.130
TcCLB.507711.235	TcCLB.511211.230	TcCLB.506891.20	TcCLB.507011.120	TcCLB.503749.30	TcCLB.503637.20
TcCLB.504049.20	TcCLB.507669.120	TcCLB.503885.50	TcCLB.510015.10	TcCLB.510517.80	TcCLB.509893.20
TcCLB.508207.50	TcCLB.507603.160	TcCLB.506949.20	TcCLB.510535.10	TcCLB.507089.30	TcCLB.503955.50
TcCLB.507839.40	TcCLB.511211.220	TcCLB.510877.20	TcCLB.510307.180	TcCLB.508381.20	TcCLB.509893.10
TcCLB.510533.190	TcCLB.510719.50	TcCLB.508837.100	TcCLB.503479.20	TcCLB.506739.209	TcCLB.511695.10
TcCLB.507905.39	TcCLB.508103.30	TcCLB.506203.90	TcCLB.508081.110	TcCLB.510911.45	TcCLB.509773.9
TcCLB.507809.170	TcCLB.506265.190	TcCLB.509951.10	TcCLB.507925.30	TcCLB.507809.70	TcCLB.511401.24
TcCLB.509153.40	TcCLB.509203.60	TcCLB.508839.30	TcCLB.511107.20	TcCLB.511129.80	TcCLB.508809.30
TcCLB.506777.10	TcCLB.506457.70	TcCLB.510297.40	TcCLB.511443.10	TcCLB.511753.150	TcCLB.507429.20
TcCLB.511215.70	TcCLB.507041.90	TcCLB.510593.10	TcCLB.507559.30	TcCLB.506743.130	TcCLB.511693.30
TcCLB.509349.20	TcCLB.506777.100	TcCLB.511441.20	TcCLB.503671.20	TcCLB.506885.210	TcCLB.509959.160
TcCLB.503685.20	TcCLB.503493.20	TcCLB.507641.30	TcCLB.506859.110	TcCLB.511589.260	TcCLB.506605.170
TcCLB.505943.50	TcCLB.506737.140	TcCLB.505807.209	TcCLB.510493.4	TcCLB.506409.110	TcCLB.510773.30
TcCLB.509393.50	TcCLB.506289.74	TcCLB.511747.30	TcCLB.511689.10	TcCLB.506625.140	TcCLB.509967.150
TcCLB.510075.30	TcCLB.503755.30	TcCLB.511751.250	TcCLB.504101.50	TcCLB.508143.80	TcCLB.473633.9
TcCLB.503907.10	TcCLB.508949.40	TcCLB.510181.130	TcCLB.507875.110	TcCLB.506375.60	TcCLB.510289.99
TcCLB.503651.4	TcCLB.508515.40	TcCLB.511189.10	TcCLB.506859.230	TcCLB.510717.20	TcCLB.504053.10
TcCLB.511685.40	TcCLB.511189.90	TcCLB.509151.90	TcCLB.510275.272	TcCLB.511189.50	TcCLB.506961.40
TcCLB.507091.150	TcCLB.506287.130	TcCLB.511735.40	TcCLB.506181.20	TcCLB.509885.20	TcCLB.418069.10
TcCLB.509809.39	TcCLB.511801.40	TcCLB.507011.240	TcCLB.503855.20	TcCLB.508325.20	TcCLB.508837.80
TcCLB.506465.5	TcCLB.505843.30	TcCLB.503949.40	TcCLB.509267.70	TcCLB.469785.5	TcCLB.511197.40

TcCLB.504213.90	TcCLB.511751.110	TcCLB.509267.10	TcCLB.506259.30	TcCLB.510253.29	TcCLB.506357.10
TcCLB.510533.60	TcCLB.505365.80	TcCLB.509153.50	TcCLB.510743.50	TcCLB.504159.100	TcCLB.506859.100
TcCLB.504117.9	TcCLB.507641.40	TcCLB.511529.220	TcCLB.507677.4	TcCLB.503957.25	TcCLB.507747.130
TcCLB.509445.20	TcCLB.506563.190	TcCLB.510091.7	TcCLB.508647.260	TcCLB.507923.30	TcCLB.508157.40
TcCLB.511421.10	TcCLB.510863.9	TcCLB.511121.11	TcCLB.507993.350	TcCLB.507641.290	TcCLB.509631.140
TcCLB.503579.10	TcCLB.509769.70	TcCLB.509127.59	TcCLB.507041.76	TcCLB.506401.170	TcCLB.504129.50
TcCLB.508317.20	TcCLB.511337.30	TcCLB.511753.110	TcCLB.506567.140	TcCLB.503801.60	TcCLB.511173.160
TcCLB.507603.240	TcCLB.509135.9	TcCLB.466823.10	TcCLB.506967.70	TcCLB.509965.70	TcCLB.508347.159
TcCLB.510175.125	TcCLB.508179.30	TcCLB.508163.120	TcCLB.506455.30	TcCLB.445281.9	TcCLB.503967.4
TcCLB.508973.20	TcCLB.511417.120	TcCLB.511621.210	TcCLB.510955.50	TcCLB.506531.60	TcCLB.510531.109
TcCLB.507841.14	TcCLB.408573.9	TcCLB.508841.10	TcCLB.503897.100	TcCLB.507737.10	TcCLB.508777.190
TcCLB.507603.120	TcCLB.511127.369	TcCLB.510025.180	TcCLB.509073.100	TcCLB.503955.40	TcCLB.511587.120
TcCLB.509139.20	TcCLB.508377.9	TcCLB.508989.60	TcCLB.511545.100	TcCLB.508239.40	TcCLB.508607.50
TcCLB.510535.40	TcCLB.506227.190	TcCLB.510857.50	TcCLB.511043.45	TcCLB.404711.10	TcCLB.422041.9
TcCLB.504033.110	TcCLB.510023.30	TcCLB.507795.80	TcCLB.506973.140	TcCLB.508209.100	TcCLB.506179.10
TcCLB.511537.24	TcCLB.505983.40	TcCLB.507041.50	TcCLB.506221.129	TcCLB.503981.90	TcCLB.511395.120
TcCLB.507641.220	TcCLB.509157.40	TcCLB.504883.70	TcCLB.510773.70	TcCLB.510381.50	TcCLB.510349.80
TcCLB.503915.9	TcCLB.506227.250	TcCLB.503885.80	TcCLB.504427.230	TcCLB.507465.10	TcCLB.507179.14
TcCLB.510769.70	TcCLB.510029.10	TcCLB.506743.220	TcCLB.509965.240	TcCLB.503867.5	TcCLB.511585.310
TcCLB.503749.10	TcCLB.508387.110	TcCLB.511199.20	TcCLB.510565.50	TcCLB.504227.10	TcCLB.506375.20
TcCLB.511715.40	TcCLB.511585.30	TcCLB.509735.30	TcCLB.511407.80	TcCLB.511599.90	TcCLB.510379.20
TcCLB.507683.10	TcCLB.509265.50	TcCLB.503643.39	TcCLB.510351.140	TcCLB.508375.70	TcCLB.507041.80
TcCLB.510101.160	TcCLB.509247.10	TcCLB.510603.20	TcCLB.504427.10	TcCLB.508727.30	TcCLB.510535.70
TcCLB.507389.30	TcCLB.506847.50	TcCLB.507257.180	TcCLB.510141.20	TcCLB.511361.4	TcCLB.506499.130
TcCLB.503479.70	TcCLB.510213.39	TcCLB.511619.5	TcCLB.506529.160	TcCLB.507953.130	TcCLB.506375.70
TcCLB.507559.40	TcCLB.511421.30	TcCLB.507677.160	TcCLB.507105.120	TcCLB.508787.14	TcCLB.511253.10
TcCLB.511529.170	TcCLB.511593.51	TcCLB.506661.50	TcCLB.508609.20	TcCLB.505807.160	TcCLB.420293.31
TcCLB.509153.139	TcCLB.508805.30	TcCLB.503885.20	TcCLB.511751.163	TcCLB.508123.10	TcCLB.511435.49
TcCLB.506963.40	TcCLB.511871.130	TcCLB.508317.129	TcCLB.511733.30	TcCLB.438059.10	TcCLB.511589.170
TcCLB.506755.60	TcCLB.510123.30	TcCLB.506227.120	TcCLB.507817.70	TcCLB.511711.60	TcCLB.504769.190
TcCLB.508647.170	TcCLB.508207.194	TcCLB.506251.40	TcCLB.507529.40	TcCLB.511911.81	TcCLB.507933.10
TcCLB.506859.150	TcCLB.510491.50	TcCLB.506025.70	TcCLB.510105.120	TcCLB.506749.10	TcCLB.504099.30
TcCLB.510345.20	TcCLB.507841.30	TcCLB.504159.10	TcCLB.510767.30	TcCLB.510769.30	TcCLB.511587.90
TcCLB.510769.80	TcCLB.510855.4	TcCLB.509763.20	TcCLB.504431.40	TcCLB.511751.140	TcCLB.508991.10
TcCLB.508181.90	TcCLB.506179.59	TcCLB.510021.180	TcCLB.511609.10	TcCLB.504835.29	TcCLB.507631.10
TcCLB.504625.40	TcCLB.503601.10	TcCLB.511423.50	TcCLB.506605.100	TcCLB.509697.15	TcCLB.509979.320
TcCLB.506829.50	TcCLB.506683.240	TcCLB.509669.110	TcCLB.504133.15	TcCLB.480997.10	TcCLB.507447.19
TcCLB.504099.50	TcCLB.510005.20	TcCLB.507711.220	TcCLB.511529.50	TcCLB.508973.80	TcCLB.503453.40
TcCLB.507099.70	TcCLB.511589.209	TcCLB.507009.80	TcCLB.510603.119	TcCLB.511051.10	TcCLB.506251.50
TcCLB.507527.35	TcCLB.509819.10	TcCLB.506203.10	TcCLB.506579.50	TcCLB.507537.40	TcCLB.507941.100
TcCLB.503981.80	TcCLB.511127.50	TcCLB.404975.30	TcCLB.511745.50	TcCLB.506227.240	TcCLB.504343.30
TcCLB.510989.20	TcCLB.511625.130	TcCLB.506529.620	TcCLB.507091.4	TcCLB.510349.70	TcCLB.510591.20
TcCLB.511445.160	TcCLB.511395.50	TcCLB.507485.50	TcCLB.508047.60	TcCLB.510963.60	TcCLB.503717.60
TcCLB.506531.9	TcCLB.507211.30	TcCLB.510073.40	TcCLB.506563.4	TcCLB.510737.119	TcCLB.509927.10
TcCLB.504041.4	TcCLB.511365.4	TcCLB.506625.130	TcCLB.509007.20	TcCLB.507805.30	TcCLB.509835.30

TcCLB.511201.19	TcCLB.506517.149	TcCLB.509443.60	TcCLB.511041.30	TcCLB.507257.130	TcCLB.510773.9
TcCLB.506529.100	TcCLB.506925.270	TcCLB.511369.5	TcCLB.503595.19	TcCLB.506933.89	TcCLB.511181.110
TcCLB.507547.50	TcCLB.511215.110	TcCLB.507925.10	TcCLB.510599.40	TcCLB.507163.70	TcCLB.507627.20
TcCLB.510747.80	TcCLB.510955.60	TcCLB.511047.70	TcCLB.506493.20	TcCLB.506289.160	TcCLB.509099.5
TcCLB.503487.20	TcCLB.508569.130	TcCLB.506625.160	TcCLB.508783.10	TcCLB.507611.131	TcCLB.510027.10
TcCLB.509157.230	TcCLB.506527.4	TcCLB.504203.29	TcCLB.506529.669	TcCLB.508489.10	TcCLB.506597.40
TcCLB.425311.9	TcCLB.506179.20	TcCLB.508569.100	TcCLB.506345.90	TcCLB.506865.20	TcCLB.506543.50
TcCLB.506647.20	TcCLB.434259.10	TcCLB.508921.40	TcCLB.510755.40	TcCLB.510345.9	TcCLB.506949.10
TcCLB.506779.180	TcCLB.506757.50	TcCLB.507941.10	TcCLB.410961.10	TcCLB.510853.30	TcCLB.506265.170
TcCLB.509537.80	TcCLB.503755.60	TcCLB.508831.150	TcCLB.510877.40	TcCLB.509267.6	TcCLB.511575.120
TcCLB.508717.60	TcCLB.509233.239	TcCLB.505073.10	TcCLB.507213.40	TcCLB.507083.20	TcCLB.511211.90
TcCLB.506599.10	TcCLB.511441.10	TcCLB.508161.23	TcCLB.506021.20	TcCLB.511415.40	TcCLB.441241.10
TcCLB.506865.79	TcCLB.509133.10	TcCLB.507477.10	TcCLB.507283.10	TcCLB.507105.132	TcCLB.445635.10
TcCLB.506579.100	TcCLB.506467.50	TcCLB.438923.10	TcCLB.509011.10	TcCLB.507551.9	TcCLB.510261.21
TcCLB.507641.90	TcCLB.506801.20	TcCLB.509631.60	TcCLB.507121.40	TcCLB.509767.130	TcCLB.506625.35
TcCLB.506285.80	TcCLB.507045.5	TcCLB.508097.10	TcCLB.504009.4	TcCLB.506799.20	TcCLB.428999.20
TcCLB.510175.110	TcCLB.434945.10	TcCLB.507237.150	TcCLB.511557.50	TcCLB.507037.89	TcCLB.509011.30
TcCLB.507467.10	TcCLB.510279.276	TcCLB.506197.10	TcCLB.506717.20	TcCLB.506737.151	TcCLB.511587.75
TcCLB.507737.40	TcCLB.510055.30	TcCLB.510275.405	TcCLB.504127.20	TcCLB.507559.110	TcCLB.509269.4
TcCLB.504277.20	TcCLB.510279.160	TcCLB.442427.10	TcCLB.506895.80	TcCLB.409139.9	TcCLB.504227.5
TcCLB.504171.20	TcCLB.510847.10	TcCLB.510377.120	TcCLB.510269.73	TcCLB.511043.60	TcCLB.507035.30
TcCLB.507905.30	TcCLB.509349.30	TcCLB.504055.130	TcCLB.507275.61	TcCLB.504421.10	TcCLB.510449.10
TcCLB.510609.60	TcCLB.510275.120	TcCLB.506499.200	TcCLB.510565.170	TcCLB.509667.31	TcCLB.510021.120
TcCLB.503801.79	TcCLB.509925.5	TcCLB.510283.20	TcCLB.511253.31	TcCLB.508325.30	TcCLB.506671.70
TcCLB.506717.80	TcCLB.505965.10	TcCLB.510283.30	TcCLB.509631.100	TcCLB.510533.210	TcCLB.506487.24
TcCLB.506717.100	TcCLB.506973.10	TcCLB.510607.29	TcCLB.511349.80	TcCLB.508163.360	TcCLB.507719.9
TcCLB.509197.10	TcCLB.511181.160	TcCLB.508283.69	TcCLB.511315.44	TcCLB.510557.73	TcCLB.506129.30
TcCLB.511705.10	TcCLB.509195.10	TcCLB.506653.40	TcCLB.508605.40	TcCLB.504703.20	TcCLB.510307.250
TcCLB.506047.13	TcCLB.507869.20	TcCLB.509265.90	TcCLB.506375.35	TcCLB.508321.21	TcCLB.508325.25
TcCLB.506673.69	TcCLB.507527.40	TcCLB.509635.9	TcCLB.509925.70	TcCLB.510773.40	TcCLB.506501.300
TcCLB.506811.160	TcCLB.506375.50	TcCLB.508439.40	TcCLB.509123.10	TcCLB.511611.11	TcCLB.509993.15
TcCLB.507041.100	TcCLB.511421.224	TcCLB.403153.10	TcCLB.511691.20	TcCLB.506763.210	TcCLB.508209.10
TcCLB.508087.10	TcCLB.511421.230	TcCLB.404001.20	TcCLB.506375.100	TcCLB.506519.40	TcCLB.510755.98
TcCLB.504117.50	TcCLB.510877.10	TcCLB.509103.4	TcCLB.504423.40	TcCLB.511253.20	TcCLB.510073.10
TcCLB.479517.30	TcCLB.507275.70	TcCLB.508541.140	TcCLB.459793.20	TcCLB.509871.80	TcCLB.506951.80
TcCLB.509919.9	TcCLB.508805.120	TcCLB.511621.200	TcCLB.510313.10	TcCLB.510913.10	TcCLB.504277.11
TcCLB.508775.60	TcCLB.511217.209	TcCLB.506399.20	TcCLB.506609.15	TcCLB.506961.35	TcCLB.507757.10
TcCLB.510317.34	TcCLB.509667.10	TcCLB.509805.260	TcCLB.504039.130	TcCLB.506623.30	TcCLB.506591.120
TcCLB.510989.40	TcCLB.511585.230	TcCLB.511621.80	TcCLB.511587.40	TcCLB.510451.10	TcCLB.504769.160
TcCLB.506345.70	TcCLB.511421.217	TcCLB.508163.371	TcCLB.504055.100	TcCLB.510815.9	TcCLB.507237.320
TcCLB.508681.20	TcCLB.506591.81	TcCLB.509205.90	TcCLB.506467.40	TcCLB.509847.50	
TcCLB.509449.10	TcCLB.511587.33	TcCLB.503777.4	TcCLB.505941.20	TcCLB.507179.20	
TcCLB.507975.70	TcCLB.511173.470	TcCLB.508285.5	TcCLB.466965.10	TcCLB.510247.10	
TcCLB.508605.20	TcCLB.511339.10	TcCLB.507979.40	TcCLB.511401.95	TcCLB.509871.39	

Table S5. Correlation analysis between protein abundances and transcripts enriched in the cytoplasmic and nuclear fractions of *T. cruzi* epimastigotes

Fraction / Set	Transcriptome*	Proteome**	Spearman	Significance
CET	444	440	0.5847	<0.0001
NET	738	44	0.2327	0.1854

* Number of genes with more than 10 reads per transcript in the corresponding fraction or set

** Number of transcripts for each transcriptome set whose encoded proteins were quantitatively determined by de Godoy *et al.*

5. Búsqueda de señales involucradas en la compartimentación núcleo-citoplasma de transcriptos en epimastigotas de *T. cruzi*

La distribución diferencial de mensajeros tiene un amplio efecto sobre el proteoma de las células y la comprensión más profunda de todos los elementos en *cis* que gobiernan la retención y/o la exportación nuclear nos ayudará a entender cómo estos ARNs son exportados o retenidos en el núcleo.

Para intentar comprender el/los mecanismos implícitos en la distribución diferencial de mensajeros entre el núcleo y el citoplasma realizamos la búsqueda de motivos lineales enriquecidos en los transcritos diferencialmente localizados entre el citoplasma y el núcleo (CET y NET) utilizando el programa DREME del paquete *meme-suite*. La identificación de motivos enriquecidos fue realizada utilizando como controles, el resto de los transcritos del parásito (NDT = transcriptoma de *T. cruzi* una vez excluidos los transcritos integrantes de las bases CET y NET), así como los controles cruzados CET respecto a NET y viceversa. Debido a un posible sesgo de los resultados, los transcritos pertenecientes a familias multigénicas fueron excluidos de la búsqueda. Mediante este análisis pudimos determinar motivos enriquecidos en los 5' y 3'UTRs de ambas bases, (tabla 1 y 2).

Tabla 1. Motivos lineales significativamente enriquecidos en las 5' y 3' UTRs de los transcritos diferencialmente localizados en el citoplasma (CET). Se presentan los motivos enriquecidos respecto a la base NET y NDT, el logo generado a partir de las secuencias y el *E-value*.

Transcriptos enriquecidos en citoplasma											
5'UTR						3'UTR					
Transcriptos enriquecidos en núcleo			Transcriptos no diferenciales			Transcriptos enriquecidos en núcleo			Transcriptos no diferenciales		
Motivo	Logo	<i>E-value</i>	Motivo	Logo	<i>E-value</i>	Motivo	Logo	<i>E-value</i>	Motivo	Logo	<i>E-value</i>
TAGAAD		1.5e-017	TAGAAR		2.2e-009	AAAAANA		1.3e-016	AAAA		5.8e-006
TAHNA		2.1e-014	WACAAC		1.6e-003	TAWWA		9.6e-013	ARTA		2.8e-004
WTYAA		1.1e-010	TATMTT		1.5e-002	AWAFA		6.1e-012	GRAA		6.4e-003
AAAWR		2.6e-007	CATM		1.5e-002	WTTTAT		5.4e-007	TWTTTTA		2.0e-002
TAMM		3.4e-003	CKTA		3.7e-002	AAAG		2.2e-006	TTTATTM		1.2e-002
SAACAMC		3.1e-003				TARR		1.1e-006			
TATCEI		4.1e-002				AACRAMAA		2.5e-004			
						TTTTTKTT		8.8e-003			
						ATWT		3.3e-003			
						GGGRGA		1.8e-002			

Tabla 2. Motivos lineales significativamente enriquecidos en los 5' y 3' UTRs de los transcritos diferencialmente localizados en el núcleo (NET). Se presentan los motivos enriquecidos respecto a la base CET y NDT, el logo generado a partir de las secuencias y el *E-value*.

Transcritos enriquecidos en núcleo											
5'UTR						3'UTR					
Transcritos enriquecidos en citoplasma			Transcritos no diferenciales			Transcritos enriquecidos en citoplasma			Transcritos no diferenciales		
Motivo	Logo	<i>E-value</i>	Motivo	Logo	<i>E-value</i>	Motivo	Logo	<i>E-value</i>	Motivo	Logo	<i>E-value</i>
GCTGMY		8.0e-011	GNGTGGC		4.6e-007	CSNCC		4.2e-008	GCGSCS		4.8e-011
GMSWGG		4.0e-007	TGACGBC		5.8e-005	SYGCTG		2.6e-007	YGRCCG		2.0e-006
KTGQYQ		7.9e-006	GSVTGCT		2.8e-004	CYSCG		3.8e-005	CCYGCCT		1.1e-005
GCGMS		6.3e-005	TTTGGTS		5.8e-004	RKQGC		8.1e-005	CGCTGGK		1.1e-005
GCKSTG		2.6e-003	GMGGC		1.4e-003	CASCAS		5.9e-004	SGYGG		4.6e-005
WQGG		2.1e-002	GCYGTCC		2.1e-003	CSTGS		4.0e-003	GNTCTGTG		6.1e-004
			CGCTGCBT		3.9e-003				CBCCG		7.0e-004
			CCMTGAGG		4.4e-003				SSGTGTC		9.9e-004
			GCDGGCG		9.1e-003				GCCYGGG		2.6e-003
			MATCCAC		1.1e-002				ATTKAGGC		4.1e-003
			TCCAMTT		1.1e-002				CRGGTC		5.0e-003
			ROCGGATT		1.8e-002				CRGTGC		6.5e-003
			GTGAGTY		2.9e-002				CTGTCT		9.0e-003
			STGGACG		3.6e-002				TKGARCC		1.1e-002
			TGCGKITG		4.9e-002				CDCCAG		3.3e-002
									CTGCGG		3.3e-002
									CGCTGC		3.3e-002
									SSGTATC		4.6e-002

A partir de los motivos lineales significativamente enriquecidos en las regiones transcritas y traducidas de los genes de las bases CET y NET se buscaron las proteínas que reconocen motivos similares a los encontrados en la base de datos publicada por Ray *et al.* (Ray *et al.*, 2013), empleando el programa Tom-Tom del paquete *meme-suite*. Esto permitió la identificación de varias proteínas de diferentes organismos. Para determinar las proteínas de *T. cruzi* que potencialmente reconocerían estos motivos se hizo una búsqueda de ortólogos. A continuación, presentamos las características más sobresalientes de estas proteínas.

5.1. Proteínas de *T. cruzi* que reconocerían los motivos identificados en los transcritos enriquecidos en el citoplasma

En las regiones 5' y 3'UTR de los genes de la base CET identificamos 27 motivos, de los cuales, solamente 11 mostraron coincidencias a distintos niveles con motivos reconocidos por proteínas de unión al ARN (Ray *et al.*, 2013) para las que se pudo definir ortólogo de *T. cruzi* (tabla 3).

Tabla 3. Proteínas que reconocen motivos lineales enriquecidos en las UTRs de los transcritos diferencialmente localizados en citoplasma. Se muestra el ID de la proteína, la especie a la que pertenece y la descripción; el ID del gen ortólogo en *T. cruzi*; la descripción del gen y los motivos lineales, blancos potenciales.

ID	Organismo	Descripción	ID <i>T. cruzi</i>	Descripción	Motivos reconocidos
CEF98978.1	<i>Ostreococcus tauri</i>	Nucleotide-binding, α - β plait protein	TcCLB.506885.70	Polyadenylate-binding protein 1 (PABP1)	SAACAMC, TAMAAR, AACAAAMAA, AAAAAANA, TWTTTTAA, TTTATTTM
LmjF.35.5040	<i>L. major</i>	polyadenylate-binding protein 1			
Tb927.9.8740	<i>T. brucei</i>	Double RNA binding domain protein 3	TcCLB.506649.80	Double RNA binding domain protein 3 (DRBP3)	TATCBT
NP_705947.3	<i>Danio rerio</i>	splicing factor 3b subunit 4	TcCLB.510143.80	RNA-binding protein	AAAG
NP_502566.1	<i>Caenorhabditis elegans</i>	Zinc finger protein mex-5	TcCLB.504929.5	zinc finger CCCH domain containing protein 11 (ZFN11)	ARTA
EIE75345.1	<i>Rhizopus oryzae</i>	hypothetical	TcCLB.510149.140	Triple RNA binding domain protein 3 (TRRM3)	GGGRGA, GRAA
XP_001347353.1	<i>Plasmodium falciparum</i>	RNA binding protein	TcCLB.510755.120	RNA binding protein	GRAA

Tabla 4. Motivos lineales enriquecidos en las UTRs de 394 transcritos diferencialmente localizados en el citoplasma que constituyen potenciales blancos de proteínas de *T. cruzi*.

DREME +, indica el número de secuencias que generaron los motivos; - indica el número de secuencias que poseen el motivo en el grupo control (NET 569 genes, NDT 4379 genes). FiMo indica el resultado de la búsqueda de los motivos generados mediante DREME en la base de datos que lo generó. Conteo, indica el número de veces que aparece el motivo en la base. Se indica también el número de genes que presentan el motivo.

Muestra	Control	DREME			FiMo	
		Motivo	+	-	Conteo	N° Genes
5'UTR CET	NET	SAACAMC	95	62	97	56
		TATCBT	84	56	102	90
	NDT	TAMAAR	199	1370	167	130
3'UTR CET	NET	AAAAAANA	265	207	250	44
		AAAG	297	312	1326	349
		AACAAMAA	63	27	250	95
		GGGRGA	162	146	350	208
	NDT	ARTA	352	3399	988	331
		GRAA	366	3663	1324	349
		TWTTTTTA	161	1248	163	240
		TTTATTTM	108	728	308	173

La proteína de unión a poli(A) 1 (TcCLB.506885.70, PABP1) reconocería los motivos "SAACAMC, TAMAAR, AACAAMAA, AAAAAANA, TWTTTTTA, TTTATTTM" presentes en 392 de los 394 mensajeros CET (tabla 4). Es ortóloga a las proteínas PABP1 de *Ostreococcus tauri* y *L. major* (tabla 3). Estas PABPs presentan 4 dominios RRM típicos hacia el extremo N-terminal. Los primeros dos dominios RRM son suficientes para la unión específica a poli(A) y se propone que el tercer y cuarto dominio RRM podrían reconocer secuencias no poli(A). La región C terminal posee un motivo denominado PABC (*Poly(A)-binding protein C-terminal domain*) que media la interacción de la nucleasa PAN con la poli(A) (da Costa Lima *et al.*, 2010). En *Trypanosoma* tendría un rol en los procesos de estabilización postranscripcional de mensajeros (Coller y Wickens, 2007). El análisis de los datos publicados de traductómica, y proteómica (de Godoy *et al.*, 2012; Smircich *et al.*, 2015) muestra que la proteína se expresa en el estadio epimastigota, y los datos de ciclo de microarreglos del vida muestran que se expresaría también en

altos niveles en el estadio amastigota (Minning *et al.*, 2009), siendo ambos estadios replicativos, y traduccionalmente activos (tabla 5). Los genes blanco de la proteína (casi el 100% de CET), en promedio, se expresan más en el estadio epimastigota (línea rosada en datos de genes blanco del transcriptoma del ciclo de vida, tabla 5). La categorización por ontología indica que la proteína tendría una localización tanto nuclear como citoplasmática. Los datos de TrypTag, muestran una localización citoplasmática de la proteína, aunque no se puede descartar su localización nuclear (Dean *et al.*, 2017). Además, en *T. brucei*, el mensajero que codifica para esta proteína es tres veces más estable en el estadio procíclico (del insecto) que en el sanguíneo (Fadda *et al.*, 2014). En conjunto, estos datos sugieren que la PABP1 podría establecer interacciones con los mensajeros enriquecidos en CET, promoviendo un aumento de la estabilidad de los mismos, que justificaría su enriquecimiento en el citoplasma. Se requerirían estudios funcionales para verificar esta inferencia.

Tabla 5. Características de las proteínas que reconocerían motivos enriquecidos en transcritos con localización citoplasmática. Para cada gen se indica el ID, el FC entre los estadios epimastigota y tripomastigota metacíclico del ARNm poliadenilado, el ARNm activamente traducido y la proteína. Se grafican los datos de expresión por microarreglos durante el ciclo de vida del gen y de los genes blancos Datos tomados de: *Smircich *et al.* (Smircich *et al.*, 2015), **Minning *et al.* (Minning *et al.*, 2009), ***de Godoy *et al.* (de Godoy *et al.*, 2012).

ID	Transcriptoma (RNA-Seq)*		Transcriptoma (Microarreglos)**				Proteoma*** log2FC(MT/E)
	log2FC(MT/E) polyA	log2FC(MT/E) RFP	Gen		Blancos		
			Ama	Tripo	Epi	Meta	
TcCLB.506885.70	-0,79	-1,01					-1,41
TcCLB.506649.80	-0,82	-2,08					-1,37
TcCLB.510143.80	-1,02	-3,59					N/A
TcCLB.504929.5	1,50	-0,67					N/A
TcCLB.510149.140	-0,88	-3,99					N/A
TcCLB.510755.120	-0,64	-3,20					-0,51

La proteína TcCLB.506649.80 (DRBD3) reconocería el motivo "TATCBT" presente en 90 mensajeros de CET (tabla 4) y es ortóloga a la Proteína de unión a ARN doble hebra 3 de *T. brucei* (Tb927.9.8740) (tabla 3). Posee dos dominios RRM dentro del dominio homólogo al de unión a trectos de polipirimidinas. Los datos de expresión del gen indican que la proteína se expresa en los estadios replicativos del parásito y a su vez, su transcripto está enriquecido en el citoplasma. Sus genes blanco, tienen en promedio, un perfil de expresión similar al del transcripto (tabla 5). La categorización por ontología indica una localización dual en núcleo y citoplasma. Se ha reportado que la proteína ortóloga en *T. brucei* reconoce el motivo consenso "TTCCCTCT", uniéndose principalmente a la región 3'UTR de mensajeros que codifican para proteínas involucradas en traducción y oxidación de glucosa promoviendo su traducción (Das *et*

al., 2015). Un reporte anterior indica que la proteína se expresa en los dos estadios del parásito siendo esencial en ambos, estando localizada principalmente en el citoplasma, aunque no homogéneamente distribuida, e interaccionando con mensajeros que codifican para proteínas asociadas a membrana (Estevez, 2008). Estos antecedentes apoyan la posibilidad del involucramiento de esta proteína en la compartimentación de un grupo de genes enriquecidos en el citoplasma.

Otra de las proteínas identificadas mediante la búsqueda por TOM-TOM es la subunidad 4 del factor de *splicing* 3b de *Danio rerio*, esta, reconocería motivos similares "AAAG" presentes en 349 mensajeros de la base CET (tabla 4). El ortólogo de *T. cruzi* es la proteína de unión al ARN TcCLB.510143.80 (tabla 3). Esta proteína tiene dos motivos RRM. Los datos de expresión indican que el mensajero se expresa en los estadios replicativos del parásito, y sus genes blanco, tienen, en promedio, un perfil de expresión similar (tabla 5). Los datos del proyecto TrypTag muestran que la proteína homóloga de *T. brucei* (Tb927.3.5280, *spliceosome-associated protein 49, putative*, DRBD1) presenta localización nuclear (Dean *et al.*, 2017). Esta proteína interacciona con las proteínas del spliceosoma p14, SAP130, SAP145 y SAP155 y su *knockdown* produce defectos en el *splicing* (Manful *et al.*, 2009). Los datos sugieren que esta proteína tiene un rol en el *splicing* de los pre-mensajeros proceso que podría afectar la compartimentación núcleo-citoplasma, aunque no podemos descartar que posea también un rol directo en la compartimentación de mensajeros.

El análisis de TOM-TOM indica que la proteína con dedos de zinc mex-5 de *Caenorhabditis elegans* reconocería el motivo "ARTA". Este motivo, está presente en 331 mensajeros de CET. La proteína *zinc finger CCCH domain containing protein 11* (ZFN11, TcCLB.504929.5) es la ortóloga en *T. cruzi*. El análisis de ontología la categoriza como una proteína de involucrada en la estabilización de mensajeros con localización citoplasmática. El ARN mensajero que codifica para esta proteína se expresa mayormente en el estadio tripomastigota metacíclico, sin embargo, la proteína se traduciría más en epimastigotas (tabla 5). Carecemos de datos de proteómica que contribuyan a resolver esta aparente contradicción. No obstante, es interesante destacar que el mensajero se encuentra enriquecido en citoplasma (dentro de los genes CET) en epimastigotas de *T. cruzi*, por lo que podría ser expresado en este estadio del parásito. Los genes blanco,

tienen, en promedio, un perfil de expresión similar al de los mensajeros CET (tabla 5). En conjunto, estos resultados no son suficientes para respaldar un posible rol en la compartimentación de ARN en epimastigotas de *T. cruzi*.

La proteína con dominio de unión a triple ARN 3, (TcCLB.510149.140, TRRM3) es ortóloga a una proteína hipotética de *Rhizopus oryzae* (tabla 6) y reconocería los motivos "GGGRGA, GRAA" de 383 mensajeros de CET (tabla 7). Presenta tres dominios RRM y un dominio trans membrana. La proteína posee una señal de localización nuclear y el TrypTag muestra que el ortólogo de *T. brucei* (Tb927.3.3670) se localiza en el nucléolo (Dean *et al.*, 2017). Los datos de expresión por microarreglos (ver tabla 5) muestran que el mensajero se expresa mayormente en los estadios epimastigota y tripomastigota metacíclico, sin embargo, por RNA-Seq parece estar en mayor nivel y traducirse más en epimastigotas (ver tabla 5). No contamos con datos de proteómica para esta proteína. Por otro lado, los niveles de los mensajeros blanco no parecen acompañar el perfil de expresión del gen. Esta proteína no parece ser por sí misma, responsable de la estabilización de los mensajeros, aunque, podría participar como co-factor.

Por último, la proteína de unión al ARN TcCLB.510755.120 es ortóloga a una proteína de unión al ARN de *Plasmodium falciparum* (tabla 6) y reconocería el motivo "GRAA" en 349 mensajeros de CET (tabla 7). La proteína posee dos dominios RRM. El mensajero se expresa mayormente en los estadios amastigota y tripomastigota sanguíneo (tabla 5), aunque el carácter comparativo del estudio de ciclo de vida de Minning *et al.* 2009 no permite aseverar que el mensajero no se exprese en los otros estadios. De hecho, los datos de RNA-Seq muestran que el mensajero se expresa y se traduce más en epimastigotas que en tripomastigotas metacíclicos; el proteoma acompaña este resultado. Por otro lado, los niveles de los mensajeros blanco no parecen acompañar el perfil de expresión del gen. La proteína homóloga de *T. brucei* Tb927.9.13990 (*RNA-binding protein*, putative; DRBD2) disminuye la traducción y/o la estabilidad de mensajeros (Erben *et al.*, 2014) y se localiza principalmente en el citoplasma (Dean *et al.*, 2017). En conjunto estos datos, no permiten inferir un rol en la compartimentación núcleo citoplasma para esta proteína.

Considerando los antecedentes de las proteínas que podrían reconocer los motivos enriquecidos en CET, la proteína PABP1, cuyo rol es la estabilización de mensajeros, podría ser responsable de los altos niveles de estos ARNs en el citoplasma.

5.2. Proteínas de *T. cruzi* que reconocerían los motivos identificados en los transcritos enriquecidos en el núcleo.

En las regiones 5' y 3'UTR de los genes de la base NET identificamos 45 motivos, de los cuales, 30 mostraron coincidencias a distintos niveles con motivos reconocidos por proteínas de unión al ARN (Ray *et al.*, 2013) para las que se pudo definir ortólogo de *T. cruzi* (tabla 6).

La proteína de unión al ARN TcCLB.511621.50 (TcSR62) es ortóloga a una proteína de humano y a otra de *Drosophila melanogaster* (tabla 6) y reconocería los motivos "GMGWGG, GCKSTG, GNGTGGC, GMGGC, GCDGGCG, TGCGGKTG, CGTGS, CGCTGGCK, GCGSCS, SGYGCG, GWTCTGTG, GCCGYGGG, CRCGGTC, CGCTGCR, RKCGC, CGNCC". En conjunto, estos motivos están presentes en 557 de los 569 genes de la base NET (sin familias multigénicas) (tabla 7), pudiendo tener un rol en la compartimentación núcleo-citoplasma en epimastigotas de *T. cruzi*. Sin embargo, como la proteína de *T. cruzi* posee tres dominios RRM y dos dominios de dedos de zinc (CCHC), mientras que la de humano y de *Drosophila* poseen dos y uno respectivamente, los motivos de reconocimiento podrían estar afectados. Los resultados de traductómica y el análisis del ciclo de vida sugieren que la proteína se expresa mayoritariamente en el estadio epimastigota, aunque la proteína no es detectada en el proteoma probablemente porque no sea de alta expresión. La proteína se encuentra en el núcleo en epimastigotas de *T. cruzi*, sin embargo, el tratamiento de los parásitos con actinomicina D produce su relocalización, al igual que la de los mensajeros poliadenilados, en el nucléolo (Nazer y Sanchez, 2011; Nazer *et al.*, 2011), sugiriendo un rol de la misma en la regulación global de mensajeros. Los resultados de relocalización de las proteínas de unión al ARN al nucléolo se reproducen en *L. mexicana*, sin embargo, en *T. brucei* este fenómeno no sucede, sugiriendo que los mecanismos serían diferentes entre las especies (Nazer y Sanchez, 2011).

Respecto a la proteína homóloga en *T. brucei* (el gen Tb927.2.4710 codifica para la proteína TbRRM1), los datos de estabilidad de mensajeros publicados por Fadda *et al.* muestran que no hay cambios en la estabilidad del ARN que codifica para la proteína en

los distintos estadios (Fadda *et al.*, 2014). El *knockdown* de la proteína conduce a la detención del ciclo celular y la muerte por apoptosis en *T. brucei*, lo que sugiere un rol en la regulación de la progresión del ciclo celular del parásito (Levy *et al.*, 2015). Otro reporte del mismo año demuestra que es una nucleoproteína que se asocia al ARN a proteínas RHS e histonas y al factor auxiliar de *splicing*, así como a la proteína de unión a trectos de polipirimidinas 2, pero no a componentes nucleares del spliceosoma (Naguleswaran *et al.*, 2015). En el estadio procíclico (del insecto), interacciona aproximadamente con 1500 mensajeros estando sobre-representados los que codifican proteínas de interacción con el ARN (incluida RRM1), varios transcritos RHS, mensajeros con regiones codificantes largas y una alta proporción de mensajeros estadio específicos que son más expresados en el estadio sanguíneo del parásito. A su vez, los autores destacan que esta proteína no interactúa con transcritos que codifican para proteínas involucradas en la traducción y proteínas específicas del estadio procíclico. Es interesante destacar la coincidencia de las funciones de los genes regulados, con las características de los genes cuyos transcritos están enriquecidos en el núcleo de epimastigotas de *T. cruzi*, descritos en el capítulo 4.1 y de los transcritos excluidos con los de los genes enriquecidos en el citoplasma. Esto nos lleva a hipotetizar un rol similar al de la proteína de *T. brucei*. El *knockdown* de TbRRM1 produce cambios en la abundancia de los ARNm que no correlacionan con la unión de la proteína a los transcritos, y no se afecta el *splicing*. Los autores proponen también un rol de esta proteína en la compactación de la cromatina, sugiriendo que cumpliría un doble papel en la regulación de la expresión génica: transcripcional, permitiendo la transcripción por la ARN polimerasa II regulando la compactación de la cromatina y post-transcripcionalmente por interacción con los mensajeros (Naguleswaran *et al.*, 2015). La proteína de *T. brucei* tiene un dominio SR que podría ser responsable de su rol en el *splicing* (De Gaudenzi *et al.*, 2005). Sin embargo, la región de la proteína de *T. cruzi* es divergente (figura 7) alterando, posiblemente, la funcionalidad de la proteína.

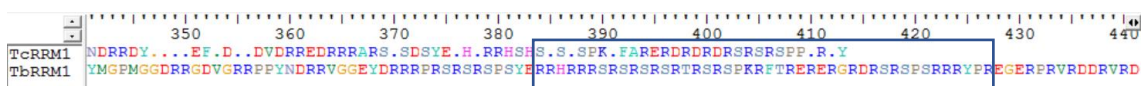


Figura 7. Alineamiento de las regiones carboxilo terminal de las proteínas RRM1 de *T. cruzi* y *T. brucei*. Se observa la divergencia del motivo SR presente en TbRRM1.

Tabla 6. Proteínas que reconocen motivos lineales enriquecidos en las UTRs de los transcritos con localización nuclear. Se muestra el ID de la proteína, la especie a la que pertenece y la descripción; el ID del gen ortólogo en *T. cruzi*, la descripción del gen y los motivos lineales, blancos potenciales.

ID	Organismo	Descripción	ID <i>T. cruzi</i>	Descripción	Motivos reconocidos
NP_002887.2	<i>Homo sapiens</i>	RNA-binding protein 4 isoform 1	TcCLB.511621.50	RNA-binding protein (TcSR62)	GMGWGG, GCKSTG, GNGTGGC, GMGGC, GCDGGCG, TGCGGKTG, CGTGS, CGCTGGCK, GCGSCS, SGYGCG, GWTCTGTG, GCCGYGGG, CRCGGTC, CGCTGCR, RKC GC, CGNCC
NP_523957.1	<i>Drosophila melanogaster</i>	lark, isoform A			
ACR44247.1	<i>Drosophila melanogaster</i>	GM17572p	TcCLB.509581.50	hypothetical protein	GMGWGG, GCGWS, TTTGCGTS, TGCGGKTG, RKC GC, SGYGCG, GSGTGTTT, GCCGYGGG
NP_492958.3	<i>Caenorhabditis elegans</i>	Uncharacterized protein CELE_C17D12.2	TcCLB.507093.229	RNA-binding protein UBP-2 (TcUBP2)	GCKSTG, GWTCTGTG
AAH28101.1	<i>Drosophila melanogaster</i>	Bruno-like 5, RNA binding protein	TcCLB.507093.220	RNA-binding protein UBP-1 (TcUBP1)	GNGTGGC, TGCGGKTG, GWTCTGTG, GSGTGTTT
NP_078950.1	<i>Homo sapiens</i>	protein lin-28 homolog A	TcCLB.509007.40	hypothetical protein	GMGGC, CCMTGAGG, GCGSCS, GCCGYGGG
XP_001579752.1	<i>Trichomonas vaginalis</i>	hypothetical protein	TcCLB.503683.30	RNA-binding protein (DRBD11)	CCMTGAGG
NP_573130.1	<i>Drosophila melanogaster</i>	eukaryotic translation initiation factor 2 α , isoform A	TcCLB.506693.4	EIF2 α	GCDGGCG
NP_511045.1	<i>Drosophila melanogaster</i>	sans fille	TcCLB.504431.90	U2 small nuclear ribonucleoprotein B (snRNP U2b)	MATCCCAC
NP_060547.2	<i>Homo sapiens</i>	RNA-binding protein 28 isoform 1	TcCLB.507037.20	RNA-binding protein	GTGCAGTY
NP_477001.2	<i>Drosophila melanogaster</i>	repressor splicing factor 1, isoform A	TcCLB.506625.70	RNA-binding protein	STGCGACG
XP_011774037.1	<i>Trypanosoma brucei</i>	RNA-binding protein, putative	TcCLB.509965.180	RNA-binding protein	SYGCTG, CTGCGG
NP_116175.1	<i>Homo sapiens</i>	zinc finger CCH domain-containing protein 10	TcCLB.508831.66	hypothetical protein	CYSCG, RKC GC, CASCAS, GCGSCS, SGYGCG, GCCGYGGG, CRGTGC
XP_003722648.1	<i>Leishmania major</i>	putative RNA binding protein	TcCLB.510747.80	Double RNA binding domain protein 9 (DRBD9)	YGRCCG, TKGAACCC

Tabla 7. Motivos lineales enriquecidos en las UTRs de 569 genes de NET que constituyen potenciales blancos de proteínas de *T. cruzi* DREME +, indica el número de secuencias que generaron los motivos; - indica el número de secuencias que poseen el motivo en el grupo control (NET 569 genes, NDT 4379 genes). FiMo indica el resultado de la búsqueda de los motivos generados mediante DREME en la base de datos que lo generó. Conteo, indica el número de veces que aparece el motivo en la base. Se indica también el número de genes que presentan el motivo.

Muestra	control	DREME			FiMo	
		Motivo	+	-	Conteo	N° Genes
5UTR NET	CET	GMGWGG	328	139	521	331
		GCGWS	492	277	333	251
		GCKSTG	271	119	525	318
	NDT	GNGTGGC	160	698	109	101
		TTTGCGTS	47	127	65	59
		GMGGC	470	3155	571	346
		CCMTGAGG	19	24	53	50
		GCDGGCG	101	439	111	89
		MATCCCAC	22	36	25	25
		GTGCAGTY	29	66	67	65
		STGCGACG	33	84	123	112
		TGCGGKTG	39	113	137	116
3UTR NET	CET	CGNCC	283	106	147	118
		SYGCTG	265	98	454	268
		CYSCG	337	154	ND	ND
		RKCGC	299	130	276	216
		CASCAS	171	57	241	175
		CGTGS	277	124	234	191
	NDT	GCGSCS	196	836	375	237
		YGRCCG	129	525	264	191
		CGCTGGCK	20	16	38	36
		SGYGCG	186	915	527	289
		GWTCTGTG	23	32	33	31
		GSGTGTTT	22	30	36	35
		GCCGYGGG	20	26	33	31
		CRCGGTC	22	34	25	24
		CRGTGC	101	436	300	202
		TKGAACCC	14	12	14	14
		CTGCYGG	59	212	61	60
CGCTGCR	46	145	50	50		

Tabla 8. Características de las proteínas que reconocerían motivos enriquecidos en NET Para cada gen se indica el ID, el FC entre los estadios epimastigota y tripomastigota metacíclico del ARNm poliadenilado, el ARNm activamente traducido y la proteína. Se grafican los datos de expresión por microarreglos durante el ciclo de vida del gen y de los genes blancos. Datos tomados de: *Smircich *et al.* (Smircich *et al.*, 2015), **Minning *et al.* (Minning *et al.*, 2009), ***de Godoy *et al.* (de Godoy *et al.*, 2012).

ID	Transcriptoma (RNA/Seq)*		Transcriptoma (Microarreglos)**				Proteoma*** log2FC(MT/E)				
	log2FC(MT/E) polyA	log2FC(MT/E) RFP	Gen		Blancos						
			Ama	Tripo	Epi	Meta	Ama	Tripo	Epi	Meta	
TcCLB.511621.50	-0,43	-2,81					N/A				
TcCLB.509581.50	4,48	6,15					N/A				
TcCLB.507093.229	-0,83	-2,48					-0,44				
TcCLB.507093.220	-0,69	-2,84	N/D				N/A				
TcCLB.509007.40	3,38	3,40					N/A				
TcCLB.503683.30	3,36	3,55					N/A				
TcCLB.506693.4	-1,70	-3,06	N/D				N/A				
TcCLB.504431.90	N/D	N/D					N/D				
TcCLB.507037.20	-0,46	-1,75					N/A				
TcCLB.506625.70	-1,51	-2,74					0,95				
TcCLB.509965.180	-0,72	-3,29					N/A				
TcCLB.508831.66	-0,10	-2,86					N/A				
TcCLB.510747.80	-0,08	-0,97					N/A				

La proteína hipotética TcCLB.509581.50 es ortóloga a una proteína *Drosophila melanogaster* (tabla 6) y reconocería motivos similares a "GMGWGG, GCGWS, TTTGCGTS, TGCGGKTG, RKCGC, SGYGCG, GSGTGTTC, GCCGYGGG". En conjunto, estos motivos están presentes en 525 de los 569 genes de la base NET (sin proteínas multigénicas) (tabla 7). Esta proteína posee dos dominios RRM y no tiene homólogo en *T. brucei* pero sí en *T. rangeli* y *T. grayi*. El perfil de expresión de este gen (tabla 8) sugiere que la proteína se expresa en los estadios infectivos del parásito, por lo tanto, podría estar implicada en la regulación de la expresión de los genes NET en estos estadios.

Otra de las proteínas que podría estar involucrada en la compartimentación núcleo citoplasma de mensajeros es la proteína TcUBP2 (TcCLB.507093.229). Esta fue identificada por el grupo del Dr. Frasch como una proteína de interacción con el ARN, denominada así por su capacidad de interactuar con elementos ricos en U (*U-rich RNA-binding protein*) (D'Orso y Frasch, 2002). Es ortóloga a una proteína no caracterizada de *Caenorhabditis elegans* que podría reconocer motivos similares "GCKSTG y GWTCTGTG". presentes en 332 de los 569 genes de la base NET (tabla 6). La proteína TcUBP2 posee un dominio RRM y dos auxiliares, hacia el extremo amino un dominio rico en glutamina y hacia el carboxilo un dominio rico en glicina, involucrados en interacciones proteína-proteína. Esta proteína, es capaz de unirse *in vitro* a secuencias de ARN ricas en AU y en GU. En el estadio epimastigota, interacciona *in vivo*, con los mensajeros *Tcsmug*, que poseen un elemento rico en AU en el 3'UTR, la interacción no sucede en el estadio tripomastigota (D'Orso y Frasch, 2002). Esta proteína forma un complejo ribonucleoproteico *in vivo* con las proteínas TcUBP1 y PABP que estabiliza los mensajeros *Tcsmug* permitiendo su expresión en el estadio epimastigota (D'Orso y Frasch, 2002; De Gaudenzi *et al.*, 2005). TcUBP2 se expresa mayormente en el estadio epimastigota y en bajos niveles en amastigotas. Los datos de traductómica apoyan este hallazgo (tabla 7). La localización citoplásmica de esta proteína (De Gaudenzi *et al.*, 2005) cuestiona la posibilidad de proponer un rol directo en la compartimentación núcleo-citoplasma, aunque, tal vez, los niveles de proteína en el núcleo pueden no ser suficientemente elevados para su detección por *western blot*.

Otra de las proteínas de la misma familia TcCLB.507093.220 (TcUBP1) es ortóloga a la proteína Bruno-like 5 de *D. melanogaster* y reconocería los motivos "NGTGGC,

TGCGGKTG, GWTCTGTG, GSGTGTTTC" que están presentes en 234 de los 569 genes de la base NET (tabla 6). Esta proteína posee un dominio RRM y tres dominios auxiliares de interacción con proteínas y se la describió como una proteína desestabilizadora del ARN, de unión a elementos ricos en AU (D'Orso y Frasch, 2001b). TcUBP1 se expresa en todos los estadios del ciclo de vida del parásito, aunque a mayor nivel en el estadio tripomastigota, donde los niveles de mensajero *Tcsmug* son bajos. Los datos de traductómica (tabla 7) son discordantes. Esta discordancia puede ser explicada teniendo en cuenta que los resultados de *western blot* en el ciclo de vida del parásito muestran una banda inespecífica, de aproximadamente 90kDa, en el estadio epimastigota, cuyos *reads* en el traductoma podrían ser asignados de forma equivocada a TcUBP1 durante el mapeo. Además del mensajero *Tcsmug*, TcUBP1 reconoce los mensajeros que codifican para las amastinas y el TSAA (*trypanosome small surface antigen*), cumpliendo un rol diferente que depende de la señal en el ARN, de su nivel y de los interactores presentes en cada estadio. Si bien esta proteína presenta localización citoplasmática (De Gaudenzi *et al.*, 2005), presenta una señal de localización nuclear (D'Orso y Frasch, 2002) pudiendo encontrarse en bajos niveles en el núcleo. Es interesante recalcar que las proteínas homólogas de *T. brucei* (TbUBP1 y 2) se localizan tanto en núcleo como en citoplasma, estabilizando los mensajeros de forma independiente de secuencia, sugiriendo que serían proteínas que interactúan con el ARN inespecíficamente (Hartmann, 2007).

La proteína hipotética TcCLB.509007.40 es ortóloga a la proteína homóloga A de lin-28 de humanos (tabla 6). Esta reconocería los motivos "GMGGC, CCMTGAGG, GCGSCS, GCCGYGGG" que en conjunto se encuentran en 456 genes de la base NET (tabla 7). La proteína posee un dominio CSD (*Cold-shock' DNA-binding domain*) y por ontología se la agrupa con las proteínas reguladoras de la transcripción. El ortólogo de *T. brucei* (Tb927.8.7820) disminuye la expresión de los mensajeros, impidiendo su traducibilidad y/o desestabilizándolos (Erben *et al.*, 2014). Los datos de expresión de este gen indican que su mensajero se expresa en epimastigotas y tripomastigotas metacíclicos y es traducido principalmente en el estadio tripomastigota metacíclico (tabla 8). Considerando nuestros resultados y los antecedentes, esta proteína podría estar involucrada en la compartimentación de ARN en *T. cruzi*.

De los análisis de TOM-TOM surge también la proteína de unión al ARN TcCLB.503683.30, ortóloga a una proteína de *Trichomonas vaginalis*, que reconocería el motivo "CCMTGAGG" presente en el 5'UTR de 50 mensajeros (tablas 6 y 7). Esta proteína posee dos dominios RRM. No hay datos de la función de esta proteína, pero el ortólogo en *T. brucei* (Tb927.3.3940, DRBD11) fue reportado como posible factor represor de la expresión de mensajeros (Lueong *et al.*, 2016). La localización citoplasmática de esta proteína según TrypTag (Dean *et al.*, 2017), y los datos de expresión del gen en *T. cruzi*, que indican que no se expresaría en epimastigotas, pero sí en tripomastigotas metacíclicos (tabla 8), no apoyan una función en la compartimentación núcleo-citoplasma de transcritos reconocidos por esta proteína, en epimastigotas de *T. cruzi*.

El factor de inicio de la traducción 2 α TcCLB.506693.4 es ortólogo al factor de inicio de la traducción eucariota 2 α de *D. melanogaster*, que reconocería el motivo "GCDGGCG" en el 5'UTR de 89 mensajeros de NET (tabla 6). Esta proteína posee un dominio de unión al ARN S1 (identificado en la proteína ribosomal S1). Los datos de traductoma indican que la proteína se expresa mayormente en el estadio epimastigota. El ortólogo de *T. brucei* exhibe un patrón reticulado en el citoplasma y se localiza también en el nucléolo (Dean *et al.*, 2017). Estos antecedentes apoyan débilmente un rol en la localización de los mensajeros en el núcleo para esta proteína.

Además, la *small nuclear ribonucleoprotein* U2 B (snRNP U2b), TcCLB.504431.90, ortóloga de la proteína Sans flie de *D. melanogaster*, posee un dominio RRM y reconocería el motivo "MATCCCAC" presente en 25 transcritos de la base. Durante el ciclo de vida del parásito, se expresa en tripomastigotas sanguíneos, pero no en epimastigotas (tabla 8). Por lo que esta proteína no es una buena candidata a regular la compartimentación núcleo-citoplasma a nivel general.

La proteína de unión al ARN TcCLB.507037.20, ortóloga a la proteína 28 de unión al ARN de humanos (*RNA-binding protein 28*), posee dos dominios RRM y reconocería el motivo "GTGCAGTY" presente en el 5'UTR de un set de 65 mensajeros de la base NET (tabla 6 y 7). Los datos de expresión del gen por RNA-Seq no son concordantes con los de microarreglos (tabla 8). Mientras que a partir del primero, el mensajero está en mayor nivel y se traduce más en epimastigotas, los datos del segundo muestran lo contrario.

No hay datos sobre la función de esta proteína, pero la ortóloga de *T. brucei* (Tb927.11.16020, DRBD10) se localiza en el nucléolo (Dean *et al.*, 2017). Con estos antecedentes, no es posible sacar conclusiones respecto al rol de esta proteína en el parásito.

La proteína de unión al ARN TcCLB.506625.70, ortóloga del factor represor de *splicing* 1 de *D. melanogaster*, posee un dominio RRM e interaccionaría con el motivo "STGCGACG" presente en 112 mensajeros de la base NET (tabla 6 y tabla 7). Los datos de RNA-Seq, indican que el gen se expresa mayoritariamente en epimastigotas, pero los niveles de proteína entre epimastigotas y tripomastigotas metacíclicos son similares (tabla 8).

La proteína de unión al ARN TcCLB.509965.180, posee tres dominios RRM, es ortóloga a la proteína de unión al ARN de *T. brucei* (Tb927.6.2550, TRRM2) y reconocería los motivos "SYGCTG, CTGCYGG" presente en 271 mensajeros de la base NET (tabla 6 y 7). Los datos de expresión del gen muestran que el mensajero está en mayor nivel y se traduce más en epimastigota que en tripomastigotas metacíclicos, aunque estos son los dos estadios de menor expresión del gen (tabla 8). No hay datos de la función de esta proteína.

La proteína hipotética TcCLB.508831.66, posee dos dominios tipo dedos de zinc (C3H1), es ortóloga a la proteína con dedos de zinc 10 de humanos y reconocería los motivos "CYSCG, RKCGC, CASCAS, GCGSCS, SGYGCG, GCCGYGGG, CRGTGC" presentes en 454 mensajeros de la base NET (tablas 6 y 7). Los datos de expresión por microarreglos muestran que el mensajero está en mayor nivel en el estadio tripomastigota metacíclico, sin embargo, los datos de RNA-Seq indican lo contrario y, de hecho, la proteína se traduciría más en epimastigotas (tabla 8). No hay datos de la función de esta proteína.

La proteína de unión al ARN TcCLB.510747.80 (DRBD9), posee dos dominios RRM, es ortóloga de la proteína de unión al ARN doble hebra 9 (DRBD9) de *L. major* y reconocería los motivos "YGRCCG, TKGAAC" de 200 mensajeros de la base NET. En *T. brucei* se expresa constitutivamente durante el ciclo de vida, sin embargo, se traduciría solamente en el estadio procíclico de parásitos obtenidos a partir de animales y no en los de cultivo (Jensen *et al.*, 2014). Por otro lado, en *T. brucei* la proteína se localiza en el nucléolo (Dean *et al.*, 2017). La proteína de *T. cruzi*, posee una señal de localización nuclear. En este caso también los datos de expresión por microarreglos y por RNA-Seq son contradictorios,

mostrando mayor nivel de mensajero para el estadio tripomastigota metacíclico, en el primer caso y lo contrario en el segundo (tabla 8). No hay datos de la función de esta proteína. Es interesante destacar que el transcripto se encuentra enriquecido en el núcleo.

Considerando los antecedentes de las proteínas analizadas, surge que la proteína TcCLB.511621.50 podría ser responsable de la compartimentación núcleo-citoplasma de los mensajeros de NET. A su vez, es interesante destacar que la proteína TcCLB.509581.50 podría estar involucrada en la expresión de estos mensajeros en el estadio infectivo del parásito.

Para finalizar, es importante destacar que la búsqueda de motivos fue realizada utilizando las secuencias de las UTRs obtenidas a partir del TriTrypDB (cuyos largos promedio fueron definidos según el promedio del largo de las UTRs determinadas experimentalmente), sin embargo, aproximadamente el 95% de los motivos identificados en el 5' UTR de los transcriptos se encuentran en las UTRs determinadas a partir del transcriptoma citoplásmico y 85% aproximadamente en el 3' UTR.

Las predicciones mediante TOM-TOM de proteínas que pueden reconocer motivos enriquecidos en nuestros grupos de transcriptos con localización diferencial y la búsqueda de ortólogos en nuestro organismo de estudio, nos permitió identificar varias proteínas que podrían estar involucradas en el proceso de compartimentación núcleo-citoplasma en epimastigotas de *T. cruzi*. De este modo, logramos conectar los motivos de secuencias con posibles proteínas reguladoras proporcionándonos una herramienta útil para seleccionar candidatos en los que podemos profundizar.

Materiales y Métodos

Determinación de UTRs

Para delimitar las regiones no traducidas de los mensajeros utilizamos el programa *online* SLaP mapper (Spliced-Leader addition y Polyadenylation) (Fiebiget *et al.*, 2014) utilizando la versión 6 del genoma de *T. cruzi* cepa Esmeraldo como referencia.

Debido a los requerimientos de los programas empleados para la definición de motivos estructurales y de secuencia, se generaron bases de datos de las regiones 3' y 5'UTRs en función de los largos promedio que determinamos para las mismas a partir del transcriptoma citoplasmático (ver capítulo 4.1). Para las regiones 5'UTR se usó un largo de 100 nt corriente arriba de la adenina del codón de inicio que se completó con 290 nt corriente debajo (390nt total). Para la región 3'UTR se usó un largo de 390nt corriente abajo del codón de terminación anotado (figura 8). Para las búsquedas se generaron seis bases de datos correspondientes a los 3' y 5' UTRs de los genes cuyos transcritos se encontraban diferencialmente localizados en núcleo y citoplasma y los genes correspondientes a transcritos no diferenciales.

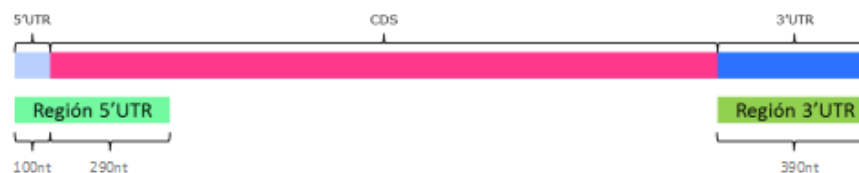


Figura 8. Representación de las regiones empleadas en los análisis de búsqueda de secuencia

Análisis de motivos lineales en el ARN

Una vez definidas las bases de datos de regiones UTR para cada agrupamiento, y para el resto del transcriptoma en general, procedimos a la búsqueda de motivos de secuencia. Se utilizó la suite DREME (Bailey, 2011), perteneciente al paquete MEME-Suite (Bailey *et al.*, 2009) para buscar motivos enriquecidos en las bases de genes diferencialmente localizados en cada uno de los compartimentos respecto a las bases correspondientes a transcritos no diferenciales y del otro compartimento. Solamente se tuvieron en cuenta aquellos motivos que estuvieran presentes en la hebra transcripta y con un *E-value* mayor

a 0,01. En los casos pertinentes se definió la ubicación de los motivos empleando la herramienta FIMO en la base de datos que generó el motivo.

Para la comparación de motivos lineales encontrados en este trabajo con respecto a los presentes en otros organismos y la búsqueda, *in silico*, de proteínas que los reconozcan, se empleó la suite TOM-TOM (Gupta *et al.*, 2007) y la base de datos de motivos de ARN proporcionada por (Ray *et al.*, 2013).

6. Análisis comparativo de transcritos poliadenilados y totales de epimastigotas de *T. cruzi*

Se presenta a continuación un manuscrito en preparación.

“Gene expression regulation in *Trypanosoma cruzi*: clues derived from comparison of transcriptomic data”

Lucía Pastro, Pablo Smircich, María Ana Duhagon, José Sotelo-Silveira and Beatriz Garat.

El análisis comparativo del transcriptoma realizado a partir de ARNs poliadenilados de epimastigotas de la cepa Dm28c (Smircich *et al.*, 2015) y los transcriptomas de ARN total aquí presentados, permite definir una tasa de ARN poliadenilado (maduro) *vs* ARN total (incluye ARNs no procesados) para cada gen y a partir de los datos obtenidos, proponer posibles mecanismos de regulación de los niveles de mensajeros.

Para realizar este análisis comparamos los resultados de expresión del transcriptoma poliadenilado de epimastigotas de *T. cruzi* (Smircich *et al.*, 2015) con los datos de los transcriptomas total, nuclear y citoplasmático presentados previamente en esta tesis (capítulo 4.1). En primer lugar, observamos que en epimastigotas de *T. cruzi* existen mensajeros regulados diferencialmente entre el transcriptoma procesado y los transcriptomas no procesados (figura 1 A, C y D). Los mensajeros que codifican para proteínas involucradas en la traducción están enriquecidos en el transcriptoma poliadenilado (figura 1), mientras que genes cuyos transcritos codifican para proteínas involucradas en procesos de patogenicidad están disminuidos. Además, en el transcriptoma nuclear, están enriquecidos mensajeros que codifican para la vía de señalización del AMPc, involucrada también en la infección (Hamedi *et al.*, 2015) (figura 1F).

La comparación con el transcriptoma poliadenilado de epimastigotas de *T. cruzi*, nos permitió definir grupos de genes sujetos a regulación de estado estacionario de mensajeros, maduración nuclear y/o retención nuclear que regularían su expresión.

En este trabajo mi contribución fue el análisis de los datos, la identificación de los genes diferencialmente procesados y su categorización. Para el grupo de genes con marcada regulación por estabilidad, el Dr. Pablo Smircich hizo la búsqueda de ortólogos en *T. brucei* y analizamos los resultados de estabilidad de transcritos publicados por Fadda *et al.* 2014. Participé en el análisis y la discusión de los resultados, así como en la redacción del manuscrito.

**Gene expression regulation in *Trypanosoma cruzi*: clues derived from
comparison of transcriptomic data**

Running Header:

Comparative analysis of T. cruzi transcriptomes

Abstract

In *Trypanosoma cruzi*, the kinetoplastid responsible for Chagas' disease, no control of transcription initiation of protein coding genes has yet been reported. Gene expression regulation occurs post-transcriptionally, including mechanisms such as nucleus-cytoplasm partitioning, mRNA stability, availability to the translation machinery and translation efficiency. Here we present the comparative analysis of two transcriptome data sets from *T. cruzi* Dm28c epimastigotes: one obtained after polyA transcript isolation and the other from total RNA purification after subcellular fractionation. Genes exhibiting striking polyA to total transcript content rate were identified. Depending on the content of total RNA at each compartment, altered mechanisms of degradation/stabilization, maturation and compartmentation are proposed. The global comparative analysis of the transcriptomes here presented, enabled the identification of sets of genes that remarkably undergo particular post-transcriptional regulatory steps establishing the level of transcript available for expression.

Keywords:

Nucleus-cytoplasm partitioning, transcript maturation, mRNA stability

Trypanosoma cruzi (Kinetoplastidae, Trypanosomatidae) is the protozoan parasite that causes Chagas' disease (Chagas, 1909), affecting approximately 6 to 8 millions of people, mostly from poor rural regions of 21 countries of Central and South America suffering this disease and several millions are in risk of infection (WHO, 2016).

Among other biological peculiarities, the expression of protein coding genes is remarkable distinctive in trypanosomatids. No control of transcription initiation has yet been reported, and the primary transcript produced is polycistronic carrying consecutive non-functionally related genes. The maturation processes (capping, splicing and polyadenylation) which occur co-transcriptionally and are tightly regulated in eukaryotes (Halbeisen *et al.*, 2008), are also unusual in trypanosomatids. Indeed, while the *cis*-splicing only exceptionally occurs (Ivens *et al.*, 2005), the maturation of the primary transcript necessarily involves the addition of a capped short sequence named spliced leader at the 5' end (*trans*-splicing) and the coordinated polyadenylation at the 3' end of the individual messenger. Though these processes may be regulated, a global pervasive maturation has been suggested for *T. cruzi* (Smircich *et al.*, 2015). Nevertheless, the mature messengers, even if derived from the same polycistronic transcript, are present in different levels (Haile and Papadopoulou, 2007; Minning *et al.*, 2009; Kramer and Carrington, 2011; Smircich *et al.*, 2015).

Studies on post-transcriptional mechanisms have mainly focused on cytoplasmic processes such as: mRNA stability (Javier De Gaudenzi, Frasch and Clayton, 2005; Araújo *et al.*, 2011; Bayer-Santos *et al.*, 2012; Guerra-Slompo *et al.*, 2012; Li *et al.*, 2012), mRNA localization (Cassola, De Gaudenzi and Frasch, 2007; Holetz *et al.*, 2007; Cassola, 2011; Kramer, 2014) and translation control (da Silva Augusto *et al.*, 2015; Smircich *et al.*, 2015). More recently, we have suggested that nucleus – cytoplasm partitioning may play a major role controlling the amount of transcript available for translation in *T. cruzi* (Pastro *et al.*, 2017).

Considering that the comparison of the transcriptome data sets of *T. cruzi* Dm28c epimastigotes obtained after polyA transcript isolation (Smircich *et al.*, 2015) and the one obtained from total RNA purification from whole cell, cytoplasm and nucleus enriched fractions (Pastro *et al.*, 2017), may reveal genes undergoing specific mechanisms affecting the transcript maturation process and subcellular localization, we performed an *in silico* analysis. Indeed, while all the polyA content for a specific transcript may be

detected in the total RNA approach, the relative content deviation for a specific gene from the population's average gives evidence for the existence of an altered processes of maturation and degradation or stability.

To determine RNA transcript levels, the annotated transcripts of *T. cruzi* CLBrener Esmeraldo-like, genome release 4.2 from the TriTrypDB were used for the polyA transcriptome data obtained by SOLiD RNA-Seq (ID: PRJNA260933) (Smircich *et al.*, 2015), and for the total RNA data obtained by Illumina HiSeq2000 platform with a paired-end run (Pastro *et al.*, 2017). Only genes with $p < 0.05$ and 20 reads or more per transcript, reaching a total of 5367 genes were used for the comparative analysis. The RPKM ratios of the polyA to the total RNA were obtained for each gene, the median of the population were calculated for whole cell, cytoplasmic and nuclear fractions and the genes with ratios higher and lower than 5 times the median were further analysed. Even when the methods to obtain the transcriptomes were different (Smircich *et al.* 2015; Pastro *et al.* 2017), the stringent conditions of selection undoubtedly drive to a limited but confident list of genes that remarkably undergo particular post-transcriptional regulatory steps. .

In accordance with the pervasive maturation previously proposed, we found a good correlation (Pearson's r 0.63) between the data derived from those two approaches Figure 1A. Nonetheless, the existence of genes whose transcripts exhibit a remarkable deviation from the average were also detected. We found 98 genes with polyA transcript levels strikingly high in comparison with the total transcript level (ratio > 5 medians) (Supplementary Table 1). This pattern can be explained by an increased mRNA stability (which increases the amount of poly adenylated RNA) and/or quick maturation (which decreases the amount of non-poly adenylated RNA) in comparison with the average. On the contrary, 314 genes showed a very low polyA transcript proportion when compared to the total transcript (ratio < 0.2 medians) (Supplementary Table 1). For these genes, the existence of either a slower maturation leading to less polyA RNA or an increase in the deadenylation process could be foreseen.

The ontology analysis, using the Blast2GO tool (Conesa and Gotz, 2008) with a gene cut off of 10 and the DAVID server (Database for Annotation, Visualization and Integrated Discovery) (Dennis Jr. *et al.*, 2003) for overrepresentation analyses, revealed several enriched functional categories. Among the genes whose transcripts are enriched in polyA-RNA, the most overrepresented functional category (Bonferroni corrected p -value < 0.05) corresponds to translation (Figure 1B). This result is in good agreement with

previous reported polyA transcriptomic data and even with proteomic data about the high abundance of translation machinery components in the *T. cruzi* epimastigotes stage (Atwood *et al.*, 2005; Minning *et al.*, 2009; de Godoy *et al.*, 2012; Smircich *et al.*, 2015). Meanwhile, among the ones that are mostly deprived from polyA tails, protein phosphorylation and microtubule-based processes are the most overrepresented categories (Bonferroni corrected p-value < 0.0001) (Figure 1B). This finding suggests a high turnover for transcripts whose products are implied in processes that may rapidly change to enable the structural and functional transformation that the parasite undergoes during life cycle and/or in response to environmental alterations.

Assuming that the polyA RNA is mainly found in the cytoplasm the comparison of the transcriptomes derived from polyA-RNA and the one derived from total cytoplasmic RNA could give information about the genes that are markedly regulated at this subcellular location separating the influence of nuclear phenomena from the analysis. A similar good correlation (r 0.61) was observed between the data derived from these two transcriptomes (Figure 1 C) suggesting a negligible contribution of the nuclear compartment to the global transcript content. We found that up to 118 genes have a high level of polyA transcripts in comparison with total RNA (ratio>5 medians), meanwhile 307 genes characteristically showed a relative low level of polyA transcripts (ratio<0.2 medians) (Supplementary Table 2). While a regulation by stabilization is predicted for the first ones, the later ones are distinctively losing their polyA tail and therefore being actively degraded or stored as non-polyA RNA in the cytoplasm. The ontology analysis revealed that the categories of genes whose transcripts are enriched or mostly deprived in polyA-RNA with respect to total RNA in the *T. cruzi* cytoplasm, are the same as the ones outlined for the analysis with the whole cell transcriptome (Figure 1 D). As data derived from a global analysis of the *T. brucei* transcript half-life is available (Fadda *et al.*, 2014), we analysed the behaviour of the orthologous genes in both categories. The *T. brucei* genes orthologous to the genes exhibiting a high polyA to total cytoplasmic RNA ratio (56) present higher half-life (t test p 0.007) in contrast to the ones exhibiting a low polyA to total cytoplasmic RNA ratio (202) (See Supplementary Table 3 for the *T. brucei* orthologous gene ID). This finding supports our interpretation of the degradation rates in the current analysis of *T. cruzi* genes.

Since the comparison of the polyA to total RNA from whole cell and cytoplasm are not absolutely coincident, the nuclear transcript content may have a non-negligible contribution to the stationary RNA level as commonly assumed. Therefore, we compared the transcriptomes derived from polyA-RNA and the one derived from the nuclear fraction. In this case, a poor correlation was observed (r 0.29, Figure 1E). We found that

up to 299 genes have a high level of polyA transcripts in comparison with total RNA from the nuclear enriched fraction (ratio>5 medians), meanwhile 479 genes characteristically showed a relative low level of polyA transcripts (ratio<0.2 medians) (Supplementary Table 4). While rapid maturation can be predicted for the first ones, a slow maturation process could justify the high contribution of non polyA tailed transcripts in the nucleus observed for the later ones. As previously, the ontology analysis revealed genes involved in translation as the ones that are characteristically enriched in polyA species and genes involved in protein phosphorylation and microtubule-based processes as the ones that are mostly enriched in non-mature transcript species (Figure 1F). Nonetheless, genes involved in cAMP metabolism were also found to have a remarkable high proportion of non-mature transcript in the nucleus in comparison to the polyA species (Figure 1F). The cAMP signalling pathway has been related to parasite infection in *T. cruzi* and others trypanosomatids (Hamedi *et al.*, 2015; Makin and Gluenz, 2015; Tagoe, Kalejaiye and de Koning, 2015) and thus, the expression of the involved genes is not expected to occur at the non-infective epimastigote stage.

It is worth to recall that trans-splicing has also been detected at the cytoplasm in *T. cruzi* (Jager *et al.*, 2007). Though the frequency of such process is not clear, the existence of these alternative maturation processes introduces another mechanism that can be acting to explain the distinct proportion of polyA to total RNA found at the cytoplasmic fraction for some genes. In addition, in eukaryotes the nuclear retention of spliced polyadenylated mRNA has been demonstrated (Gondran *et al.*, 1999; Weil *et al.*, 2000; Bahar Halpern *et al.*, 2015). Since such occurrence has not yet been investigated in *T. cruzi*, we cannot rule out the possibility of the presence of mature mRNA in the nucleus thus contributing to the polyA content. These two considerations add additional terms in the analysis of the mechanisms responsible for the polyA to total transcript observed for the nucleus and cytoplasm fraction. However, the data here presented unveils set of genes that undergo processes such as cytoplasmic stabilization/degradation processes, nuclear maturation and/or subcellular partitioning processes, assuring the level of transcript available for expression.

Acknowledgments

This work was supported by CSIC and CAP UdelaR, ANII and PEDECIBA, Uruguay. We acknowledge all members of the Laboratorio de Interacciones Moleculares at UDELAR and the Department of Genomics at IIBCE for constant discussion.

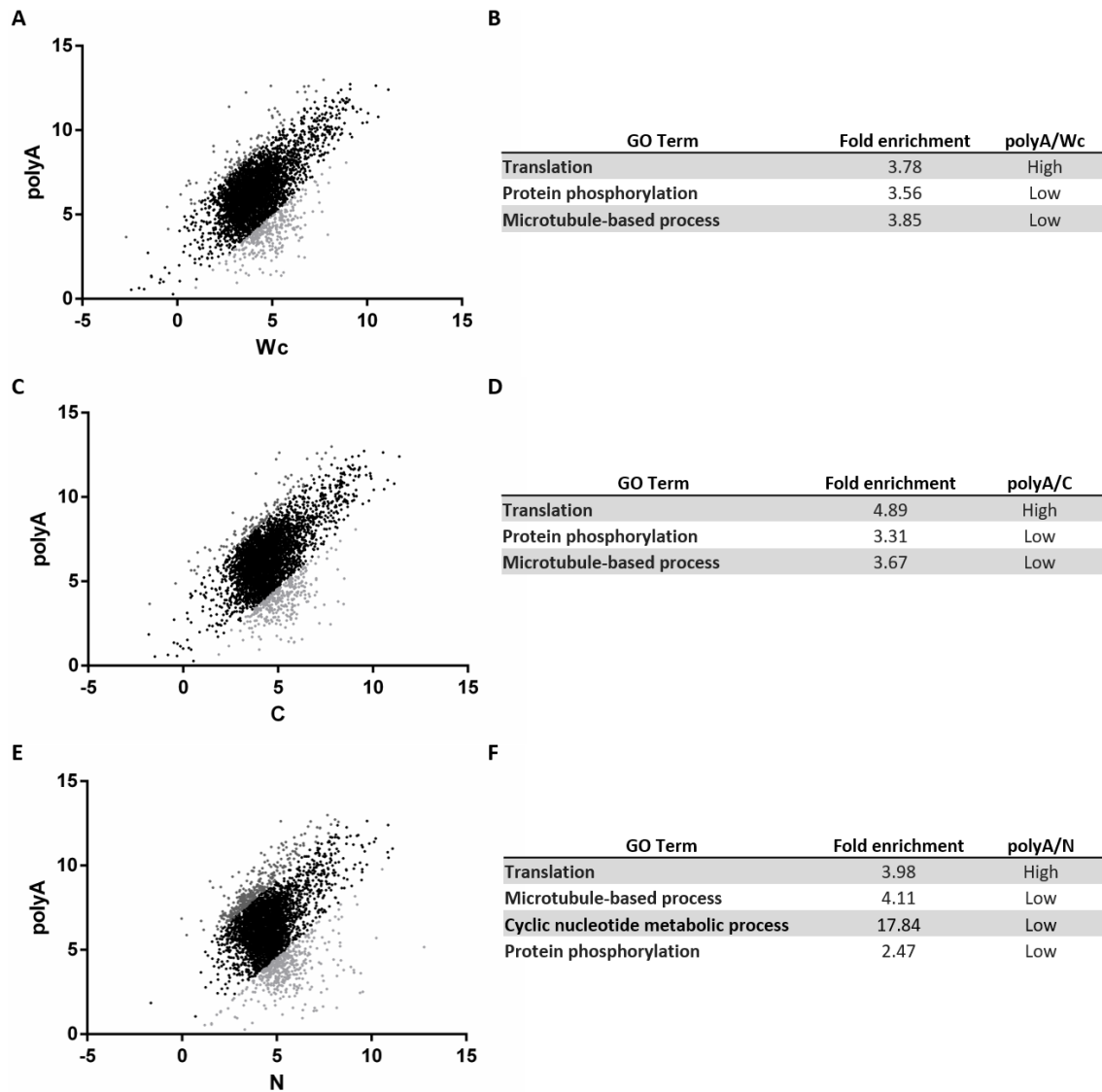


Fig. 1. Comparison of data derived from transcriptomes of *T. cruzi* Dm28c epimastigotes (Smircich *et al.*, 2015) and (Pastro *et al.*, 2017). Scatter plot of the estimated expression levels derived from polyA-RNA transcriptome (polyA) and total RNA from whole cell (Wc) transcriptome (**A**), and total RNA from cytoplasmic (C) transcriptome (**C**) and total RNA from nuclear (N) transcriptome (**E**); gene clustering of transcripts with low (ratio < 0.2 medians) or high (ratio > 5 medians) of polyA mRNA abundance compared to total RNA from whole cell (**B**), cytoplasm (**D**) and nucleus (**F**). See Supplementary Table 1, 2 and 3 for list of genes. Grouping of genes according to Blast2GO biological processes.

Supplementary Table 1 List of genes with remarkable proportion of polyA to total RNA transcript in whole epimastigote cells of *T. cruzi* Dm28c

High: Genes whose polyA to total whole cell RNA transcriptomes ratio is >5 medians. Low: Genes whose polyA to total whole cell RNA transcriptomes ratio is < 0.2 medians. Data derived from reported transcriptomes (Smircich *et al.*, 2015) and (Pastro *et al.*, 2017).

High Wc vs polyA						
TcCLB.507447.19	TcCLB.510409.10	TcCLB.507465.10	TcCLB.510535.20	TcCLB.510741.70	TcCLB.504103.30	TcCLB.511127.60
TcCLB.507603.260	TcCLB.504625.70	TcCLB.511467.20	TcCLB.508811.30	TcCLB.510149.80	TcCLB.511071.130	TcCLB.503835.39
TcCLB.484311.10	TcCLB.441241.10	TcCLB.508997.20	TcCLB.475705.9	TcCLB.507885.10	TcCLB.508541.190	TcCLB.510055.150
TcCLB.504423.40	TcCLB.510877.10	TcCLB.511365.4	TcCLB.510609.60	TcCLB.461515.10	TcCLB.511557.50	TcCLB.510283.160
TcCLB.504703.20	TcCLB.507875.10	TcCLB.508165.130	TcCLB.503895.19	TcCLB.506811.180	TcCLB.509167.110	TcCLB.506779.80
TcCLB.506673.69	TcCLB.448567.9	TcCLB.511001.160	TcCLB.508829.9	TcCLB.511621.110	TcCLB.511431.90	TcCLB.508215.3
TcCLB.412093.10	TcCLB.509473.10	TcCLB.508567.100	TcCLB.479379.9	TcCLB.510073.30	TcCLB.511189.70	TcCLB.503409.10
TcCLB.511443.10	TcCLB.511445.10	TcCLB.506937.40	TcCLB.504425.110	TcCLB.510231.10	TcCLB.506165.4	TcCLB.398751.10
TcCLB.511253.31	TcCLB.441061.9	TcCLB.509247.60	TcCLB.510749.4	TcCLB.509137.20	TcCLB.506691.30	TcCLB.510565.70
TcCLB.510283.10	TcCLB.510247.10	TcCLB.509171.100	TcCLB.506179.59	TcCLB.504883.70	TcCLB.506655.7	TcCLB.507929.10
TcCLB.511441.10	TcCLB.506213.120	TcCLB.509589.50	TcCLB.510769.80	TcCLB.508989.90	TcCLB.511417.80	TcCLB.503955.110
TcCLB.508837.80	TcCLB.510533.10	TcCLB.504929.5	TcCLB.509537.50	TcCLB.507809.140	TcCLB.507809.14	TcCLB.468941.9
TcCLB.506925.550	TcCLB.506401.24	TcCLB.506743.60	TcCLB.511001.60	TcCLB.508239.30	TcCLB.506775.30	TcCLB.509269.4
TcCLB.509891.70	TcCLB.507531.19	TcCLB.511253.10	TcCLB.506399.40	TcCLB.511425.49	TcCLB.506857.10	TcCLB.504147.110
TcCLB.506525.120	TcCLB.506859.204	TcCLB.506947.10	TcCLB.511131.80	TcCLB.438059.10	TcCLB.507011.250	TcCLB.511695.10
TcCLB.510531.109	TcCLB.509585.10	TcCLB.508137.40	TcCLB.510125.10	TcCLB.510427.10	TcCLB.511867.130	TcCLB.511911.81
TcCLB.511867.209	TcCLB.504107.10	TcCLB.511115.10	TcCLB.424195.9	TcCLB.506203.120	TcCLB.506925.150	TcCLB.510861.140
TcCLB.508439.40	TcCLB.508307.140	TcCLB.504053.10	TcCLB.506529.595	TcCLB.510317.7	TcCLB.507429.60	TcCLB.508277.360
TcCLB.511141.10	TcCLB.508569.70	TcCLB.508181.4	TcCLB.506167.50	TcCLB.506407.90	TcCLB.458153.10	TcCLB.510997.50
TcCLB.511527.9	TcCLB.511685.60	TcCLB.504255.30	TcCLB.509279.10	TcCLB.506249.70	TcCLB.507809.80	TcCLB.510311.100
TcCLB.509013.19	TcCLB.511535.10	TcCLB.510295.59	TcCLB.506503.90	TcCLB.506529.480	TcCLB.504449.14	TcCLB.503865.20
TcCLB.404001.20	TcCLB.408573.9	TcCLB.507083.109	TcCLB.507709.90	TcCLB.509631.140	TcCLB.507737.40	TcCLB.510861.150
TcCLB.507933.10	TcCLB.504117.50	TcCLB.506963.70	TcCLB.503629.10	TcCLB.504125.70	TcCLB.507639.110	TcCLB.508347.90
TcCLB.445281.9	TcCLB.510773.9	TcCLB.509775.30	TcCLB.503751.30	TcCLB.508211.50	TcCLB.505007.10	TcCLB.507867.100
TcCLB.509173.9	TcCLB.474341.9	TcCLB.507641.14	TcCLB.507639.90	TcCLB.506773.10	TcCLB.509149.60	TcCLB.510283.60
TcCLB.506973.140	TcCLB.506401.350	TcCLB.509233.120	TcCLB.507099.80	TcCLB.426397.9	TcCLB.508347.40	TcCLB.507275.20
TcCLB.510309.70	TcCLB.508723.70	TcCLB.506941.9	TcCLB.503613.10	TcCLB.511181.110	TcCLB.509775.20	TcCLB.511113.9
TcCLB.506849.20	TcCLB.506859.230	TcCLB.506863.80	TcCLB.508541.230	TcCLB.511127.300	TcCLB.510525.90	TcCLB.506625.60
TcCLB.509583.19	TcCLB.511327.20	TcCLB.511869.9	TcCLB.503975.40	TcCLB.445553.10	TcCLB.511407.50	TcCLB.511325.50
TcCLB.509599.20	TcCLB.508831.4	TcCLB.510747.170	TcCLB.430061.9	TcCLB.506529.600	TcCLB.508385.30	TcCLB.511555.110
TcCLB.510533.210	TcCLB.462617.10	TcCLB.506825.40	TcCLB.509801.30	TcCLB.506357.130	TcCLB.508307.180	TcCLB.506251.130
TcCLB.509103.39	TcCLB.510515.20	TcCLB.509805.210	TcCLB.505007.30	TcCLB.506859.240	TcCLB.504125.100	TcCLB.511427.19
TcCLB.506925.170	TcCLB.507099.40	TcCLB.506907.14	TcCLB.507677.150	TcCLB.507795.10	TcCLB.508781.10	TcCLB.506637.10
TcCLB.506401.12	TcCLB.509247.30	TcCLB.508731.60	TcCLB.506825.10	TcCLB.508569.10	TcCLB.507805.30	TcCLB.506753.230
TcCLB.511749.9	TcCLB.507831.110	TcCLB.509979.40	TcCLB.510565.179	TcCLB.510509.9	TcCLB.510247.20	TcCLB.511253.20
TcCLB.504133.15	TcCLB.510861.14	TcCLB.480785.10	TcCLB.510159.19	TcCLB.507529.20	TcCLB.510773.120	TcCLB.511545.80
TcCLB.504153.90	TcCLB.433273.10	TcCLB.477079.10	TcCLB.503795.10	TcCLB.508041.10	TcCLB.507011.230	TcCLB.510755.19
TcCLB.508347.159	TcCLB.506543.50	TcCLB.503683.11	TcCLB.506811.160	TcCLB.506505.10	TcCLB.511425.40	TcCLB.511435.49
TcCLB.509007.40	TcCLB.504087.20	TcCLB.506925.10	TcCLB.511559.70	TcCLB.508133.19	TcCLB.503479.40	TcCLB.510595.4
TcCLB.511751.14	TcCLB.503989.50	TcCLB.510297.10	TcCLB.509455.9	TcCLB.510855.4	TcCLB.480997.10	TcCLB.511421.80
TcCLB.509013.10	TcCLB.405179.10	TcCLB.510603.130	TcCLB.508277.100	TcCLB.510431.190	TcCLB.511751.70	TcCLB.507083.20
TcCLB.506401.290	TcCLB.511313.10	TcCLB.509067.60	TcCLB.510859.50	TcCLB.503531.49	TcCLB.506543.100	TcCLB.506697.20
TcCLB.503419.10	TcCLB.503479.70	TcCLB.510771.9	TcCLB.508541.240	TcCLB.506401.110	TcCLB.510609.100	TcCLB.511001.90
TcCLB.506829.90	TcCLB.506533.34	TcCLB.506503.80	TcCLB.506949.20	TcCLB.506465.40	TcCLB.503975.50	TcCLB.503671.39
TcCLB.511431.10	TcCLB.503955.100	TcCLB.506825.120	TcCLB.508209.10	TcCLB.511337.4	TcCLB.511127.390	
Low Wc vs polyA						
TcCLB.503579.90	TcCLB.511729.40	TcCLB.506859.130	TcCLB.510755.89	TcCLB.506779.60	TcCLB.510029.70	TcCLB.509141.40
TcCLB.509667.20	TcCLB.506925.130	TcCLB.507611.280	TcCLB.509965.200	TcCLB.504077.30	TcCLB.506963.14	TcCLB.510349.50
TcCLB.506801.80	TcCLB.508675.29	TcCLB.507841.34	TcCLB.506405.110	TcCLB.510407.20	TcCLB.510381.10	TcCLB.506249.10
TcCLB.510971.5	TcCLB.509157.220	TcCLB.503899.130	TcCLB.508285.50	TcCLB.510911.45	TcCLB.504099.30	TcCLB.508039.70
TcCLB.505009.49	TcCLB.510275.405	TcCLB.506945.290	TcCLB.484299.10	TcCLB.510295.30	TcCLB.506943.130	TcCLB.503453.40
TcCLB.506717.80	TcCLB.504429.70	TcCLB.511729.30	TcCLB.503893.90	TcCLB.438147.18	TcCLB.511369.5	TcCLB.510591.20

TcCLB.507105.120	TcCLB.506147.120	TcCLB.506887.100	TcCLB.506775.140	TcCLB.508037.10	TcCLB.506749.20	TcCLB.511685.30
TcCLB.508605.20	TcCLB.510243.80	TcCLB.510005.20	TcCLB.509349.30	TcCLB.503957.20	TcCLB.509735.30	TcCLB.508321.21
TcCLB.511621.90	TcCLB.507041.50	TcCLB.507389.50	TcCLB.508209.100	TcCLB.505843.24	TcCLB.510165.110	TcCLB.503953.14
TcCLB.510525.40	TcCLB.511217.209	TcCLB.506483.69	TcCLB.509129.10	TcCLB.507093.339	TcCLB.506533.106	TcCLB.506265.50
TcCLB.509993.15	TcCLB.508473.10	TcCLB.510755.50	TcCLB.509147.50	TcCLB.503855.20	TcCLB.508285.20	TcCLB.404711.10
TcCLB.482471.10	TcCLB.506341.30	TcCLB.507257.120	TcCLB.504433.10	TcCLB.510451.10	TcCLB.508541.220	TcCLB.511577.180
TcCLB.511071.80	TcCLB.510145.20	TcCLB.507875.110	TcCLB.509233.190	TcCLB.508727.18	TcCLB.511369.15	TcCLB.507045.5
TcCLB.506671.70	TcCLB.503579.70	TcCLB.508727.54	TcCLB.507611.389	TcCLB.506563.50	TcCLB.509815.90	TcCLB.504089.60

Supplementary Table 2 List of genes with remarkable proportion of polyA to total RNA transcript in cytoplasm fraction of epimastigote cells of *T. cruzi* Dm28c

High: Genes whose polyA to total cytoplasm RNA transcriptomes ratio is >5 medians. Low: Genes whose polyA to total cytoplasm RNA transcriptomes ratio is < 0.2 medians. Data derived from reported transcriptomes (Smircich *et al.*, 2015) and (Pastro *et al.*, 2017).

High N vs polyA						
TcCLB.507447.19	TcCLB.448567.9	TcCLB.509589.50	TcCLB.506811.180	TcCLB.509893.10	TcCLB.503671.39	TcCLB.511113.9
TcCLB.510283.10	TcCLB.511535.10	TcCLB.506963.70	TcCLB.506863.80	TcCLB.507805.30	TcCLB.508277.360	TcCLB.504625.70
TcCLB.504423.40	TcCLB.506907.14	TcCLB.511427.19	TcCLB.504425.110	TcCLB.511001.90	TcCLB.511189.70	TcCLB.506625.60
TcCLB.504703.20	TcCLB.509585.10	TcCLB.506753.230	TcCLB.509631.140	TcCLB.511337.4	TcCLB.503835.39	TcCLB.511445.10
TcCLB.484311.10	TcCLB.462617.10	TcCLB.511115.10	TcCLB.506399.40	TcCLB.510149.80	TcCLB.506775.30	TcCLB.510409.10
TcCLB.506973.140	TcCLB.506543.50	TcCLB.510535.20	TcCLB.510125.10	TcCLB.506357.130	TcCLB.506223.120	TcCLB.506401.290
TcCLB.507603.260	TcCLB.510533.10	TcCLB.508731.60	TcCLB.507809.140	TcCLB.506859.240	TcCLB.506773.10	TcCLB.507875.10
TcCLB.412093.10	TcCLB.506213.120	TcCLB.510769.80	TcCLB.510317.7	TcCLB.510431.190	TcCLB.509351.6	TcCLB.507465.10
TcCLB.506673.69	TcCLB.507531.19	TcCLB.509171.100	TcCLB.510859.50	TcCLB.508989.90	TcCLB.511693.30	TcCLB.509805.210
TcCLB.511441.10	TcCLB.474341.9	TcCLB.430061.9	TcCLB.503613.10	TcCLB.508777.40	TcCLB.510863.9	TcCLB.480785.10
TcCLB.511253.31	TcCLB.503479.70	TcCLB.507083.20	TcCLB.511001.60	TcCLB.505007.10	TcCLB.503479.40	TcCLB.511001.160
TcCLB.511443.10	TcCLB.503419.10	TcCLB.438059.10	TcCLB.508209.10	TcCLB.428999.10	TcCLB.509269.4	TcCLB.506925.10
TcCLB.508837.80	TcCLB.508307.140	TcCLB.508097.10	TcCLB.511131.80	TcCLB.506691.30	TcCLB.434931.20	TcCLB.503955.100
TcCLB.508439.40	TcCLB.508723.70	TcCLB.506941.9	TcCLB.508277.100	TcCLB.508995.10	TcCLB.510773.120	TcCLB.509233.120
TcCLB.510531.109	TcCLB.506829.90	TcCLB.510159.19	TcCLB.509149.60	TcCLB.509801.30	TcCLB.506505.10	TcCLB.510609.60
TcCLB.506925.550	TcCLB.509473.10	TcCLB.510747.170	TcCLB.511559.70	TcCLB.445553.10	TcCLB.510057.4	TcCLB.510525.90
TcCLB.509891.70	TcCLB.506401.350	TcCLB.510855.4	TcCLB.503975.40	TcCLB.511621.110	TcCLB.508215.3	TcCLB.506503.90
TcCLB.510533.210	TcCLB.503989.50	TcCLB.506825.40	TcCLB.506529.595	TcCLB.422041.9	TcCLB.511911.81	TcCLB.510755.19
TcCLB.509013.19	TcCLB.504107.10	TcCLB.505007.30	TcCLB.508307.180	TcCLB.506465.40	TcCLB.504147.110	TcCLB.509137.20
TcCLB.511527.9	TcCLB.508831.4	TcCLB.506637.10	TcCLB.479379.9	TcCLB.507709.90	TcCLB.506825.120	TcCLB.509247.60
TcCLB.509173.9	TcCLB.506251.130	TcCLB.509067.60	TcCLB.507529.20	TcCLB.511407.50	TcCLB.479253.9	TcCLB.504883.70
TcCLB.506525.120	TcCLB.441061.9	TcCLB.506947.10	TcCLB.507099.80	TcCLB.503975.50	TcCLB.506401.110	TcCLB.506949.20
TcCLB.404001.20	TcCLB.510861.14	TcCLB.511557.50	TcCLB.507639.110	TcCLB.506925.150	TcCLB.506179.59	TcCLB.510297.10
TcCLB.509583.19	TcCLB.509007.40	TcCLB.509979.40	TcCLB.504137.50	TcCLB.485889.16	TcCLB.503531.49	TcCLB.509537.50
TcCLB.506849.20	TcCLB.507099.40	TcCLB.503751.30	TcCLB.511417.80	TcCLB.507993.230	TcCLB.509167.110	TcCLB.508347.90
TcCLB.511867.209	TcCLB.511555.110	TcCLB.510603.130	TcCLB.510509.9	TcCLB.511071.130	TcCLB.509455.9	TcCLB.511695.10
TcCLB.509599.20	TcCLB.506859.230	TcCLB.504255.30	TcCLB.511043.60	TcCLB.506203.40	TcCLB.511545.80	TcCLB.510247.20
TcCLB.507933.10	TcCLB.511253.20	TcCLB.506811.160	TcCLB.510073.30	TcCLB.507795.10	TcCLB.398751.10	TcCLB.504111.49
TcCLB.511141.10	TcCLB.508181.4	TcCLB.511869.9	TcCLB.508239.30	TcCLB.508211.50	TcCLB.508811.30	TcCLB.511867.130
TcCLB.445281.9	TcCLB.405179.10	TcCLB.511435.49	TcCLB.506167.50	TcCLB.506529.600	TcCLB.506937.40	TcCLB.510231.10
TcCLB.506925.170	TcCLB.508569.70	TcCLB.504929.5	TcCLB.506249.70	TcCLB.504125.100	TcCLB.511467.20	TcCLB.508089.10
TcCLB.511751.14	TcCLB.507831.110	TcCLB.424195.9	TcCLB.508829.9	TcCLB.503795.10	TcCLB.508137.40	TcCLB.506857.10
TcCLB.508347.159	TcCLB.511327.20	TcCLB.475705.9	TcCLB.511421.80	TcCLB.507011.230	TcCLB.507641.14	TcCLB.509279.10
TcCLB.504117.50	TcCLB.504133.15	TcCLB.508567.100	TcCLB.510295.59	TcCLB.506655.7	TcCLB.506401.24	TcCLB.511425.49
TcCLB.506401.12	TcCLB.511685.60	TcCLB.503683.11	TcCLB.504449.14	TcCLB.507627.9	TcCLB.511325.50	TcCLB.426397.9
TcCLB.509103.39	TcCLB.511313.10	TcCLB.503629.10	TcCLB.510565.179	TcCLB.504103.30	TcCLB.511253.10	TcCLB.511127.300
TcCLB.510309.70	TcCLB.510515.20	TcCLB.503895.19	TcCLB.508541.240	TcCLB.507011.250	TcCLB.408573.9	TcCLB.507809.14
TcCLB.510247.10	TcCLB.508997.20	TcCLB.507083.109	TcCLB.506529.480	TcCLB.508347.40	TcCLB.477079.10	TcCLB.506519.140
TcCLB.510877.10	TcCLB.433273.10	TcCLB.511365.4	TcCLB.508541.230	TcCLB.461515.10	TcCLB.509775.20	TcCLB.506165.4

TcCLB.511431.10	TcCLB.504087.20	TcCLB.506503.80	TcCLB.506407.90	TcCLB.503409.10	TcCLB.405165.19	TcCLB.504125.70
TcCLB.504153.90	TcCLB.506533.34	TcCLB.510771.9	TcCLB.506203.120	TcCLB.506413.89	TcCLB.506859.204	TcCLB.509775.30
TcCLB.441241.10	TcCLB.508165.130	TcCLB.507639.90	TcCLB.511181.110	TcCLB.511127.390	TcCLB.511131.30	TcCLB.506825.10
TcCLB.509013.10	TcCLB.506743.60	TcCLB.510595.4	TcCLB.508973.80	TcCLB.511751.70	TcCLB.510741.70	TcCLB.509247.30
TcCLB.511749.9	TcCLB.504053.10	TcCLB.506697.20	TcCLB.507677.150	TcCLB.508133.19	TcCLB.510773.9	
Low N vs polyA						
TcCLB.508037.10	TcCLB.511071.80	TcCLB.509233.150	TcCLB.509157.220	TcCLB.510349.50	TcCLB.503449.14	TcCLB.506289.60
TcCLB.503653.70	TcCLB.507689.50	TcCLB.507041.50	TcCLB.509129.10	TcCLB.484299.10	TcCLB.511621.90	TcCLB.511729.40
TcCLB.506413.30	TcCLB.507257.120	TcCLB.511589.50	TcCLB.510243.80	TcCLB.508991.10	TcCLB.506945.290	TcCLB.508723.90
TcCLB.511867.90	TcCLB.510749.90	TcCLB.506925.130	TcCLB.509967.10	TcCLB.509147.50	TcCLB.508989.60	TcCLB.508039.70
TcCLB.506249.10	TcCLB.510737.40	TcCLB.510755.50	TcCLB.506405.110	TcCLB.508285.20	TcCLB.507611.389	TcCLB.510145.20
TcCLB.510351.140	TcCLB.507105.120	TcCLB.510971.5	TcCLB.482471.10	TcCLB.510029.70	TcCLB.504429.70	TcCLB.506413.20
TcCLB.438147.18	TcCLB.510609.20	TcCLB.504089.60	TcCLB.507099.130	TcCLB.506671.70	TcCLB.507611.10	TcCLB.503893.90
TcCLB.510173.70	TcCLB.508943.40	TcCLB.509267.60	TcCLB.508209.100	TcCLB.507045.5	TcCLB.507093.339	TcCLB.504433.10
TcCLB.507011.240	TcCLB.508207.110	TcCLB.508515.20	TcCLB.506775.140	TcCLB.504099.30	TcCLB.510525.40	TcCLB.509233.190
TcCLB.510759.10	TcCLB.506779.60	TcCLB.508727.54	TcCLB.506341.30	TcCLB.511369.5	TcCLB.508781.120	TcCLB.510591.20
TcCLB.510005.20	TcCLB.509965.200	TcCLB.508605.20	TcCLB.508727.18	TcCLB.506963.14	TcCLB.503899.119	TcCLB.510295.30
TcCLB.505843.24	TcCLB.509649.20	TcCLB.506563.50	TcCLB.503855.20	TcCLB.508541.220	TcCLB.504867.70	TcCLB.509349.30
TcCLB.510323.4	TcCLB.506367.130	TcCLB.503953.14	TcCLB.507875.110	TcCLB.506749.20	TcCLB.507709.80	TcCLB.510755.89
TcCLB.508675.29	TcCLB.510293.30	TcCLB.506943.130	TcCLB.510911.45	TcCLB.510381.10	TcCLB.511545.20	TcCLB.510451.10
TcCLB.507467.94	TcCLB.507673.20	TcCLB.504077.30	TcCLB.508285.50	TcCLB.503579.70	TcCLB.509815.90	TcCLB.508321.21
TcCLB.506147.120	TcCLB.507677.100	TcCLB.509735.30	TcCLB.508473.10	TcCLB.511577.180	TcCLB.509141.40	TcCLB.503453.40
TcCLB.505807.20	TcCLB.510407.20	TcCLB.506181.50	TcCLB.506533.106	TcCLB.510165.110	TcCLB.511685.30	

Supplementary Table 3 List of *T. brucei* genes orthologous to genes exhibiting remarkable proportion of polyA to total RNA transcript in the cytoplasm fraction of epimastigotes cells of *T. cruzi* Dm28c

Data derived from transcriptomes of *T. brucei* (Fadda et al. 2014)

High: *T. brucei* orthologous to *T. cruzi* genes, whose polyA to total cytoplasm RNA transcriptomes ratio is >5 medians, with stability values. Low: *T. brucei* orthologous to *T. cruzi* genes, whose polyA to total cytoplasm RNA transcriptomes ratio is < 0.2 medians, with stability values.

x	[Product Description]	[Input Ortholog(s)]	BS mRNA half-life (min)	PC mRNA half-life (min)
Tb927.10.10000	hypothetical protein, conserved	TcCLB.508307.180	11,44371986	90,73662043
Tb927.10.11440	G1 cyclin CycE4, putative	TcCLB.507677.150	25,66391945	60,31494856
Tb927.10.13430	citrate synthase, putative	TcCLB.509801.30	29,12211418	NA
Tb927.10.15310	hypothetical protein, conserved	TcCLB.506505.10	24,15436745	79,46971178
Tb927.10.3900	protein kinase, putative	TcCLB.511001.60	19,55801964	35,89804888
Tb927.10.5150	zinc finger protein family member, putative	TcCLB.510295.59	7,138719559	5,086061954
Tb927.10.5250	zinc finger protein family member, putative	TcCLB.503795.10	58,89180422	65,94602823
Tb927.10.6050	clathrin heavy chain	TcCLB.506167.50	Stable	122,8546643
Tb927.10.7700	ATP-binding cassette protein subfamily G	TcCLB.506249.70	16,22159004	104,3081117
Tb927.11.12400	hypothetical protein, conserved	TcCLB.511127.300	15,83525658	20,20211935
Tb927.11.15840	L-Lysine transport protein, putative	TcCLB.511545.80	27,89408684	70,56244612
Tb927.11.16310	Unconventional myosin	TcCLB.504103.30	30,61591387	10,61374426
Tb927.11.6120	ABC transporter, putative	TcCLB.510149.80	19,92524147	59,67329264
Tb927.11.6810	tubulin-tyrosine ligase-like protein, putative	TcCLB.506775.30	13,12941551	NA
Tb927.11.8240	TPR repeat, putative	TcCLB.506165.4	23,29632759	NA
Tb927.2.5480	hypothetical protein, conserved	TcCLB.503975.50	25,52384377	0,052449703
Tb927.3.3730	ABC transporter, putative	TcCLB.510149.80	61,87969923	26,45495653
Tb927.3.3960	Double RNA binding domain protein 6A	TcCLB.503683.11	35,22626638	21,00469351
Tb927.3.4720	dynamin-1-like protein	TcCLB.503671.39	26,89360619	45,56990862
Tb927.3.4760	dynamin-1-like protein	TcCLB.503671.39	24,24286842	51,50050879
Tb927.4.2590	hypothetical protein, conserved	TcCLB.508133.19	13,72092247	NA
Tb927.4.4230	RNA-binding protein, putative	TcCLB.505007.10	11,46454811	17,90870905
Tb927.5.810	zinc finger CCCH domain containing protein 11	TcCLB.504929.5	Stable	171,4017463
Tb927.6.4470	Prozyme	TcCLB.509167.110	26,84199333	95,64692259
Tb927.7.1220	hypothetical protein, conserved	TcCLB.508211.50	15,73494911	16,58838987
Tb927.7.2420	glycogen synthase kinase-3 alpha, putative	TcCLB.506857.10	19,27763939	17,66256094
Tb927.7.2680	zinc finger protein family member, putative	TcCLB.506859.240	7,713289261	188,4129167
Tb927.7.5380	RNA-binding protein, putative	TcCLB.506825.10	14,11536217	NA
Tb927.7.960	protein kinase, putative	TcCLB.477079.10, TcCLB.508215.3	28,09710503	59,80957747
Tb927.8.4500	Eukaryotic translation initiation factor 4 gamma 5	TcCLB.508989.90	33,73628855	65,43426275
Tb927.8.4700	amino acid transporter, putative	TcCLB.511545.80	15,63054085	16,66669607
Tb927.8.6650	RNA-binding protein, putative	TcCLB.506399.40	28,22436333	NA
Tb927.9.2240	protein kinase, putative	TcCLB.503409.10	12,0779705	21,1478734
Tb927.9.2730	hypothetical protein, conserved	TcCLB.508347.40	17,5231266	NA

Supplementary Table 4 List of genes with remarkable proportion of polyA to total RNA transcript in nuclear fraction of epimastigote cells of *T. cruzi* Dm28c

High: Genes whose polyA to total nuclear RNA transcriptomes ratio is >5 medians. Low: Genes whose polyA to total nuclear RNA transcriptomes ratio is < 0.2 medians.

Data derived from reported transcriptomes (Smircich *et al.*, 2015) and (Pastro *et al.*, 2017).

High N vs polyA						
TcCLB.508439.40	TcCLB.507099.40	TcCLB.511685.60	TcCLB.508809.30	TcCLB.506825.120	TcCLB.511139.10	TcCLB.510105.200
TcCLB.504703.20	TcCLB.506743.60	TcCLB.510595.4	TcCLB.509267.70	TcCLB.510599.40	TcCLB.506399.20	TcCLB.506375.50
TcCLB.507447.19	TcCLB.511749.9	TcCLB.441061.9	TcCLB.510431.190	TcCLB.508829.9	TcCLB.504835.29	TcCLB.506655.7
TcCLB.504423.40	TcCLB.509805.210	TcCLB.507641.14	TcCLB.511071.130	TcCLB.508777.40	TcCLB.509769.130	TcCLB.509247.60
TcCLB.511441.10	TcCLB.507979.40	TcCLB.506591.120	TcCLB.511189.70	TcCLB.506925.150	TcCLB.506753.230	TcCLB.508317.40
TcCLB.506673.69	TcCLB.510855.4	TcCLB.508307.140	TcCLB.509927.10	TcCLB.509925.70	TcCLB.506503.80	TcCLB.506543.10
TcCLB.506973.140	TcCLB.506625.60	TcCLB.505007.30	TcCLB.507809.14	TcCLB.509809.39	TcCLB.506205.10	TcCLB.511445.160
TcCLB.484311.10	TcCLB.506925.170	TcCLB.510741.70	TcCLB.506963.70	TcCLB.511337.4	TcCLB.508779.90	TcCLB.511313.40
TcCLB.511253.31	TcCLB.511113.9	TcCLB.509455.9	TcCLB.508277.360	TcCLB.506685.50	TcCLB.511421.224	TcCLB.507603.120
TcCLB.510283.10	TcCLB.508723.70	TcCLB.504107.10	TcCLB.508239.30	TcCLB.511127.390	TcCLB.506967.70	TcCLB.511545.100
TcCLB.510533.210	TcCLB.506129.30	TcCLB.509067.60	TcCLB.507083.109	TcCLB.506531.9	TcCLB.509133.10	TcCLB.508137.40
TcCLB.507603.260	TcCLB.503989.50	TcCLB.503895.19	TcCLB.508831.150	TcCLB.507711.160	TcCLB.508325.30	TcCLB.507841.30
TcCLB.441241.10	TcCLB.510769.80	TcCLB.503629.10	TcCLB.511421.80	TcCLB.511131.30	TcCLB.511365.90	TcCLB.510231.10
TcCLB.412093.10	TcCLB.506401.350	TcCLB.506697.20	TcCLB.507467.10	TcCLB.506885.210	TcCLB.510073.80	TcCLB.506891.10
TcCLB.510247.10	TcCLB.508831.4	TcCLB.509979.40	TcCLB.507477.20	TcCLB.509805.100	TcCLB.510251.19	TcCLB.506525.70
TcCLB.508837.80	TcCLB.507275.70	TcCLB.508717.60	TcCLB.509205.90	TcCLB.503813.10	TcCLB.504149.20	TcCLB.507807.10

TcCLB.510531.109	TcCLB.445635.10	TcCLB.510771.9	TcCLB.508347.90	TcCLB.506625.160	TcCLB.511621.240	TcCLB.509669.50
TcCLB.511443.10	TcCLB.511611.11	TcCLB.503751.30	TcCLB.433273.10	TcCLB.506859.150	TcCLB.511621.60	TcCLB.509007.70
TcCLB.404001.20	TcCLB.509233.120	TcCLB.506941.9	TcCLB.506529.480	TcCLB.507277.20	TcCLB.511211.230	TcCLB.511733.120
TcCLB.510877.10	TcCLB.511327.20	TcCLB.506811.180	TcCLB.510247.20	TcCLB.508163.360	TcCLB.505983.40	TcCLB.506635.30
TcCLB.507933.10	TcCLB.510055.30	TcCLB.506413.89	TcCLB.503419.10	TcCLB.507551.9	TcCLB.507737.10	TcCLB.503901.18
TcCLB.504117.50	TcCLB.509775.30	TcCLB.428999.10	TcCLB.507993.230	TcCLB.506829.90	TcCLB.504117.9	TcCLB.509151.90
TcCLB.508097.10	TcCLB.511365.4	TcCLB.506529.600	TcCLB.511127.369	TcCLB.511325.50	TcCLB.503651.4	TcCLB.511127.60
TcCLB.508347.159	TcCLB.504883.70	TcCLB.507831.110	TcCLB.508921.30	TcCLB.507627.9	TcCLB.505977.13	TcCLB.506739.180
TcCLB.511253.20	TcCLB.509589.50	TcCLB.508277.100	TcCLB.511315.9	TcCLB.504425.110	TcCLB.511051.10	TcCLB.507925.20
TcCLB.506925.550	TcCLB.511555.110	TcCLB.507809.140	TcCLB.506749.10	TcCLB.508731.60	TcCLB.510055.70	TcCLB.503531.49
TcCLB.445281.9	TcCLB.510525.90	TcCLB.506407.90	TcCLB.506773.10	TcCLB.509965.70	TcCLB.509153.40	TcCLB.511867.200
TcCLB.510773.9	TcCLB.506717.20	TcCLB.510565.179	TcCLB.509595.20	TcCLB.510125.10	TcCLB.510105.120	TcCLB.510285.80
TcCLB.509891.70	TcCLB.422041.9	TcCLB.509137.20	TcCLB.511427.19	TcCLB.505555.10	TcCLB.503981.80	TcCLB.507795.10
TcCLB.509599.20	TcCLB.448567.9	TcCLB.504087.20	TcCLB.504147.250	TcCLB.506647.20	TcCLB.511617.9	TcCLB.507757.10
TcCLB.506543.50	TcCLB.506533.34	TcCLB.507627.20	TcCLB.510143.50	TcCLB.510101.290	TcCLB.510311.100	TcCLB.510989.20
TcCLB.509013.19	TcCLB.508973.80	TcCLB.510509.9	TcCLB.511215.70	TcCLB.509149.60	TcCLB.504213.90	TcCLB.506825.180
TcCLB.511253.10	TcCLB.506949.20	TcCLB.506465.40	TcCLB.503955.100	TcCLB.511133.30	TcCLB.509167.180	TcCLB.511867.130
TcCLB.507465.10	TcCLB.509893.10	TcCLB.480997.10	TcCLB.506637.10	TcCLB.510769.10	TcCLB.509805.260	TcCLB.508129.9
TcCLB.507083.20	TcCLB.504421.10	TcCLB.506203.40	TcCLB.405165.19	TcCLB.504153.150	TcCLB.510515.20	TcCLB.507011.230
TcCLB.511527.9	TcCLB.507805.30	TcCLB.510309.70	TcCLB.511417.80	TcCLB.510747.60	TcCLB.504769.160	TcCLB.509993.51
TcCLB.506849.20	TcCLB.506375.100	TcCLB.511535.10	TcCLB.503579.10	TcCLB.504255.30	TcCLB.506227.240	TcCLB.507677.129
TcCLB.509173.9	TcCLB.510557.73	TcCLB.506213.120	TcCLB.509449.10	TcCLB.509153.139	TcCLB.506401.140	TcCLB.503755.30
TcCLB.506859.204	TcCLB.510773.40	TcCLB.405179.10	TcCLB.434931.20	TcCLB.506401.170	TcCLB.503855.50	TcCLB.511751.70
TcCLB.504625.70	TcCLB.462617.10	TcCLB.510863.9	TcCLB.509009.10	TcCLB.509967.150	TcCLB.507035.30	TcCLB.506863.80
TcCLB.509583.19	TcCLB.507737.40	TcCLB.506503.90	TcCLB.503967.19	TcCLB.510773.30	TcCLB.506203.120	TcCLB.509631.140
TcCLB.511043.60	TcCLB.510535.20	TcCLB.506531.60	TcCLB.482097.10	TcCLB.508089.10	TcCLB.506697.80	TcCLB.511181.110
TcCLB.510609.60	TcCLB.409139.9	TcCLB.509247.30	TcCLB.503697.30	TcCLB.510307.10	TcCLB.507941.100	TcCLB.510533.10
TcCLB.506401.12	TcCLB.509473.10	TcCLB.509667.31	TcCLB.506691.30	TcCLB.510057.4	TcCLB.505965.10	TcCLB.474341.9
TcCLB.509269.4	TcCLB.507237.320	TcCLB.504449.14	TcCLB.504111.49	TcCLB.509667.10	TcCLB.511315.44	TcCLB.509871.80
TcCLB.408573.9	TcCLB.509585.10	TcCLB.511467.20	TcCLB.510055.150	TcCLB.510073.30	TcCLB.510349.40	TcCLB.511911.81
TcCLB.511557.50	TcCLB.508181.4	TcCLB.511001.90	TcCLB.508549.30	TcCLB.508783.10	TcCLB.439803.9	TcCLB.506179.59
TcCLB.504053.10	TcCLB.506763.210	TcCLB.506623.30	TcCLB.509819.10	TcCLB.511709.19	TcCLB.508165.130	TcCLB.507477.10
TcCLB.506487.24	TcCLB.511695.10	TcCLB.506737.151	TcCLB.507099.80	TcCLB.504089.14	TcCLB.506289.170	TcCLB.508805.30
TcCLB.511431.10	TcCLB.506895.80	TcCLB.504137.50	TcCLB.509773.9	TcCLB.480785.10	TcCLB.511585.230	TcCLB.506227.250
TcCLB.511751.14	TcCLB.509195.10	TcCLB.506467.40	TcCLB.509775.20	TcCLB.510317.7	TcCLB.505073.10	TcCLB.506519.140
TcCLB.509013.10	TcCLB.506925.10	TcCLB.508569.70	TcCLB.503955.50	TcCLB.509669.20	TcCLB.426397.9	TcCLB.506223.120
TcCLB.506811.160	TcCLB.398751.10	TcCLB.506937.40	TcCLB.511313.10	TcCLB.511115.10	TcCLB.507641.40	TcCLB.479379.9
TcCLB.506859.230	TcCLB.506401.290	TcCLB.510861.14	TcCLB.507837.100	TcCLB.506961.25	TcCLB.504035.149	TcCLB.510719.80
TcCLB.507719.9	TcCLB.507179.20	TcCLB.510159.19	TcCLB.504247.79	TcCLB.507929.10	TcCLB.506563.160	TcCLB.507839.40
TcCLB.509103.39	TcCLB.511693.30	TcCLB.510449.10	TcCLB.509171.100	TcCLB.506599.10	TcCLB.511127.320	TcCLB.507639.110
TcCLB.503479.70	TcCLB.510755.19	TcCLB.511425.49	TcCLB.508609.20	TcCLB.509351.6	TcCLB.510075.60	TcCLB.511211.10
TcCLB.504133.15	TcCLB.511001.160	TcCLB.511869.9	TcCLB.503627.20	TcCLB.507611.131	TcCLB.511873.10	TcCLB.511559.70
TcCLB.504153.90	TcCLB.509537.50	TcCLB.510815.9	TcCLB.511705.10	TcCLB.503697.100	TcCLB.503675.40	TcCLB.507277.30
TcCLB.508209.10	TcCLB.508997.20	TcCLB.509893.20	TcCLB.511407.50	TcCLB.508567.100	TcCLB.503915.9	TcCLB.507389.30
TcCLB.506525.120	TcCLB.434945.10	TcCLB.508719.9	TcCLB.511215.20	TcCLB.508551.30	TcCLB.506357.80	TcCLB.506683.240
TcCLB.511435.49	TcCLB.510603.130	TcCLB.507011.250	TcCLB.430061.9	TcCLB.479253.9	TcCLB.507867.80	TcCLB.504125.70
TcCLB.511445.10	TcCLB.511181.160	TcCLB.506375.70	TcCLB.509007.40	TcCLB.510379.20	TcCLB.503613.10	TcCLB.506529.595
TcCLB.506401.24	TcCLB.511141.10	TcCLB.507905.30	TcCLB.508811.30	TcCLB.506357.130	TcCLB.510533.190	TcCLB.510859.50
TcCLB.511867.209	TcCLB.510409.10	TcCLB.510747.170	TcCLB.510297.10	TcCLB.506527.4	TcCLB.506401.110	TcCLB.507531.19
TcCLB.506251.130	TcCLB.475705.9	TcCLB.509007.20	TcCLB.425311.9	TcCLB.510859.40	TcCLB.506181.20	TcCLB.508181.70
TcCLB.438059.10	TcCLB.442427.10	TcCLB.507677.4	TcCLB.507257.50	TcCLB.509151.120	TcCLB.510311.60	TcCLB.507669.50
TcCLB.506907.14	TcCLB.508995.10	TcCLB.503637.20	TcCLB.507631.10	TcCLB.506825.40	TcCLB.506493.90	TcCLB.511131.80

TcCLB.485889.16	TcCLB.507875.10					
Low N vs polyA						
TcCLB.510073.17	TcCLB.510101.190	TcCLB.506649.80	TcCLB.508549.10	TcCLB.419417.4	TcCLB.506931.4	TcCLB.503453.30
TcCLB.506811.120	TcCLB.506457.20	TcCLB.509769.36	TcCLB.505843.20	TcCLB.506945.290	TcCLB.507485.60	TcCLB.503779.20
TcCLB.511145.60	TcCLB.511555.40	TcCLB.509965.200	TcCLB.508923.10	TcCLB.511139.20	TcCLB.503893.70	TcCLB.511071.190
TcCLB.506605.200	TcCLB.506949.50	TcCLB.506885.430	TcCLB.510615.4	TcCLB.511001.100	TcCLB.503953.14	TcCLB.506779.60
TcCLB.508185.10	TcCLB.506953.30	TcCLB.511711.100	TcCLB.511071.171	TcCLB.503689.4	TcCLB.506401.120	TcCLB.506205.20
TcCLB.504147.70	TcCLB.508389.10	TcCLB.511627.60	TcCLB.510523.29	TcCLB.509599.100	TcCLB.511729.40	TcCLB.510429.50
TcCLB.510737.40	TcCLB.507603.70	TcCLB.510515.40	TcCLB.504163.80	TcCLB.508285.20	TcCLB.511145.20	TcCLB.510243.40
TcCLB.504035.70	TcCLB.510525.40	TcCLB.506413.80	TcCLB.507093.140	TcCLB.506753.150	TcCLB.508805.80	TcCLB.509267.60
TcCLB.438147.18	TcCLB.510073.70	TcCLB.511867.90	TcCLB.511729.30	TcCLB.510741.20	TcCLB.507389.50	TcCLB.504147.120
TcCLB.506811.140	TcCLB.509599.40	TcCLB.506777.70	TcCLB.506473.14	TcCLB.511867.100	TcCLB.511589.50	TcCLB.507611.389
TcCLB.432027.9	TcCLB.508075.7	TcCLB.506407.110	TcCLB.508277.120	TcCLB.482471.10	TcCLB.506341.30	TcCLB.509011.90
TcCLB.507521.50	TcCLB.504017.10	TcCLB.508723.100	TcCLB.506637.120	TcCLB.506811.90	TcCLB.511407.70	TcCLB.506739.99
TcCLB.506495.6	TcCLB.510603.80	TcCLB.507611.10	TcCLB.510761.18	TcCLB.506405.149	TcCLB.506945.240	TcCLB.504033.80
TcCLB.510089.39	TcCLB.509805.30	TcCLB.507013.30	TcCLB.510293.30	TcCLB.469435.4	TcCLB.510349.50	TcCLB.506931.10
TcCLB.507625.200	TcCLB.506945.210	TcCLB.504431.118	TcCLB.511365.94	TcCLB.506811.150	TcCLB.504163.50	TcCLB.506885.14
TcCLB.510743.30	TcCLB.508955.10	TcCLB.508039.70	TcCLB.407477.69	TcCLB.503991.30	TcCLB.511041.10	TcCLB.508321.21
TcCLB.506779.120	TcCLB.506979.10	TcCLB.507711.50	TcCLB.510165.20	TcCLB.506887.100	TcCLB.508515.20	TcCLB.508727.18
TcCLB.506677.10	TcCLB.506943.23	TcCLB.508241.20	TcCLB.511739.10	TcCLB.510971.5	TcCLB.509967.10	TcCLB.511577.180
TcCLB.507083.50	TcCLB.509497.30	TcCLB.504179.20	TcCLB.503527.4	TcCLB.503893.90	TcCLB.510737.79	TcCLB.508717.40
TcCLB.504011.40	TcCLB.510101.240	TcCLB.511529.210	TcCLB.506753.140	TcCLB.506405.110	TcCLB.508995.60	TcCLB.509591.10
TcCLB.506865.10	TcCLB.510285.74	TcCLB.511355.30	TcCLB.506375.80	TcCLB.506775.140	TcCLB.511361.50	TcCLB.510381.10
TcCLB.507611.270	TcCLB.511627.130	TcCLB.506473.6	TcCLB.447255.10	TcCLB.511539.18	TcCLB.504255.50	TcCLB.511469.20
TcCLB.506691.70	TcCLB.506519.70	TcCLB.506147.60	TcCLB.511543.60	TcCLB.510057.30	TcCLB.506563.79	TcCLB.506529.324
TcCLB.506251.120	TcCLB.506657.50	TcCLB.510717.40	TcCLB.509067.40	TcCLB.509805.190	TcCLB.506289.60	TcCLB.503449.14
TcCLB.507037.10	TcCLB.507963.44	TcCLB.504643.20	TcCLB.506801.80	TcCLB.503899.130	TcCLB.507677.70	TcCLB.508207.70
TcCLB.503893.30	TcCLB.507933.29	TcCLB.506563.50	TcCLB.508037.10	TcCLB.506755.250	TcCLB.506211.250	TcCLB.510165.110
TcCLB.506869.40	TcCLB.511369.5	TcCLB.508137.30	TcCLB.503779.40	TcCLB.508785.4	TcCLB.509979.281	TcCLB.508943.40
TcCLB.510763.90	TcCLB.506929.30	TcCLB.506213.80	TcCLB.510259.50	TcCLB.511469.40	TcCLB.506553.10	TcCLB.506413.30
TcCLB.508357.20	TcCLB.509171.80	TcCLB.511127.274	TcCLB.506925.300	TcCLB.511071.80	TcCLB.505807.20	TcCLB.508647.110
TcCLB.511421.90	TcCLB.503449.10	TcCLB.507809.39	TcCLB.510101.90	TcCLB.506529.50	TcCLB.508285.50	TcCLB.508183.20
TcCLB.510173.70	TcCLB.508241.80	TcCLB.507099.130	TcCLB.507257.120	TcCLB.509069.20	TcCLB.509499.14	TcCLB.510855.10
TcCLB.507143.20	TcCLB.510759.60	TcCLB.507689.50	TcCLB.509551.70	TcCLB.510259.30	TcCLB.508265.10	TcCLB.508727.54
TcCLB.507521.110	TcCLB.507257.150	TcCLB.511127.40	TcCLB.507709.80	TcCLB.506567.90	TcCLB.506859.10	TcCLB.506943.130
TcCLB.506859.130	TcCLB.510243.80	TcCLB.504163.100	TcCLB.507091.30	TcCLB.504203.40	TcCLB.510407.20	TcCLB.509649.20
TcCLB.506925.130	TcCLB.509147.50	TcCLB.504867.70	TcCLB.503899.119	TcCLB.505989.110	TcCLB.509233.190	TcCLB.507467.94
TcCLB.509205.60	TcCLB.511685.30	TcCLB.504077.30	TcCLB.509233.150	TcCLB.506265.50	TcCLB.511067.20	TcCLB.509499.29
TcCLB.509129.10	TcCLB.509141.40	TcCLB.407477.40	TcCLB.506963.14	TcCLB.504005.54	TcCLB.503579.70	TcCLB.506749.20
TcCLB.506413.20	TcCLB.508675.29	TcCLB.507105.40	TcCLB.510029.70	TcCLB.511145.43	TcCLB.510145.20	TcCLB.504089.60
TcCLB.506495.9	TcCLB.505843.24	TcCLB.506945.200	TcCLB.484299.10	TcCLB.506943.69	TcCLB.503453.40	TcCLB.506367.130
TcCLB.506529.40	TcCLB.510295.30	TcCLB.507093.339	TcCLB.511369.15	TcCLB.507105.50	TcCLB.507611.280	TcCLB.510609.20
TcCLB.509157.220	TcCLB.509815.90	TcCLB.511867.140	TcCLB.506533.106	TcCLB.504089.70	TcCLB.508781.120	TcCLB.508541.220
TcCLB.508723.90	TcCLB.507673.20	TcCLB.506167.40	TcCLB.506655.20	TcCLB.510089.210	TcCLB.504433.10	
TcCLB.508473.10	TcCLB.510755.50	TcCLB.508207.110	TcCLB.511545.20	TcCLB.510755.89	TcCLB.506181.50	

Bibliography

- Agabian, N. (1990) 'Trans splicing of nuclear pre-mRNAs', *Cell*, 1990/06/29, 61(7), pp. 1157–1160. doi: 0092-8674(90)90674-4 [pii].
- Akef, A., Lee, E. S. and Palazzo, A. F. (no date) 'Splicing promotes the nuclear export of β -globin mRNA by overcoming nuclear retention elements'. doi: 10.1261/rna.051987.115.
- Allison, D. B., Cui, X., Page, G. P. and Sabripour, M. (2006) 'Microarray data analysis: from disarray to consolidation and consensus', *Nature Reviews Genetics*, 7(1), pp. 55–65. doi: 10.1038/nrg1749.
- Alvarez, F., Robello, C. and Vignali, M. (1994) 'Evolution of codon usage and base contents in kinetoplastid protozoans.', *Molecular biology and evolution*, 11(5), pp. 790–802. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7968492> (Accessed: 9 March 2017).
- Alves, L. R., Oliveira, C. and Goldenberg, S. (2015) 'Eukaryotic translation elongation factor-1 alpha is associated with a specific subset of mRNAs in *Trypanosoma cruzi*', *BMC Microbiol*, 15, p. 104. doi: 10.1186/s12866-015-0436-2.
- Alves, L. R., Oliveira, C., Morking, P. A., Kessler, R. L., Martins, S. T., Romagnoli, B. A., Marchini, F. K. and Goldenberg, S. (2014) 'The mRNAs associated to a zinc finger protein from *Trypanosoma cruzi* shift during stress conditions', *RNA Biol*, 11(7), pp. 921–933. doi: 10.4161/rna.29622.
- Antwi, E. B., Haanstra, J. R., Ramasamy, G., Jensen, B., Droll, D., Rojas, F., Minia, I., Terrao, M., Mercé, C., Matthews, K., Myler, P. J., Parsons, M., Clayton, C., Merce, C., Matthews, K., Myler, P. J., Parsons, M. and Clayton, C. (2016) 'Integrative analysis of the *Trypanosoma brucei* gene expression cascade predicts differential regulation of mRNA processing and unusual control of ribosomal protein expression', *BMC Genomics*, 17(1), p. 306. doi: 10.1186/s12864-016-2624-3.
- Araújo, F. S., Coelho, L. M., Silva, L. do C., da Silva Neto, B. R., Parente-Rocha, J. A., Bailão, A. M., de Oliveira, C. M. A., Fernandes, G. da R., Hernández, O., Ochoa, J. G. M., Soares, C. M. de A., Pereira, M. and Pereira, M. (2016) 'Effects of Argentilactone on the Transcriptional Profile, Cell Wall and Oxidative Stress of *Paracoccidioides* spp.', *PLOS Neglected Tropical Diseases*. Edited by J. M. Vinetz. Public Library of Science, 10(1), p. e0004309. doi: 10.1371/journal.pntd.0004309.
- Araújo, P. R., Teixeira, S. M., Araujo, P. R. and Teixeira, S. M. (2011) 'Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review', *Mem Inst Oswaldo Cruz*. 2011/06/10, 106(3), pp. 257–266. doi: S0074-02762011000300002 [pii].
- Archer, S. K., Luu, V. D., de Queiroz, R. A., Brems, S. and Clayton, C. (2009) '*Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle', *PLoS Pathog*. 2009/08/29, 5(8), p. e1000565. doi: 10.1371/journal.ppat.1000565.
- Arner, E., Kindlund, E., Nilsson, D., Farzana, F., Ferella, M., Tammi, M. T. and Andersson, B. (2007) 'Database of *Trypanosoma cruzi* repeated genes: 20,000 additional gene variants', *BMC Genomics*, 8, p. 391. doi: 10.1186/1471-2164-8-391.
- Atayde, V. D., Tschudi, C. and Ullu, E. (2011) 'The emerging world of small silencing RNAs in protozoan parasites.', *Trends in parasitology*, 27(7), pp. 321–7. doi: 10.1016/j.pt.2011.03.002.
- Atwood, J. A., Weatherly, D. B., Minning, T. A., Bundy, B., Cavola, C., Opperdoes, F. R., Orlando, R. and Tarleton, R. L. (2005) 'The *Trypanosoma cruzi* proteome.', *Science (New York, N.Y.)*, 309(5733), pp. 473–6. doi: 10.1126/science.1110289.
- Avila, A. R., Yamada-Ogatta, S. F., da Silva Monteiro, V., Krieger, M. A., Nakamura, C. V., de Souza, W. and Goldenberg, S. (2001) 'Cloning and characterization of the metacyclogenin gene, which is specifically expressed during *Trypanosoma cruzi* metacyclogenesis', *Mol Biochem Parasitol*, 117(2), pp. 169–177. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11606227.
- Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., Itzkovitz, S., Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I. and Itzkovitz, S. (2015) 'Nuclear Retention of mRNA in Mammalian Tissues', *Cell Rep*, 13(12), pp. 2653–2662. doi: 10.1016/j.celrep.2015.11.036.
- Bailey, T. L. (2011) 'DREME: motif discovery in transcription factor ChIP-seq data.', *Bioinformatics (Oxford, England)*. Oxford University Press, 27(12), pp. 1653–9. doi: 10.1093/bioinformatics/btr261.
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W.

- and Noble, W. S. (2009) 'MEME SUITE: tools for motif discovery and searching.', *Nucleic acids research*, 37(Web Server issue), pp. W202-8. doi: 10.1093/nar/gkp335.
- Balagopal, V. and Parker, R. (2009) 'Stm1 modulates mRNA decay and Dhh1 function in *Saccharomyces cerevisiae*.' *Genetics*, 181(1), pp. 93–103. doi: 10.1534/genetics.108.092601.
- Bandziulis, R. J., Swanson, M. S. and Dreyfuss, G. (1989) 'RNA-binding proteins as developmental regulators', *Genes Dev*, 3(4), pp. 431–437. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2470643.
- Bartholomeu, D. C., Cerqueira, G. C., Leao, A. C., daRocha, W. D., Pais, F. S., Macedo, C., Djikeng, A., Teixeira, S. M. and El-Sayed, N. M. (2009) 'Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*', *Nucleic Acids Res.* 2009/04/02, 37(10), pp. 3407–3417. doi: gkp172 [pii]10.1093/nar/gkp172.
- Bayer-Santos, E., Gentil, L. G., Cordero, E. M., Correa, P. R., da Silveira, J. F., Corrêa, P. R. C. and da Silveira, J. F. (2012) 'Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes', *Acta Trop*, 123(3), pp. 230–233. doi: 10.1016/j.actatropica.2012.03.014.
- Bern, C., Kjos, S., Yabsley, M. J. and Montgomery, S. P. (2011) 'Trypanosoma cruzi and Chagas' Disease in the United States', *Clin Microbiol Rev*, 24(4), pp. 655–681. doi: 10.1128/CMR.00005-11.
- Berná, L., Chiribao, M. L., Greif, G., Rodriguez, M., Alvarez-Valin, F. and Robello, C. (2017) 'Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*'. PeerJ Inc. doi: 10.7287/PEERJ.PREPRINTS.2782V1.
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renaud, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E. and El-Sayed, N. M. (2005) 'The genome of the African trypanosome *Trypanosoma brucei*', *Science*. 2005/07/16, 309(5733), pp. 416–422. doi: 10.1126/science.1112642.
- Blackwell, J. M. and Melville, S. E. (1999) 'Status of protozoan genome analysis: trypanosomatids', *Parasitology*, 118 Suppl, pp. S11-4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10466133>.
- Boiani, M., Piacenza, L., Hernández, P., Boiani, L., Cerecetto, H., González, M. and Denicola, A. (2010) 'Mode of action of Nifurtimox and N-oxide-containing heterocycles against *Trypanosoma cruzi*: Is oxidative stress involved?', *Biochemical Pharmacology*, 79(12), pp. 1736–1745. doi: 10.1016/j.bcp.2010.02.009.
- Boucher, N., Wu, Y., Dumas, C., Dube, M., Sereno, D., Breton, M. and Papadopoulos, B. (2002) 'A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element', *J Biol Chem*, 277(22), pp. 19511–19520. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11912202.
- Boulon, S., Westman, B. J., Hutten, S., Boisvert, F.-M. and Lamond, A. I. (2010) 'The nucleolus under stress.', *Molecular cell*. Elsevier, 40(2), pp. 216–27. doi: 10.1016/j.molcel.2010.09.024.
- Branche, C., Ochaya, S., Aslund, L. and Andersson, B. (2006) 'Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*', *Mol Biochem Parasitol*, 147(1), pp. 30–38. doi: 10.1016/j.molbiopara.2006.01.005.
- Brandão, A. and Jiang, T. (2009) 'The composition of untranslated regions in *Trypanosoma*

- cruzi genes.', *Parasitology international*. Elsevier Ireland Ltd, 58(3), pp. 215–9. doi: 10.1016/j.parint.2009.06.001.
- Bringaud, F., Biteau, N., Melville, S. E., Hez, S., El-Sayed, N. M., Leech, V., Berriman, M., Hall, N., Donelson, J. E. and Baltz, T. (2002) 'A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*', *Eukaryot Cell*, 1(1), pp. 137–151. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12455980>.
- Buscaglia, C. A., Campo, V. A., Frasc, A. C. and Di Noia, J. M. (2006) 'Trypanosoma cruzi surface mucins: host-dependent coat diversity', *Nat Rev Microbiol*, 4(3), pp. 229–236. doi: 10.1038/nrmicro1351.
- Buscaglia, C. A., Campo, V. A., Di Noia, J. M., Torrecilhas, A. C. T., De Marchi, C. R., Ferguson, M. A. J., Frasc, A. C. C. and Almeida, I. C. (2004) 'The Surface Coat of the Mammal-dwelling Infective Trypomastigote Stage of *Trypanosoma cruzi* Is Formed by Highly Diverse Immunogenic Mucins', *Journal of Biological Chemistry*, 279(16), pp. 15860–15869. doi: 10.1074/jbc.M314051200.
- Camara Mde, L., Bouvier, L. A., Canepa, G. E., Miranda, M. R., Pereira, C. A., Cámara, M. de los M., Bouvier, L. A., Canepa, G. E., Miranda, M. R. and Pereira, C. A. (2013) 'Molecular and functional characterization of a *Trypanosoma cruzi* nuclear adenylate kinase isoform', *PLoS Negl Trop Dis*. Edited by C. A. Buscaglia, 7(2), p. e2044. doi: 10.1371/journal.pntd.0002044.
- Campo, V., Di Noia, J. M., Buscaglia, C. A., Agüero, F., Sanchez, D. O. and Frasc, A. C. (2004) 'Differential accumulation of mutations localized in particular domains of the mucin genes expressed in the vertebrate host stage of *Trypanosoma cruzi*', *Mol Biochem Parasitol*, 133(1), pp. 81–91. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14668015.
- Campos, P. C., Bartholomeu, D. C., DaRocha, W. D., Cerqueira, G. C. and Teixeira, S. M. R. (2008) 'Sequences involved in mRNA processing in *Trypanosoma cruzi*', *Int J Parasitol*. 2008/08/14, 38(12), pp. 1383–1389. doi: S0020-7519(08)00232-4 [pii]10.1016/j.ijpara.2008.07.001.
- Cantey, P. T., Stramer, S. L., Townsend, R. L., Kamel, H., Ofafa, K., Todd, C. W., Currier, M., Hand, S., Varnado, W., Dotson, E., Hall, C., Jett, P. L. and Montgomery, S. P. (2012) 'The United States *Trypanosoma cruzi* Infection Study: evidence for vector-borne transmission of the parasite that causes Chagas disease among United States blood donors', *Transfusion*, 52(9), pp. 1922–1930. doi: 10.1111/j.1537-2995.2012.03581.x.
- Cassola, A. (2011) 'RNA Granules Living a Post-transcriptional Life: the Trypanosomes' Case', *Curr Chem Biol*, 5(2), pp. 108–117. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21949551>.
- Cassola, A. and Frasc, A. C. (2009) 'An RNA recognition motif mediates the nucleocytoplasmic transport of a trypanosome RNA-binding protein', *J Biol Chem*. 2009/10/06, 284(50), pp. 35015–35028. doi: M109.031633 [pii]10.1074/jbc.M109.031633.
- Cassola, A., De Gaudenzi, J. G. and Frasc, A. C. (2007) 'Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes', *Mol Microbiol*. 2007/07/20, 65(3), pp. 655–670. doi: MMI5833 [pii]10.1111/j.1365-2958.2007.05833.x.
- Cavalier-Smith, T. (2010) 'Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree', *Biology Letters*, 6(3), pp. 342–345. doi: 10.1098/rsbl.2009.0948.
- Cazzulo, J. J. (1994) 'Intermediate metabolism in *Trypanosoma cruzi*', *J Bioenerg Biomembr*, 26(2), pp. 157–165. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8056782>.
- Chagas, C. (1909) 'Nova tripanozomiaze humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi*, agente etiologico da nova entidade mórbida do homen.', *Mem. Inst. Oswaldo Cruz*, 1, pp. 159–219.
- Chávez, S., Beilharz, T., Rondón, A. G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J. Q., Lithgow, T. and Aguilera, A. (2000) 'A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*.' *The EMBO journal*. European Molecular Biology Organization, 19(21), pp. 5824–34. doi: 10.1093/emboj/19.21.5824.
- Choi, Y.-J., Aliota, M. T., Mayhew, G. F., Erickson, S. M. and Christensen, B. M. (2014) 'Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions.' *PLoS neglected tropical diseases*. Public Library of Science, 8(5), p. e2905. doi: 10.1371/journal.pntd.0002905.
- Clayton, C. (2013) 'The regulation of trypanosome gene expression by RNA-binding proteins',

- PLoS Pathog*, 9(11), p. e1003680. doi: 10.1371/journal.ppat.1003680.
- Clayton, C. E. (2002) 'Life without transcriptional control? From fly to man and back again', *EMBO J*, 21(8), pp. 1881–1888. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11953307.
- Coller, J. M., Tucker, M., Sheth, U., Valencia-Sanchez, M. A. and Parker, R. (2001) 'The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes', *RNA*, 2002/01/10, 7(12), pp. 1717–1727. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11780629.
- Coller, J. and Parker, R. (2005) 'General translational repression by activators of mRNA decapping', *Cell*, 2005/09/24, 122(6), pp. 875–886. doi: S0092-8674(05)00706-3 [pii]10.1016/j.cell.2005.07.012.
- Coller, J. and Wickens, M. (2007) 'Tethered function assays: an adaptable approach to study RNA regulatory proteins.', *Methods in enzymology*, 429(7), pp. 299–321. doi: 10.1016/S0076-6879(07)29014-7.
- Conesa, A. and Gotz, S. (2008) 'Blast2GO: A comprehensive suite for functional analysis in plant genomics', *Int J Plant Genomics*, 2008, p. 619832. doi: 10.1155/2008/619832.
- da Costa Lima, T. D., Moura, D. M. N., Reis, C. R. S., Vasconcelos, J. R. C., Ellis, L., Carrington, M., Figueiredo, R. C. B. Q. and de Melo Neto, O. P. (2010) 'Functional characterization of three Leishmania poly(A) binding protein homologues with distinct binding properties to RNA and protein partners', *Eukaryotic Cell*, 9(10), pp. 1484–1494. doi: 10.1128/EC.00148-10.
- Coughlin, B. C., Teixeira, S. M., Kirchhoff, L. V and Donelson, J. E. (2000) 'Amastin mRNA abundance in Trypanosoma cruzi is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein', *J Biol Chem*, 275(16), pp. 12051–12060. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10766837.
- Cuevas, I. C., Cazzulo, J. J. and Sanchez, D. O. (2003) 'gp63 homologues in Trypanosoma cruzi: surface antigens with metalloprotease activity and a possible role in host cell infection', *Infect Immun*, 2003/09/23, 71(10), pp. 5739–5749. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14500495.
- Cuevas, I. C., Frasch, A. C. and D'Orso, I. (2005) 'Insights into a CRM1-mediated RNA-nuclear export pathway in Trypanosoma cruzi', *Mol Biochem Parasitol*, 139(1), pp. 15–24. doi: 10.1016/j.molbiopara.2004.11.002.
- D'Orso, I. and Frasch, A. C. (2001a) 'Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in Trypanosoma cruzi by interaction with specific RNA-binding proteins', *J Biol Chem*, 276(19), pp. 15783–15793. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11278796.
- D'Orso, I. and Frasch, A. C. (2001b) 'TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes', *J Biol Chem*, 276(37), pp. 34801–34809. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11435421.
- D'Orso, I. and Frasch, A. C. C. C. (2002) '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', *J Biol Chem*, 277(52), pp. 50520–50528. doi: 10.1074/jbc.M209092200.
- D'Orso, I., De Gaudenzi, J. G. and Frasch, A. C. (2003) 'RNA-binding proteins and mRNA turnover in trypanosomes', *Trends Parasitol*, 19(4), pp. 151–155. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12689640.
- Dallagiovanna, B., Correa, A., Probst, C. M., Holetz, F., Smircich, P., de Aguiar, A. M., Mansur, F., da Silva, C. V, Mortara, R. A., Garat, B., Buck, G. A., Goldenberg, S. and Krieger, M. A. (2008) 'Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from Trypanosoma cruzi', *J Biol Chem*, 2007/12/07, 283(13), pp. 8266–8273. doi: M703097200 [pii]10.1074/jbc.M703097200.

- Dallagiovanna, B., Perez, L., Sotelo-Silveira, J., Smircich, P., Duhagon, M. A. and Garat, B. (2005) 'Trypanosoma cruzi: Molecular characterization of TcPUF6, a Pumilio protein', *Exp Parasitol*, 109(4), pp. 260–264. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15755425.
- Das, A., Bellofatto, V., Rosenfeld, J., Carrington, M., Romero-Zaliz, R., del Val, C. and Estévez, A. M. (2015) *High throughput sequencing analysis of Trypanosoma brucei DRBD3/PTB1-bound mRNAs*, *Molecular and Biochemical Parasitology*. doi: 10.1016/j.molbiopara.2015.02.003.
- Dean, S., Sunter, J. D. and Wheeler, R. J. (2017) 'TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource.', *Trends in parasitology*. Elsevier, 33(2), pp. 80–82. doi: 10.1016/j.pt.2016.10.009.
- Dennis Jr., G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. and Lempicki, R. A. (2003) 'DAVID: Database for Annotation, Visualization, and Integrated Discovery', *Genome Biol*, 4(5), p. P3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12734009>.
- Díaz de Toranzo, E. G., Castro, J. A., Franke de Cazzulo, B. M. and Cazzulo, J. J. (1988) 'Interaction of benzimidazole reactive metabolites with nuclear and kinetoplastic DNA, proteins and lipids from Trypanosoma cruzi.', *Experientia*, 44(10), pp. 880–1. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3053234> (Accessed: 18 May 2017).
- Djikeng, A., Ferreira, L., D'Angelo, M., Dolezal, P., Lamb, T., Murta, S., Triggs, V., Ulbert, S., Villarino, A., Renzi, S., Ullu, E. and Tschudi, C. (2001) 'Characterization of a candidate Trypanosoma brucei U1 small nuclear RNA gene.', *Molecular and biochemical parasitology*, 113(1), pp. 109–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11254959> (Accessed: 2 April 2017).
- Docampo, R. and Moreno, S. N. (2011) 'Acidocalcisomes', *Cell Calcium*, 50(2), pp. 113–119. doi: 10.1016/j.ceca.2011.05.012.
- Docampo, R., de Souza, W., Miranda, K., Rohloff, P. and Moreno, S. N. (2005) 'Acidocalcisomes - conserved from bacteria to man', *Nat Rev Microbiol*, 3(3), pp. 251–261. doi: 10.1038/nrmicro1097.
- Drechsler, H. and McAINSH, A. D. (2012) 'Exotic mitotic mechanisms', *Open Biol*, 2(12), p. 120140. doi: 10.1098/rsob.120140.
- Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R. and Clayton, C. (2013) 'Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein', *PLoS Pathog*, 9(4), p. e1003286. doi: 10.1371/journal.ppat.1003286.
- Duhagon, M. A., Dallagiovanna, B., Ciganda, M., Ruyechan, W., Williams, N. and Garat, B. (2003) 'A novel type of single-stranded nucleic acid binding protein recognizing a highly frequent motif in the intergenic regions of Trypanosoma cruzi', *Biochem Biophys Res Commun*, 309(1), pp. 183–188. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12943680.
- Duhagon, M. A., Dallagiovanna, B. and Garat, B. (2001) 'Unusual features of poly[dT-dG].[dC-dA] stretches in CDS-flanking regions of Trypanosoma cruzi genome', *Biochem Biophys Res Commun*, 287(1), pp. 98–103.
- Duhagon, M. A., Pastro, L., Sotelo-Silveira, J. R., Perez-Diaz, L., Maugeri, D., Nardelli, S. C., Schenkman, S., Williams, N., Dallagiovanna, B. and Garat, B. (2009) 'The Trypanosoma cruzi nucleic acid binding protein Tc38 presents changes in the intramitochondrial distribution during the cell cycle', *BMC Microbiol*. 2009/02/13, 9, p. 34. doi: 1471-2180-9-34 [pii]10.1186/1471-2180-9-34.
- Duhagon, M. A., Smircich, P., Forteza, D., Naya, H., Williams, N. and Garat, B. (2011) 'Comparative genomic analysis of dinucleotide repeats in Tritryps', *Gene*. 2011/08/10, 487(1), pp. 29–37. doi: S0378-1119(11)00358-1 [pii]10.1016/j.gene.2011.07.022.
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E.,

- Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. and Andersson, B. (2005) 'The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease', *Science*. 2005/07/16, 309(5733), pp. 409–415. doi: 309/5733/409 [pii]10.1126/science.1112631.
- El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renaud, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D. and Hall, N. (2005) 'Comparative genomics of trypanosomatid parasitic protozoa', *Science*. 2005/07/16, 309(5733), pp. 404–409. doi: 309/5733/404 [pii]10.1126/science.1112181.
- Erben, E., Chakraborty, C. and Clayton, C. (2014) 'The CAF1-NOT complex of trypanosomes.', *Frontiers in genetics*, 4, p. 299. doi: 10.3389/fgene.2013.00299.
- Erben, E. D., Fadda, A., Lueong, S., Rg, J., Hoheisel, D. and Clayton, C. (2014) 'A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in *Trypanosoma brucei*', *PLoS Pathog*, 10(6). doi: 10.1371/journal.ppat.1004178.
- Estevez, A. M. (2008) 'The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes', *Nucleic Acids Res*. 2008/07/10, 36(14), pp. 4573–4586. doi: gkn406 [pii]10.1093/nar/gkn406.
- Fadda, A., Ryten, M., Droll, D., Rojas, F., Farber, V., Haanstra, J. R., Merce, C., Bakker, B. M., Matthews, K. and Clayton, C. (2014) 'Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels', *Mol Microbiol*, 94(2), pp. 307–326. doi: 10.1111/mmi.12764.
- Fiebig, M., Gluenz, E., Carrington, M. and Kelly, S. (2014) 'SLaP mapper: a webserver for identifying and quantifying spliced-leader addition and polyadenylation site usage in kinetoplastid genomes', *Mol Biochem Parasitol*, 196(2), pp. 71–74. doi: 10.1016/j.molbiopara.2014.07.012.
- Franzen, O., Arner, E., Ferella, M., Nilsson, D., Respuela, P., Carninci, P., Hayashizaki, Y., Aslund, L., Andersson, B. and Daub, C. O. (2011) 'The short non-coding transcriptome of the protozoan parasite *Trypanosoma cruzi*', *PLoS Negl Trop Dis*, 5(8), p. e1283. doi: 10.1371/journal.pntd.0001283.
- Franzen, O., Ochaya, S., Sherwood, E., Lewis, M. D., Llewellyn, M. S., Miles, M. A. and Andersson, B. (2011) 'Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener', *PLoS Negl Trop Dis*, 5(3), p. e984. doi: 10.1371/journal.pntd.0000984.
- Franzén, O., Talavera-López, C., Ochaya, S., Butler, C. E., Messenger, L. A., Lewis, M. D., Llewellyn, M. S., Marinkelle, C. J., Tyler, K. M., Miles, M. A. and Andersson, B. (2012) 'Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*', *BMC Genomics*, 13(1), p. 531. doi: 10.1186/1471-2164-13-531.
- Gale, M., Carter, V. and Parsons, M. (1994) 'Cell cycle-specific induction of an 89 kDa serine/threonine protein kinase activity in *Trypanosoma brucei*.' *Journal of cell science*, pp. 1825–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7527048> (Accessed: 9 March 2017).
- Garcia-Silva, M. R., Cura das Neves, R. F., Cabrera-Cabrera, F., Sanguinetti, J., Medeiros, L. C., Robello, C., Naya, H., Fernandez-Calero, T., Souto-Padron, T., de Souza, W. and Cayota, A. (2014) 'Extracellular vesicles shed by *Trypanosoma cruzi* are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells', *Parasitology Research*. Springer Berlin Heidelberg, 113(1), pp. 285–304. doi: 10.1007/s00436-013-3655-1.
- Garcia-Silva, M. R., Frugier, M., Tosar, J. P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S. and Cayota, A. (2010) 'A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules', *Mol Biochem Parasitol*. 2010/02/17, 171(2), pp. 64–73. doi: S0166-6851(10)00035-6 [pii]10.1016/j.molbiopara.2010.02.003.
- De Gaudenzi, J., Frasch, A. C. and Clayton, C. (2005) 'RNA-binding domain proteins in Kinetoplastids: a comparative analysis', *Eukaryot Cell*. 2005/12/13, 4(12), pp. 2106–2114. doi: 4/12/2106 [pii]10.1128/EC.4.12.2106-2114.2005.

- De Gaudenzi, J., Frasch, A. C. and Clayton, C. (2005) 'RNA-Binding Domain Proteins in Kinetoplastids: a Comparative Analysis', *EUKARYOTIC CELL*, 4(12), pp. 2106–2114. doi: 10.1128/EC.4.12.2106–2114.2005.
- De Gaudenzi, J. G., Carmona, S. J., Agüero, F., Frasch, A. C., Agüero, F. and Frasch, A. C. (2013) 'Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes', *PeerJ*. 2013/08/02. PeerJ, Inc, 1, p. e118. doi: 10.7717/peerj.118118 [pii].
- De Gaudenzi, J. G., Jager, A. V., Izcovich, R., Campo, V. A., Jäger, A. V., Izcovich, R., Campo, V. A., Jäger, A. V., Izcovich, R., Campo, V. A., Jäger, A. V., Izcovich, R. and Campo, V. A. (2016) '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*. Blackwell Publishing Inc., 63(4), pp. 440–452. doi: 10.1111/jeu.12288.
- De Gaudenzi, J. G., Noe, G., Campo, V. A., Frasch, A. C. and Cassola, A. (2011) 'Gene expression regulation in trypanosomatids', *Essays Biochem*. 2011/10/26, 51, pp. 31–46. doi: bse0510031 [pii]10.1042/bse0510031.
- Gerber, A. P., Herschlag, D. and Brown, P. O. (2004) 'Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast', *PLoS Biol*. 2004/03/17, 2(3), p. E79. doi: 10.1371/journal.pbio.0020079.
- de Godoy, L. M., Marchini, F. K., Pavoni, D. P., Rampazzo Rde, C., Probst, C. M., Goldenberg, S. and Krieger, M. A. (2012) 'Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis', *Proteomics*. 2012/07/05, 12(17), pp. 2694–2703. doi: 10.1002/pmic.201200078.
- Gomez, M. L., Erijman, L., Arauzo, S., Torres, H. N. and Tellez-Inon, M. T. (1989) 'Protein kinase C in *Trypanosoma cruzi* epimastigote forms: partial purification and characterization', *Mol Biochem Parasitol*. 1989/09/01, 36(2), pp. 101–108. doi: 0166-6851(89)90182-5 [pii].
- Gondran, P., Amiot, F., Weil, D. and Dautry, F. (1999) 'Accumulation of mature mRNA in the nuclear fraction of mammalian cells.', *FEBS letters*, 458(3), pp. 324–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10570933> (Accessed: 28 June 2017).
- Greif, G., Ponce de Leon, M., Lamolle, G., Rodriguez, M., Piñeyro, D., Tavares-Marques, L. M., Reyna-Bello, A., Robello, C. and Alvarez-Valin, F. (2013) 'Transcriptome analysis of the bloodstream stage from the parasite *Trypanosoma vivax*', *BMC Genomics*, 14(1), p. 149. doi: 10.1186/1471-2164-14-149.
- Grisard, E. C., Teixeira, S. M. R., de Almeida, L. G. P., Stoco, P. H., Gerber, A. L., Talavera-Lopez, C., Lima, O. C., Andersson, B. and de Vasconcelos, A. T. R. (2014) 'Trypanosoma cruzi Clone Dm28c Draft Genome Sequence', *Genome Announcements*, 2(1), pp. e01114-13-e01114-13. doi: 10.1128/genomeA.01114-13.
- Grynberg, P., Passos-Silva, D. G., Mourao Mde, M., Hirata Jr., R., Macedo, A. M., Machado, C. R., Bartholomeu, D. C. and Franco, G. R. (2012) 'Trypanosoma cruzi gene expression in response to gamma radiation', *PLoS One*. 2012/01/17, 7(1), p. e29596. doi: 10.1371/journal.pone.0029596PONE-D-11-12240 [pii].
- Gualdron-Lopez, M., Brennand, A., Hannaert, V., Quinones, W., Caceres, A. J., Bringaud, F., Concepcion, J. L. and Michels, P. A. (2012) 'When, how and why glycolysis became compartmentalised in the Kinetoplastea. A new look at an ancient organelle', *Int J Parasitol*, 42(1), pp. 1–20. doi: 10.1016/j.ijpara.2011.10.007.
- Guerra-Giraldez, C., Quijada, L. and Clayton, C. E. (2002) 'Compartmentation of enzymes in a microbody, the glycosome, is essential in *Trypanosoma brucei*', *J Cell Sci*, 115(Pt 13), pp. 2651–2658. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12077356>.
- Guerra-Slompo, E. P., Probst, C. M., Pavoni, D. P., Goldenberg, S., Krieger, M. A. and Dallagiovanna, B. (2012) 'Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs', *Biochem Biophys Res Commun*. 2012/03/20, 420(2), pp. 302–307. doi: S0006-291X(12)00428-7 [pii]10.1016/j.bbrc.2012.02.154.
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. T. L., Noble, W. S., Maniatis, T., Goodbourn, S., Fischer, J., Pawson, T., Nash, P., Tompa, M., Li, N., Bailey, T. T. L., Church, G., Moor, B., Eskin, E., Favorov, A., Frith, M., Fu, Y., Kent, W., Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W., Lenhard, B., Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., Schacherer, F., Henikoff, S., Henikoff, J., Pietrokovski, S., Hughes, J., Estep, P., Tavazoie, S., Church, G., Wang, T., Stormo, G., Schones, D., Sumazin, P., Zhang, M., Roepcke, S., Grossmann, S., Rahmann, S., Vingron, M., Thijs, G., Marchal, K., Lescot, M., Rombauts, S., Moor, B., Rouze, P., Moreau, Y., Aerts, S., Loo, P.,

- Thijs, G., Moreau, Y., Moor, B., Choi, I., Kwon, J., Kim, S., Sandelin, A., Wasserman, W., Staden, R., Tatusov, R., Altschul, S., Koonin, E., Bailey, T. T. L., Gribskov, M., Bailey, T. T. L., Elkan, C., Grundy, W., Bailey, T. T. L., Elkan, C., Baker, M., Hanley, J., McNeil, B., Agresti, A., Crooks, G., Hon, G., Chandonia, J. and Brenner, S. (2007) 'Quantifying similarity between motifs.', *Genome biology*. BioMed Central, 8(2), p. R24. doi: 10.1186/gb-2007-8-2-r24.
- Haag, J., O'hUigin, C. and Overath, P. (1998) 'The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria', *Mol Biochem Parasitol*, 91(1), pp. 37–49. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9574924> (Accessed: 13 February 2017).
- Haile, S., Estevez, A. M. and Clayton, C. (2003) 'A role for the exosome in the in vivo degradation of unstable mRNAs', *RNA*. 2003/11/19, 9(12), pp. 1491–1501. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14624005.
- Haile, S. and Papadopoulou, B. (2007) 'Developmental regulation of gene expression in trypanosomatid parasitic protozoa', *Curr Opin Microbiol*. 2008/01/08, 10(6), pp. 569–577. doi: S1369-5274(07)00141-5 [pii]10.1016/j.mib.2007.10.001.
- Halbeisen, R. E., Galgano, A., Scherrer, T. and Gerber, A. P. (2008) 'Post-transcriptional gene regulation: from genome-wide studies to principles', *Cell Mol Life Sci*. 2007/11/29, 65(5), pp. 798–813. doi: 10.1007/s00018-007-7447-6.
- Hall, T. A. (1999) 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.', *Nucleic Acids Symposium Series*, 41, pp. 95–98.
- Hall, B. S., Bot, C. and Wilkinson, S. R. (2011) 'Nifurtimox Activation by Trypanosomal Type I Nitroreductases Generates Cytotoxic Nitrile Metabolites', *Journal of Biological Chemistry*, 286(15), pp. 13088–13095. doi: 10.1074/jbc.M111.230847.
- Hamedi, A., Botelho, L., Britto, C., Fragoso, S. P., Umaki, A. C., Goldenberg, S., Bottu, G. and Salmon, D. (2015) 'In vitro metacyclogenesis of *Trypanosoma cruzi* induced by starvation correlates with a transient adenylyl cyclase stimulation as well as with a constitutive upregulation of adenylyl cyclase expression', *Mol Biochem Parasitol*, 200(1–2), pp. 9–18. doi: 10.1016/j.molbiopara.2015.04.002.
- Hartmann, C., Benz, C., Brems, S., Ellis, L., Luu, V.-D., Stewart, M., D'Orso, I., Busold, C., Fellenberg, K., Frasnich, A. C. C., Carrington, M., Hoheisel, J. and Clayton, C. E. (2007) 'Small Trypanosome RNA-Binding Proteins TbUBP1 and TbUBP2 Influence Expression of F-Box Protein mRNAs in Bloodstream Trypanosomes', *Eukaryotic Cell*, 6(11), pp. 1964–1978. doi: 10.1128/EC.00279-07.
- Hartmann, C. and Clayton, C. (2008) 'Regulation of a transmembrane protein gene family by the small RNA-binding proteins TbUBP1 and TbUBP2', *Mol Biochem Parasitol*. 2007/11/21, 157(1), pp. 112–115. doi: S0166-6851(07)00275-7 [pii]10.1016/j.molbiopara.2007.09.009.
- Hendriks, E. F. and Matthews, K. R. (2005) 'Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein', *Mol Microbiol*, 57(3), pp. 706–716. doi: 10.1111/j.1365-2958.2005.04679.x.
- Hendriks, E. F., Robinson, D. R., Hinkins, M. and Matthews, K. R. (2001) 'A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form', *EMBO J*, 20(23), pp. 6700–6711. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11726506.
- Herold, A., Teixeira, L. and Izaurralde, E. (2003) 'Genome-wide analysis of nuclear mRNA export pathways in *Drosophila*', *EMBO J*, 22(10), pp. 2472–2483. doi: 10.1093/emboj/cdg233.
- Hershberg, R. and Petrov, D. A. (2008) 'Selection on Codon Bias', *Annual Review of Genetics*, 42(1), pp. 287–299. doi: 10.1146/annurev.genet.42.110807.091442.
- Hessle, V., Bjork, P., Sokolowski, M., Gonzalez de Valdivia, E., Silverstein, R., Artemenko, K., Tyagi, A., Maddalo, G., Ilag, L., Helbig, R., Zubarev, R. A. and Visa, N. (2009) 'The exosome associates cotranscriptionally with the nascent pre-mRNP through interactions with heterogeneous nuclear ribonucleoproteins', *Mol Biol Cell*. 2009/06/06, 20(15), pp. 3459–3470. doi: E09-01-0079 [pii]10.1091/mbc.E09-01-0079.
- Holetz, F. B., Correa, A., Avila, A. R., Nakamura, C. V., Krieger, M. A. and Goldenberg, S. (2007) 'Evidence of P-body-like structures in *Trypanosoma cruzi*', *Biochem Biophys Res Commun*. 2007/04/03, 356(4), pp. 1062–1067. doi: S0006-291X(07)00589-X [pii]10.1016/j.bbrc.2007.03.104.
- Horn, D. (2008) 'Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids', *BMC Genomics*, 9(1), p. 2. doi: 10.1186/1471-2164-9-2.
- Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D. and

- Savioli, L. (2007) 'Control of neglected tropical diseases', *N Engl J Med.* 2007/09/07, 357(10), pp. 1018–1027. doi: 357/10/1018 [pii]10.1056/NEJMra064142.
- Hotez, P., Raff, S., Fenwick, A., Richards Jr., F. and Molyneux, D. H. (2007) 'Recent progress in integrated neglected tropical disease control', *Trends Parasitol.* 2007/10/24, 23(11), pp. 511–514. doi: S1471-4922(07)00249-8 [pii]10.1016/j.pt.2007.08.015.
- Hotz, H. R., Hartmann, C., Huober, K., Hug, M. and Clayton, C. (1997) 'Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation', *Nucleic Acids Res.* 25(15), pp. 3017–3026. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9224601.
- Houston-Ludlam, G. A., Belew, A. T. and El-Sayed, N. M. (2016) 'Comparative Transcriptome Profiling of Human Foreskin Fibroblasts Infected with the Sylvio and Y Strains of *Trypanosoma cruzi*', *PLoS One*, 11(8), p. e0159197. doi: 10.1371/journal.pone.0159197.
- Ignatovichina, A. V., Takagi, Y., Liu, Y., Nagata, K. and Ho, C. K. (2015) 'The messenger RNA decapping and recapping pathway in *Trypanosoma*.' *Proceedings of the National Academy of Sciences of the United States of America.* National Academy of Sciences, 112(22), pp. 6967–72. doi: 10.1073/pnas.1424909112.
- Inoue, A. H., Serpeloni, M., Hiraiwa, P. M., Yamada-Ogatta, S. F., Muniz, J. R., Motta, M. C., Vidal, N. M., Goldenberg, S. and Avila, A. R. (2014) 'Identification of a novel nucleocytoplasmic shuttling RNA helicase of trypanosomes', *PLoS One*, 9(10), p. e109521. doi: 10.1371/journal.pone.0109521.
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B. and Myler, P. J. (2005) 'The genome of the kinetoplastid parasite, *Leishmania major*', *Science*, 309(5733), pp. 436–442. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16020728.
- Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J. M., Dacks, J. B., Acosta-Serrano, A., Field, M. C., Ginger, M. L. and Berriman, M. (2016) 'Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism', *Curr Biol*, 26(2), pp. 161–172. doi: 10.1016/j.cub.2015.11.055.
- Jager, A. V., De Gaudenzi, J. G., Cassola, A., D'Orso, I. and Frasch, A. C. (2007) 'mRNA maturation by two-step trans-splicing/polyadenylation processing in trypanosomes', *Proc Natl Acad Sci U S A.* 2007/02/03, 104(7), pp. 2035–2042. doi: 0611125104 [pii]10.1073/pnas.0611125104.
- Jensen, B. C., Ramasamy, G., Vasconcelos, E. J., Ingolia, N. T., Myler, P. J. and Parsons, M. (2014) 'Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*', *BMC Genomics*, 15, p. 911. doi: 10.1186/1471-2164-15-911.
- Jimeno, S. and Aguilera, A. (2010) 'The THO complex as a key mRNP biogenesis factor in development and cell differentiation', *J Biol*, 9(1), p. 6. doi: 10.1186/jbiol217.
- Jimeno, S., Rondón, A. G., Luna, R. and Aguilera, A. (2002) 'The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability.', *The EMBO journal.* European Molecular Biology Organization, 21(13), pp. 3526–35. doi: 10.1093/emboj/cdf335.
- Jonas, S. and Izaurralde, E. (2015) 'Towards a molecular understanding of microRNA-mediated gene silencing', *Nature Reviews Genetics.* Nature Research, 16(7), pp. 421–433. doi: 10.1038/nrg3965.

- Jung, H., Gkogkas, C. G., Sonenberg, N. and Holt, C. E. (2014) 'Remote control of gene function by local translation', *Cell*, 157(1), pp. 26–40. doi: 10.1016/j.cell.2014.03.005.
- Kawashita, S. Y., da Silva, C. V., Mortara, R. A., Burleigh, B. A. and Briones, M. R. (2009) 'Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins', *Mol Biochem Parasitol.* 2009/04/28, 165(1), pp. 19–31. doi: S0166-6851(09)00003-6 [pii]10.1016/j.molbiopara.2008.12.010.
- Keene, J. D. (2007) 'RNA regulons: coordination of post-transcriptional events', *Nat Rev Genet.* 2007/06/19, 8(7), pp. 533–543. doi: nrg2111 [pii]10.1038/nrg2111.
- Kohler, A. and Hurt, E. (2007) 'Exporting RNA from the nucleus to the cytoplasm', *Nat Rev Mol Cell Biol.* 2007/09/06, 8(10), pp. 761–773. doi: nrm2255 [pii]10.1038/nrm2255.
- Kolev, N. G., Franklin, J. B., Carmi, S., Shi, H., Michaeli, S. and Tschudi, C. (2010) 'The Transcriptome of the Human Pathogen *Trypanosoma brucei* at Single-Nucleotide Resolution', *PLoS Pathogens.* Edited by S. M. Beverley. Public Library of Science, 6(9), p. e1001090. doi: 10.1371/journal.ppat.1001090.
- Konarska, M. M., Grabowski, P. J., Padgett, R. A. and Sharp, P. A. (1985) 'Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors', *Nature*, 313(6003), pp. 552–557. doi: 10.1038/313552a0.
- Kramer, S. (2012) 'Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids', *Mol Biochem Parasitol.* 2011/10/25, 181(2), pp. 61–72. doi: S0166-6851(11)00225-8 [pii]10.1016/j.molbiopara.2011.10.002.
- Kramer, S. (2014) 'RNA in development: how ribonucleoprotein granules regulate the life cycles of pathogenic protozoa', *Wiley Interdiscip Rev RNA*, 5(2), pp. 263–284. doi: 10.1002/wrna.1207.
- Kramer, S. and Carrington, M. (2011) 'Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids', *Trends Parasitol.* 2010/07/09, 27(1), pp. 23–30. doi: S1471-4922(10)00121-2 [pii]10.1016/j.pt.2010.06.011.
- Kramer, S., Piper, S., Estevez, A. and Carrington, M. (2016) 'Polycistronic trypanosome mRNAs are a target for the exosome', *Mol Biochem Parasitol*, 205(1–2), pp. 1–5. doi: 10.1016/j.molbiopara.2016.02.009.
- Kramer, S., Queiroz, R., Ellis, L., Hoheisel, J. D., Clayton, C. and Carrington, M. (2010) 'The RNA helicase DHH1 is central to the correct expression of many developmentally regulated mRNAs in trypanosomes', *J Cell Sci.* 2010/02/04, 123(Pt 5), pp. 699–711. doi: jcs.058511 [pii]10.1242/jcs.058511.
- Kramer, S., Queiroz, R., Ellis, L., Webb, H., Hoheisel, J. D., Clayton, C. and Carrington, M. (2008) 'Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169', *J Cell Sci.* 2008/08/21, 121(Pt 18), pp. 3002–3014. doi: jcs.031823 [pii]10.1242/jcs.031823.
- LeBowitz, J. H., Smith, H. Q., Rusche, L. and Beverley, S. M. (1993) 'Coupling of poly(A) site selection and trans-splicing in *Leishmania*', *Genes Dev.* 1993/06/01, 7(6), pp. 996–1007. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8504937.
- Levy, G. V., Bañuelos, C. P., Níttolo, A. G., Ortiz, G. E., Mendiondo, N., Moretti, G., Tekiel, V. S. and Sánchez, D. O. (2015) 'Depletion of the SR-Related Protein TbRRM1 Leads to Cell Cycle Arrest and Apoptosis-Like Death in *Trypanosoma brucei*.' *PLoS one.* Public Library of Science, 10(8), p. e0136070. doi: 10.1371/journal.pone.0136070.
- Li, Y., Shah-Simpson, S., Okrah, K., Belew, A. T., Choi, J., Caradonna, K. L., Padmanabhan, P., Ndegwa, D. M., Temanni, M. R., Corrada Bravo, H., El-Sayed, N. M. and Burleigh, B. A. (2016) 'Transcriptome Remodeling in *Trypanosoma cruzi* and Human Cells during Intracellular Infection', *PLoS Pathog.* Public Library of Science, 12(4), p. e1005511. doi: 10.1371/journal.ppat.1005511.
- Li, Z. H., De Gaudenzi, J. G., Alvarez, V. E., Mendiondo, N., Wang, H., Kissinger, J. C., Frasch, A. C. and Docampo, R. (2012) 'A 43-nucleotide U-rich element in 3'-untranslated region of large number of *Trypanosoma cruzi* transcripts is important for mRNA abundance in intracellular amastigotes', *J Biol Chem.* 2012/04/14, 287(23), pp. 19058–19069. doi: M111.338699 [pii]10.1074/jbc.M111.338699.
- Liang, X. H., Haritan, A., Uliel, S. and Michaeli, S. (2003) 'trans and cis splicing in trypanosomatids: mechanism, factors, and regulation', *Eukaryot Cell.* 2003/10/14, 2(5), pp. 830–840. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14714922.

ds=14555465.

- Liao, J. Y., Ma, L. M., Guo, Y. H., Zhang, Y. C., Zhou, H., Shao, P., Chen, Y. Q. and Qu, L. H. (2010) 'Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers', *PLoS One*. 2010/05/26, 5(5), p. e10563. doi: 10.1371/journal.pone.0010563.
- Linder, J. C. and Staehelin, L. A. (1977) 'Plasma membrane specializations in a trypanosomatid flagellate', *J Ultrastruct Res*, 60(2), pp. 246–262. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/886643>.
- Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D. and Erben, E. D. (2016) 'Gene expression regulatory networks in *T. brucei*: insights into the role of the mRNA-binding proteome', *Molecular Microbiology*, 100(3), pp. 457–471. doi: 10.1111/mmi.13328.
- Luo, M. J. and Reed, R. (1999) 'Splicing is required for rapid and efficient mRNA export in metazoans.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(26), pp. 14937–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10611316> (Accessed: 12 May 2017).
- Madison-Antenucci, S., Grams, J. and Hajduk, S. L. (2002) 'Editing machines: the complexities of trypanosome RNA editing', *Cell*, 108(4), pp. 435–438. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11909515.
- Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H. O., Bishop, J. R., Falcone, F. H., Gavrilescu, C., Montgomery, J. L., Santori, M. I., Stern, L. S., Wang, Z., Ullu, E. and Tschudi, C. (2000) 'A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA', *RNA*, 6(2), pp. 163–169. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10688355.
- Makin, L. and Gluenz, E. (2015) 'cAMP signalling in trypanosomatids: role in pathogenesis and as a drug target', *Trends Parasitol*, 31(8), pp. 373–379. doi: 10.1016/j.pt.2015.04.014.
- Mallick, B., Ghosh, Z. and Chakrabarti, J. (2008) 'MicroRNA switches in *Trypanosoma brucei*', *Biochemical and Biophysical Research Communications*, 372(3), pp. 459–463. doi: 10.1016/j.bbrc.2008.05.084.
- Manful, T., Cristodero, M. and Clayton, C. (2009) *DRBD1 is the Trypanosoma brucei homologue of the Spliceosome-Associated Protein 49*, *Molecular and Biochemical Parasitology*. doi: 10.1016/j.molbiopara.2009.03.003.
- Manger, I. D. and Boothroyd, J. C. (1998) 'Identification of a nuclear protein in *Trypanosoma brucei* with homology to RNA-binding proteins from cis-splicing systems', *Mol Biochem Parasitol*, 97(1–2), pp. 1–11. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9879882.
- Martinez-Calvillo, S., Vizuet-de-Rueda, J. C., Florencio-Martinez, L. E., Manning-Cela, R. G. and Figueroa-Angulo, E. E. (2010) 'Gene expression in trypanosomatid parasites', *J Biomed Biotechnol*. 2010/02/20, 2010, p. 525241. doi: 10.1155/2010/525241.
- Matthews, K. R. (2005) 'The developmental cell biology of *Trypanosoma brucei*', *J Cell Sci*, 118(Pt 2), pp. 283–290. doi: 10.1242/jcs.01649.
- Matthews, K. R., Tschudi, C. and Ullu, E. (1994) 'A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes', *Genes Dev*. 1994/02/15, 8(4), pp. 491–501. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7907303.
- Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S. and Matthews, K. (2006) 'Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements', *Nucleic Acids Res*, 34(18), pp. 5312–5324. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17012283.
- Minning, T. A., Weatherly, D. B., Atwood 3rd, J., Orlando, R. and Tarleton, R. L. (2009) 'The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*', *BMC Genomics*. 2009/08/12, 10, p. 370. doi: 1471-2164-10-370 [pii]10.1186/1471-2164-10-370.
- Morillo, C. A., Marin-Neto, J. A., Avezum, A., Sosa-Estani, S., Rassi, A., Rosas, F., Villena, E., Quiroz, R., Bonilla, R., Britto, C., Guhl, F., Velazquez, E., Bonilla, L., Meeks, B., Rao-Melacini, P., Pogue, J., Mattos, A., Lazdins, J., Rassi, A., Connolly, S. J. and Yusuf, S. (2015)

- 'Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy', *New England Journal of Medicine*. Massachusetts Medical Society, 373(14), pp. 1295–1306. doi: 10.1056/NEJMoa1507574.
- Morking, P. A., Dallagiovanna, B. M., Foti, L., Garat, B., Picchi, G. F., Umaki, A. C., Probst, C. M., Krieger, M. A., Goldenberg, S. and Fragoso, S. P. (2004) 'TcZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro', *Biochem Biophys Res Commun*, 319(1), pp. 169–177. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15158457.
- Morking, P. A., Rampazzo Rde, C., Walrad, P., Probst, C. M., Soares, M. J., Gradia, D. F., Pavoni, D. P., Krieger, M. A., Matthews, K., Goldenberg, S., Fragoso, S. P. and Dallagiovanna, B. (2012) 'The zinc finger protein TcZFP2 binds target mRNAs enriched during *Trypanosoma cruzi* metacyclogenesis', *Mem Inst Oswaldo Cruz*. 2012/09/20, 107(6), pp. 790–799. doi: S0074-02762012000600014 [pii].
- Naguleswaran, A., Gunasekera, K., Schimanski, B., Heller, M., Hemphill, A., Ochsenreiter, T. and Roditi, I. (2015) 'Trypanosoma brucei RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure.', *mBio*. American Society for Microbiology, 6(2), p. e00114. doi: 10.1128/mBio.00114-15.
- Nardelli, S. C., Avila, A. R., Freund, A., Motta, M. C., Manhães, L., de Jesus, T. C. L., Schenkman, S., Fragoso, S. P., Krieger, M. A., Goldenberg, S. and Dallagiovanna, B. (2007) 'Small-subunit rRNA processome proteins are translationally regulated during differentiation of *Trypanosoma cruzi*.', *Eukaryotic cell*, 6(2), pp. 337–45. doi: 10.1128/EC.00279-06.
- Nardy, A. F., Freire-de-Lima, C. G., Perez, A. R. and Morrot, A. (2016) 'Role of *Trypanosoma cruzi* Trans-sialidase on the Escape from Host Immune Surveillance', *Front Microbiol*, 7, p. 348. doi: 10.3389/fmicb.2016.00348.
- Nazer, E. and Sanchez, D. O. (2011) 'Nucleolar accumulation of RNA binding proteins induced by Actinomycin D is functional in *Trypanosoma cruzi* and *Leishmania mexicana* but not in *T. brucei*', *PLoS One*, 6(8), p. e24184. doi: 10.1371/journal.pone.0024184.
- Nazer, E., Verdun, R. E. and Sanchez, D. O. (2011) 'Nucleolar localization of RNA binding proteins induced by actinomycin D and heat shock in *Trypanosoma cruzi*', *PLoS One*, 6(5), p. e19920. doi: 10.1371/journal.pone.0019920.
- Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I. and Ochsenreiter, T. (2010) 'Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*', *PLoS Pathog*. 2010/08/12, 6(8). doi: 10.1371/journal.ppat.1001037.
- Di Noia, J. M., D'Orso, I., Aslund, L., Sanchez, D. O. and Frasch, A. C. (1998) 'The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions', *J Biol Chem*, 273(18), p. 10843–50. Available at: <http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/referer?http://www.jbc.org/cgi/content/full/273/18/10843>.
- Di Noia, J. M., D'Orso, I., Sanchez, D. O. and Frasch, A. C. (2000) 'AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency', *J Biol Chem*, 275(14), pp. 10218–10227. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10744707.
- O'Connell, D. (2007) 'Neglected Diseases', *Nature*. Nature Publishing Group, 449(7159), pp. 157–157. doi: 10.1038/449157a.
- Obado, S. O., Brillantes, M., Uryu, K., Zhang, W., Ketaren, N. E., Chait, B. T., Field, M. C. and Rout, M. P. (2016) 'Interactome Mapping Reveals the Evolutionary History of the Nuclear Pore Complex', *PLoS Biol*, 14(2). doi: 10.1371/journal.pbio.1002365.
- Obado, S. O., Field, M. C. and Rout, M. P. (2017) 'Comparative interactomics provides evidence for functional specialization of the nuclear pore complex', *Nucleus*, pp. 00–00. doi: 10.1080/19491034.2017.1313936.
- Opperdoes, F. R. and Borst, P. (1977) 'Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome', *FEBS Lett*, 80(2), pp. 360–364. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/142663>.
- Ouellette, M. and Papadopoulou, B. (2009) 'Coordinated gene expression by post-transcriptional regulons in African trypanosomes', *J Biol*, 8(11), p. 100. doi: 10.1186/jbiol203.
- De Pablos, L. M., Diaz Lozano, I. M., Jercic, M. I., Quinzada, M., Gimenez, M. J., Calabuig, E., Espino, A. M., Schijman, A. G., Zulantay, I., Apt, W. and Osuna, A. (2016) 'The C-terminal

- region of *Trypanosoma cruzi* MASPs is antigenic and secreted via exovesicles', *Sci Rep*, 6, p. 27293. doi: 10.1038/srep27293.
- De Pablos, L. M. and Osuna, A. (2012) 'Multigene families in *Trypanosoma cruzi* and their role in infectivity', *Infect Immun*. 2012/03/21. doi: 10.1128/IAI.06225-11 [pii]10.1128/IAI.06225-11.
- Padilla-Mejía, N. E., Florencio-Martínez, L. E., Figueroa-Angulo, E. E., Manning-Cela, R. G., Hernández-Rivas, R., Myler, P. J. and Martínez-Calvillo, S. (2009) 'Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites', *BMC Genomics*, 10(1), p. 232. doi: 10.1186/1471-2164-10-232.
- Palazzo, A. F. and Akef, A. (2012) 'Nuclear export as a key arbiter of "mRNA identity" in eukaryotes', *Biochim Biophys Acta*, 1819(6), pp. 566–577. doi: 10.1016/j.bbagr.2011.12.012.
- Palazzo, A. F. and Akef, A. (2012) 'Nuclear export as a key arbiter of "mRNA identity" in eukaryotes', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(6), pp. 566–577. doi: 10.1016/j.bbagr.2011.12.012.
- Parker, R. and Sheth, U. (2007) 'P bodies and the control of mRNA translation and degradation', *Mol Cell*. 2007/03/14, 25(5), pp. 635–646. doi: S1097-2765(07)00111-6 [pii]10.1016/j.molcel.2007.02.011.
- Parsons, M., Stuart, K. and Smiley, B. L. (1991) 'Trypanosoma brucei: analysis of codon usage and nucleotide composition of nuclear genes.', *Experimental parasitology*, 73(1), pp. 101–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2055297> (Accessed: 9 March 2017).
- Pastro, L., Smircich, P., Di Paolo, A., Becco, L., Duhagon, M. A., Sotelo-Silveira, J. and Garat, B. (2017) 'Nuclear Compartmentalization Contributes to Stage-Specific Gene Expression Control in *Trypanosoma cruzi*', *Frontiers in Cell and Developmental Biology*. Frontiers, 5, p. 8. doi: 10.3389/fcell.2017.00008.
- Paterou, A., Walrad, P., Craddy, P., Fenn, K. and Matthews, K. (2006) 'Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development', *J Biol Chem*. 2006/10/18, 281(51), pp. 39002–39013. doi: M604280200 [pii]10.1074/jbc.M604280200.
- Perez-Díaz, L., Correa, A., Moretao, M. P., Goldenberg, S., Dallagiovanna, B. and Garat, B. (2012) 'The overexpression of the trypanosomatid-exclusive TcRBP19 RNA-binding protein affects cellular infection by *Trypanosoma cruzi*', *Mem Inst Oswaldo Cruz*. 2013/01/09, 107(8), pp. 1076–1079. doi: S0074-02762012000800021 [pii].
- Perez-Díaz, L., Duhagon, M. A., Smircich, P., Sotelo-Silveira, J., Robello, C., Krieger, M. A., Goldenberg, S., Williams, N., Dallagiovanna, B. and Garat, B. (2007) 'Trypanosoma cruzi: molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle', *Exp Parasitol*. 2007/05/04, 117(1), pp. 99–105. doi: S0014-4894(07)00084-7 [pii]10.1016/j.exppara.2007.03.010.
- Perez-Díaz, L., Pastro, L., Smircich, P., Dallagiovanna, B. and Garat, B. (2013) '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', *Biochem Biophys Res Commun*. 2013/06/08, 436(2), pp. 295–299. doi: S0006-291X(13)00903-0 [pii]10.1016/j.bbrc.2013.05.096.
- Pérez-Díaz, L., Silva, T. C. and Teixeira, S. M. R. (2017) 'Involvement of an RNA binding protein containing Alba domain in the stage-specific regulation of beta-amastin expression in *Trypanosoma cruzi*', *Molecular and Biochemical Parasitology*, 211, pp. 1–8. doi: 10.1016/j.molbiopara.2016.12.005.
- Polak, A. and Richle, R. (1978) 'Mode of action of the 2-nitroimidazole derivative benznidazole.', *Annals of tropical medicine and parasitology*, 72(1), pp. 45–54. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/418744> (Accessed: 18 May 2017).
- Portal, D., Espinosa, J. M., Lobo, G. S., Kadener, S., Pereira, C. A., De La Mata, M., Tang, Z., Lin, R. J., Kornblihtt, A. R., Baralle, F. E., Flawia, M. M. and Torres, H. N. (2003) 'An early ancestor in the evolution of splicing: a *Trypanosoma cruzi* serine-arginine-rich protein (TcSR) is functional in cis-splicing', *Mol Biochem Parasitol*, 127(1), pp. 37–46. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12615334.
- Portal, D., Lobo, G. S., Kadener, S., Prasad, J., Espinosa, J. M., Pereira, C. A., Tang, Z., Lin, R. J., Manley, J. L., Kornblihtt, A. R., Flawia, M. M. and Torres, H. N. (2003) 'Trypanosoma cruzi TcSRPK, the first protozoan member of the SRPK family, is biochemically and functionally conserved with metazoan SR protein-specific kinases', *Mol Biochem Parasitol*. 2003/03/05, 127(1), pp. 9–21. doi: S0166685102002992 [pii].
- Porto-Carreiro, I., Attias, M., Miranda, K., De Souza, W. and Cunha-e-Silva, N. (2000)

- 'Trypanosoma cruzi epimastigote endocytic pathway: cargo enters the cytostome and passes through an early endosomal network before storage in reservosomes', *Eur J Cell Biol*, 79(11), pp. 858–869. doi: 10.1078/0171-9335-00112.
- Qiagen (no date) *CLC Genomics Workbench*. Available at: <https://www.qiagenbioinformatics.com/>.
- Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J. D. and Clayton, C. (2009) 'Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons', *BMC Genomics*. 2009/10/28, 10, p. 495. doi: 1471-2164-10-495 [pii]10.1186/1471-2164-10-495.
- Rassi, A., Rassi, A. and Marcondes de Rezende, J. (2012) 'American Trypanosomiasis (Chagas Disease)', *Infectious Disease Clinics of North America*, 26(2), pp. 275–291. doi: 10.1016/j.idc.2012.03.002.
- Rastrojo, A., Carrasco-Ramiro, F., Martín, D., Crespillo, A., Reguera, R. M., Aguado, B. and Requena, J. M. (2013) 'The transcriptome of Leishmania major in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq', *BMC Genomics*, 14(1), p. 223. doi: 10.1186/1471-2164-14-223.
- Ray, D., Weirauch, H. K. K. B. C. M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Rakesh, S., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. O. F., Lei, E. P., Fraser, A. G., Blencowe, B. J., Morris, Q. D. and Hughes, T. R. (2013) 'A compendium of RNA-binding motifs for decoding gene regulation.', *Nature*, 499(7457), pp. 172–7. doi: 10.1038/nature12311.
- Reifur, L., Garcia-Silva, M. R., Poubel, S. B., Alves, L. R., Arauco, P., Buiar, D. K., Goldenberg, S., Cayota, A. and Dallagiovanna, B. (2012) 'Distinct subcellular localization of tRNA-derived fragments in the infective metacyclic forms of Trypanosoma cruzi', *Mem Inst Oswaldo Cruz*. 2012/09/20, 107(6), pp. 816–819. doi: S0074-02762012000600018 [pii].
- Remme, J. H. F., Feenstra, P., Lever, P. R., Medici, A. C., Morel, C. M., Noma, M., Ramaiah, K. D., Richards, F., Seketeli, A., Schmunis, G., van Brakel, W. H. and Vassall, A. (2006) 'Tropical Diseases Targeted for Elimination: Chagas Disease, Lymphatic Filariasis, Onchocerciasis, and Leprosy', in Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A., and Musgrove, P. (eds) *Disease Control Priorities in Developing Countries*. 2nd edn. Washington (DC). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21250324> (Accessed: 13 February 2017).
- Roberts, T. G., Sturm, N. R., Yee, B. K., Yu, M. C., Hartshorne, T., Agabian, N. and Campbell, D. A. (1998) 'Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans', *Mol Cell Biol*. 1998/07/22, 18(8), pp. 4409–4417. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9671450.
- Rout, M. P. and Field, M. C. (2001) 'Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component', *J Biol Chem*. 2001/07/31, 276(41), pp. 38261–38271. doi: 10.1074/jbc.M104024200M104024200 [pii].
- Ruvalcaba-Trejo, L. I. and Sturm, N. R. (2011) 'The Trypanosoma cruzi Sylvio X10 strain maxicircle sequence: the third musketeer', *BMC Genomics*. 2011/01/26, 12, p. 58. doi: 1471-2164-12-58 [pii]10.1186/1471-2164-12-58.
- Saas, J., Ziegelbauer, K., von Haeseler, A., Fast, B. and Boshart, M. (2000) 'A developmentally regulated aconitase related to iron-regulatory protein-1 is localized in the cytoplasm and in the mitochondrion of Trypanosoma brucei', *J Biol Chem*. 2000/01/25, 275(4), pp. 2745–2755. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10644738.
- dos Santos, S. L., Freitas, L. M., Lobo, F. P., Rodrigues-Luiz, G. F., Mendes, T. A., Oliveira, A. C., Andrade, L. O., Chiari, E., Gazzinelli, R. T., Teixeira, S. M., Fujiwara, R. T. and Bartholomeu, D. C. (2012) 'The MASP family of Trypanosoma cruzi: changes in gene expression and antigenic profile during the acute phase of experimental infection', *PLoS Negl Trop Dis*, 6(8), p. e1779. doi: 10.1371/journal.pntd.0001779.
- Schenkman, S., Eichinger, D., Pereira, M. E. and Nussenzweig, V. (1994) 'Structural and functional properties of Trypanosoma trans-sialidase', *Annu Rev Microbiol*. 1994/01/01, 48, pp. 499–523. doi: 10.1146/annurev.mi.48.100194.002435.

- Schenkman, S., Pascoalino Bdos, S. and Nardelli, S. C. (2011) 'Nuclear structure of *Trypanosoma cruzi*', *Adv Parasitol*, 75, pp. 251–283. doi: 10.1016/B978-0-12-385863-4.00012-5.
- Schwanhaussner, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) 'Global quantification of mammalian gene expression control', *Nature*, 473(7347), pp. 337–342. doi: 10.1038/nature10098.
- Serpeloni, M., Vidal, N. M., Goldenberg, S., Avila, A. R. and Hoffmann, F. G. (2011) 'Comparative genomics of proteins involved in RNA nucleocytoplasmic export', *BMC Evol Biol*, 11, p. 7. doi: 10.1186/1471-2148-11-7.
- Shapiro, T. A. and Englund, P. T. (1995) 'The structure and replication of kinetoplast DNA', *Annu Rev Microbiol*, 49(1), pp. 117–143. doi: 10.1146/annurev.mi.49.100195.001001.
- Shaw, J. M., Feagin, J. E., Stuart, K. and Simpson, L. (1988) 'Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons', *Cell*, 53(3), pp. 401–411. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2452696.
- Siegel, T. N., Gunasekera, K., Cross, G. A. and Ochsenreiter, T. (2011) 'Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing', *Trends Parasitol*, 27(10), pp. 434–441. doi: 10.1016/j.pt.2011.05.006.
- Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S. and Cross, G. A. (2010) 'Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites', *Nucleic Acids Res*. 2010/04/14, 38(15), pp. 4946–4957. doi: gkq237 [pii]10.1093/nar/gkq237.
- da Silva Augusto, L., Moretti, N. S., Ramos, T. C., de Jesus, T. C., Zhang, M., Castilho, B. A. and Schenkman, S. (2015) 'A membrane-bound eIF2 alpha kinase located in endosomes is regulated by heme and controls differentiation and ROS levels in *Trypanosoma cruzi*', *PLoS Pathog*, 11(2), p. e1004618. doi: 10.1371/journal.ppat.1004618.
- Simpson, L., Sbicego, S. and Aphasizhev, R. (2003) 'Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business', *RNA*, 9(3), pp. 265–276. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12591999>.
- Simpson, L., Thiemann, O. H., Savill, N. J., Alfonzo, J. D. and Maslov, D. A. (2000) 'Evolution of RNA editing in trypanosome mitochondria', *Proc Natl Acad Sci U S A*. National Academy of Sciences, 97(13), pp. 6986–6993. doi: 10.1073/PNAS.97.13.6986.
- Smircich, P., Eastman, G., Bispo, S., Duhagon, M. A., Guerra-Slompo, E. P., Garat, B., Goldenberg, S., Munroe, D. J., Dallagiovanna, B., Holetz, F. and Sotelo-Silveira, J. R. (2015) 'Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*', *BMC Genomics*, 16, p. 443. doi: 10.1186/s12864-015-1563-8.
- Smith, D. F. and Parsons, M. (1996) *Molecular biology of parasitic protozoa*. 1st edn, *Frontiers in molecular biology*; 13. 1st edn. Oxford; New York: IRL Press at Oxford University Press.
- Solnestam, B. W., Stranneheim, H., Hallman, J., Kaller, M., Lundberg, E., Lundeberg, J. and Akan, P. (2012) 'Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs', *BMC Genomics*. 2012/11/01, 13, p. 574. doi: 1471-2164-13-574 [pii]10.1186/1471-2164-13-574.
- de Souza, W. (2009) 'Structural organization of *Trypanosoma cruzi*', *Mem Inst Oswaldo Cruz*. 2009/09/24, 104 Suppl, pp. 89–100. doi: S0074-02762009000900014 [pii].
- De Souza, W. (2002) 'Basic cell biology of *Trypanosoma cruzi*', *Curr Pharm Des*, 8(4), pp. 269–285. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11860366.
- de Souza, W., Attias, M. and Rodrigues, J. C. F. (2009) 'Particularities of mitochondrial structure in parasitic protists (Apicomplexa and Kinetoplastida)', *The International Journal of Biochemistry & Cell Biology*, 41(10), pp. 2069–2080. doi: 10.1016/j.biocel.2009.04.007.
- De Souza, W. and Meyer, H. (1974) 'On the fine structure of the nucleus in *Trypanosoma cruzi* in tissue culture forms. Spindle fibers in the dividing nucleus', *J Protozool*. 1974/02/01, 21(1), pp. 48–52. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4594241.
- Strambio-De-Castillia, C., Niepel, M. and Rout, M. P. (2010) 'The nuclear pore complex: bridging nuclear transport and gene regulation', *Nat Rev Mol Cell Biol*, 11(7), pp. 490–501. doi:

10.1038/nrm2928.

Strom, A. C. and Weis, K. (2001) 'Importin-beta-like nuclear transport receptors', *Genome Biol*, 2(6), p. REVIEWS3008. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11423015>.

Suganuma, K., Yamasaki, S., Asada, M., Kawazu, S. and Inoue, N. (2012) 'The epimastigote stage-specific gene expression of CESP is tightly regulated by its 3' UTR', *Mol Biochem Parasitol*. 2012/10/13, 186(1), pp. 77–80. doi: S0166-6851(12)00243-5 [pii]10.1016/j.molbiopara.2012.09.010.

Tagoe, D. N., Kalejaiye, T. D. and de Koning, H. P. (2015) 'The ever unfolding story of cAMP signaling in trypanosomatids: vive la difference!', *Front Pharmacol*, 6, p. 185. doi: 10.3389/fphar.2015.00185.

Talavera-López, C. N. (2016) 'TRYPANOSOMA CRUZI GENOME PLASTICITY AND EVOLUTION'.

Tan, T. H. P., Pach, R., Crausaz, A., Ivens, A. and Schneider, A. (2002) 'tRNAs in Trypanosoma brucei: genomic organization, expression, and mitochondrial import.', *Molecular and cellular biology*, 22(11), pp. 3707–17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11997507> (Accessed: 23 February 2017).

Trask, H. W., Cowper-Sal-lari, R., Sartor, M. A., Gui, J., Heath, C. V., Renuka, J., Higgins, A. J., Andrews, P., Korc, M., Moore, J. H. and Tomlinson, C. R. (2009) 'Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs', *RNA*, 15(10), pp. 1917–1928. doi: 10.1261/rna.1677409.

Ullu, E., Matthews, K. R. and Tschudi, C. (1993) 'Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts', *Mol Cell Biol*. 1993/01/01, 13(1), pp. 720–725. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8417363.

Ulrich, P. N., Jimenez, V., Park, M., Martins, V. P., Atwood 3rd, J., Moles, K., Collins, D., Rohloff, P., Tarleton, R., Moreno, S. N., Orlando, R. and Docampo, R. (2011) 'Identification of contractile vacuole proteins in Trypanosoma cruzi', *PLoS One*, 6(3), p. e18013. doi: 10.1371/journal.pone.0018013.

Valencia, P., Dias, A. P. and Reed, R. (2008) 'Splicing promotes rapid and efficient mRNA export in mammalian cells.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(9), pp. 3386–91. doi: 10.1073/pnas.0800250105.

Vasquez, J. J., Hon, C. C., Vanselow, J. T., Schlosser, A. and Siegel, T. N. (2014) 'Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages', *Nucleic Acids Res*, 42(6), pp. 3623–3637. doi: 10.1093/nar/gkt1386.

Vazquez, M., Atorrasagasti, C., Bercovich, N., Volcovich, R. and Levin, M. J. (2003) 'Unique features of the Trypanosoma cruzi U2AF35 splicing factor', *Mol Biochem Parasitol*, 128(1), pp. 77–81. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12706799.

Vickerman, K. (1970) 'Functional aspects of the cytology of trypanosomes', *Trans R Soc Trop Med Hyg*, 64(1), pp. 180–181. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/5442056>.

Vlasova-St Louis, I. and Bohjanen, P. R. (2011) 'Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1.', *Current opinion in genetics & development*, 21(4), pp. 444–51. doi: 10.1016/j.gde.2011.03.002.

Walrad, P. B., Capewell, P., Fenn, K. and Matthews, K. R. (2012) 'The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in Trypanosoma brucei.', *Nucleic acids research*. Oxford University Press, 40(7), pp. 2869–83. doi: 10.1093/nar/gkr1106.

Wang, Y., Opperman, L., Wickens, M. and Hall, T. M. (2009) 'Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein', *Proc Natl Acad Sci U S A*. 2009/11/11, 106(48), pp. 20186–20191. doi: 0812076106 [pii]10.1073/pnas.0812076106.

Weil, D., Boutain, S., Audibert, A. and Dautry, F. (2000) 'Mature mRNAs accumulated in the nucleus are neither the molecules in transit to the cytoplasm nor constitute a stockpile for gene expression.', *RNA (New York, N.Y.)*, 6(7), pp. 962–75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10917593> (Accessed: 28 June 2017).

Wen, Y. Z., Zheng, L. L., Qu, L. H., Ayala, F. J. and Lun, Z. R. (2012) 'Pseudogenes are not pseudo any more', *RNA Biol*, 9(1), pp. 27–32. doi: 10.4161/rna.9.1.18277.

Wente, S. R. and Rout, M. P. (2010) 'The Nuclear Pore Complex and Nuclear Transport', *Cold*

- Spring Harbor Perspectives in Biology*, 2(10), pp. a000562–a000562. doi: 10.1101/cshperspect.a000562.
- Westenberger, S. J., Cerqueira, G. C., El-Sayed, N. M., Zingales, B., Campbell, D. A. and Sturm, N. R. (2006) 'Trypanosoma cruzi mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region', *BMC Genomics*. 2006/03/24, 7, p. 60. doi: 1471-2164-7-60 [pii]10.1186/1471-2164-7-60.
- Weston, D., La Flamme, A. C. and Van Voorhis, W. C. (1999) 'Expression of Trypanosoma cruzi surface antigen FL-160 is controlled by elements in the 3' untranslated, the 3' intergenic, and the coding regions', *Mol Biochem Parasitol*, 102(1), pp. 53–66. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10477176.
- Wharton, R. P. and Aggarwal, A. K. (2006) 'mRNA regulation by Puf domain proteins', *Sci STKE*. 2006/09/28, 2006(354), p. pe37. doi: stke.3542006pe37 [pii]10.1126/stke.3542006pe37.
- WHO (2009) *Global_chagas_2009*. Available at: http://gamapserver.who.int/mapLibrary/Files/Maps/Global_chagas_2009.png (Accessed: 21 February 2017).
- WHO (2016) 'Chagas disease (American trypanosomiasis)', *World Health Organ Tech Rep*. World Health Organization, (340). Available at: <http://www.who.int/mediacentre/factsheets/fs340/en/> (Accessed: 21 February 2017).
- Wickens, M., Bernstein, D. S., Kimble, J. and Parker, R. (2002) 'A PUF family portrait: 3'UTR regulation as a way of life', *Trends Genet*. 2002/02/23, 18(3), pp. 150–157. doi: S0168952501026166 [pii].
- Wittkopp, P. J. and Kalay, G. (2011) 'Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence', *Nat Rev Genet*, 13(1), pp. 59–69. doi: 10.1038/nrg3095.
- Xu, P., Wen, L., Benegal, G., Wang, X. and Buck, G. A. (2001) 'Identification of a spliced leader RNA binding protein from Trypanosoma cruzi', *Mol Biochem Parasitol*, 112(1), pp. 39–49. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11166385.
- Yao, C. (2010) 'Major surface protease of trypanosomatids: one size fits all?', *Infect Immun*, 78(1), pp. 22–31. doi: 10.1128/IAI.00776-09.
- Yoshida, N. (2008) 'Trypanosoma cruzi infection by oral route', *Parasitology International*, 57(2), pp. 105–109. doi: 10.1016/j.parint.2007.12.008.
- Zeiner, G. M., Sturm, N. R. and Campbell, D. A. (2003) 'Exportin 1 mediates nuclear export of the kinetoplastid spliced leader RNA', *Eukaryot Cell*, 2(2), pp. 222–230. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12684371>.
- Zhang, K., Shi, Z.-M., Chang, Y.-N., Hu, Z.-M., Qi, H.-X. and Hong, W. (2014) 'The ways of action of long non-coding RNAs in cytoplasm and nucleus', *Gene*, 547(1), pp. 1–9. doi: 10.1016/j.gene.2014.06.043.
- Zingales, B., Andrade, S. G., Briones, M. R., Campbell, D. A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A. M., Machado, C. R., Miles, M. A., Romanha, A. J., Sturm, N. R., Tibayrenc, M. and Schijman, A. G. (2009) 'A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI', *Mem Inst Oswaldo Cruz*. 2009/12/23, 104(7), pp. 1051–1054. doi: S0074-02762009000700021 [pii].

7. Comprendiendo el *cis-splicing* en *T. cruzi*: análisis del gen codificante de la poliA polimerasa

Las proteínas de los kinetoplastos, son codificados generalmente por una región génica sin interrupciones (un único exón), aunque hay algunas excepciones descritas en las que la región codificante está formada por más de un exón, entre ellas se destacan los genes de la poliA polimerasa (PAP) de *T. cruzi* y *T. brucei* que están formados por dos exones y para los cuales se demostró que el *cis-splicing* sucede (Mair *et al.*, 2000). Considerando que los datos de los transcriptomas nuclear y citoplasmático pueden brindar información para esclarecer el proceso de *cis-splicing* en *T. cruzi* decidimos profundizar en el estudio del gen que codifica para la proteína PAP.

7.1. Análisis comparativo *in silico* del gen que codifica para la proteína PAP en cepas de *T. cruzi*.

En la cepa Dm28c, utilizada para realizar los transcriptomas, el gen que codifica para la proteína PAP está anotado como una región codificante ininterrumpida, por lo tanto, realizamos un análisis comparativo con las secuencias reportadas para los haplotipos Esmeraldo y non-Esmeraldo *like* de CL Brener para determinar si existe conservación de *cis-splicing*.

En primer lugar, notamos que los límites anotados del intrón son diferentes entre los haplotipos. De hecho, el haplotipo non-Esmeraldo *like* de CL Brener (TcCLB.508153.440), no presenta las señales conservadas requeridas para el *splicing* (figura 9); adicionalmente, la secuencia intrónica reportada interrumpiría un codón, que se regeneraría por *splicing*, no se modifica el marco abierto de lectura, pero cambiaría el aminoácido codificado. No obstante, se observan señales canónicas de *splicing* vecinas al sitio anotado.

Por otro lado, aunque el gen reportado para la cepa Dm28c no posee un intrón anotado, sí presenta las secuencias canónicas necesarias para el *splicing*, los dinucleótidos GT y AG en el 5' y 3' del intrón respectivamente y la secuencia complementaria a U1 (GTATGGTGAGT) (Djikeng *et al.*, 2001).

Del análisis de la secuencia, se deduce, que la anotación automática del genoma identifica un único marco abierto de lectura porque la secuencia que correspondería al intrón está en fase con las secuencias exónicas. Además, en la cepa Dm28c, la inserción de una citosina en la posición 1612 del gen conduce a un cambio de fase y a la aparición de un codón de *stop*, que en conjunto con el mantenimiento de fase en el intrón llevan a predecir una proteína de tamaño acorde al esperado (figura 9).

	210	220	230	240	250
TCDM_00289	GCGAGCCTAC	CCGTTTCGGTT	CTTGC GGTTT	GGGTGCCTCC	GTTGCAGATA
TcCLB.5103	GCGAGCCTAT	CCGTTTCGGTT	CCTGC GGTTT	GGGTGCCTCC	GTTGCAGATA
TcCLB.5081	GCGAGCCTAC	CCGTTTCGGTT	CTTGC GGTTT	GGGTGCCTCC	GTTGCAGATA
	260	270	280	290	300
TCDM_00289	GCGATCTCGA	TATGTATGGT	GAGTCATTTT	CCAACCCCTC	TCTAATCCCC
TcCLB.5103	GCGATCTCGA	TATgtatggt	gagt catttt	ccaaccctc	tctaatacccc
TcCLB.5081	GCGATCTCGA	Tatgtatggt	gagt catttt	ccaaccctc	tctaatacccc
	310	320	330	340	350
TCDM_00289	<i>TCACTTATTT</i>	<i>TAACGCGTTC</i>	<i>CACACCACAC</i>	<i>TTCCCTTCTC</i>	<i>TTGTAAAGCT</i>
TcCLB.5103	tcacttat	taacgcgttc	cacaccacac	ttcccttctc	ttgtaaagct
TcCLB.5081	tcacttgttt	taacgcgttc	cacatcacac	ttcccttctc	ttgtaaagct
	360	370	380	390	400
TCDM_00289	<i>ACCGCCTCTC</i>	<i>TCGCCCTCTC</i>	<i>TCTTTTAT</i>	<i>TATTTTCTAT</i>	<i>TTATCAGTGT</i>
TcCLB.5103	accgcctctc	tcgccctctc	tctttttatt	tattttctat	ttatcagtgt
TcCLB.5081	accgcctctc	tcgccctctc	tctttttatt	tattttctat	ttatcagtgt
	410	420	430	440	450
TCDM_00289	<i>TTGAACAAC</i>	<i>CGAAAAAAA</i>	<i>AAAA-TCTGC</i>	<i>TTTCTTTCC</i>	<i>TTTTTTTTT</i>
TcCLB.5103	ttgaacaact	taaaaaaaaa	aaaa-tctgc	ttttctttcc	ttttttttt
TcCLB.5081	ttgaacaact	caaaaaaaaa	aaaaatctgc	ttttctttcc	ttttttcttt
	460	470	480	490	500
TCDM_00289	<i>TT-ATATTCC</i>	<i>GTGCCGTCG</i>	<i>TTTTCAATTG</i>	<i>AAAACGCTCG</i>	<i>CACTGTCTTG</i>
TcCLB.5103	tttatattcc	gtgccggtcg	ttttcaatta	aaaacgctcg	cactttcttg
TcCLB.5081	tt-atattcc	gtgccggtcg	ttttcaattg	aaaacgctcg	cactgtcttg
	510	520	530	540	550
TCDM_00289	<i>AATT---CT</i>	<i>TTTTTTTAC</i>	<i>ACACTGTCCA</i>	<i>CTCCGTGTGC</i>	<i>ATCTCCAAAT</i>
TcCLB.5103	aattttttatt	ttttttttac	acaccgtcca	ctccgtgtgc	atctccaaat
TcCLB.5081	aattat--tt	ttttttttac	acactgtcca	ctccgtgtgc	atctccaaat
	560	570	580	590	600
TCDM_00289	<i>GCCGTGTTTT</i>	<i>TTCTGGTAGT</i>	<u>GT</u> GCTGTTTT	GTTCTTTTAA	GGTAACGGCG
TcCLB.5103	gccgtgtttt	ttccggtagt	<u>GT</u> CTGATTT	GTCTTTTAA	GGTAACGGCG
TcCLB.5081	gccgtgtttt	ttccggtagt	<u>GT</u> GCTGTTTT	GTTCTTTTAA	GGTAACGGCG
	610	620	630	640	650
TCDM_00289	GAGGTGTTTT	TTCGCGAGTT	TCCTCCGCTG	CTGCAGTCTG	TCGTGGGGAA
TcCLB.5103	GAGGTGTTTT	TTCGCGAGTT	TCCTCCACTG	CTGCAGCCTG	TCGTGGGGAA
TcCLB.5081	GAGGTGTTTT	TTCGCGAGTT	CCCTCCGCTG	CTGCAGCCTG	TCGTGGGGAA
	1510	1520	1530	1540	1550
TCDM_00289	AATGATGGTC	CACCAAGGTC	TGTACCAAAA	TCCGGCCGTT	TTCTTAAAGG
TcCLB.5103	AATGATGTTC	CACCAAGGTC	TGTACCAAAA	CCCGGCCGTT	TTCTTAAAGG
TcCLB.5081	AATGATGGTC	CACCAAGGTC	TGCACCAAAA	TCCGGCCGTT	TTCTTAAAGG
	1560	1570	1580	1590	1600
TCDM_00289	TAGTTGTTTT	ATCGTGGGGA	TAAAAGAGAA	AGTTGGTAGG	CGCTTTCCAC
TcCLB.5103	TAGTTGTTTT	ATCGTAGGGA	TAAAAGAGAA	AGTTGGTAGG	CGCTTTCCAC
TcCLB.5081	TAGTTGTTTT	ATCGTGGGGA	TAAAAGAGAA	AGTTGGTAGG	CGTTTTCCAC
	1610	1620	1630	1640	1650
TCDM_00289	AGAAGAATCT	GTTTTT CGA	GGCGTT TGA -	-----	-----
TcCLB.5103	AGAAGAATAT	GTTTTTT- GA	GGCGTTTGAT	GAGCTAAGGT	ATACTGTGTT
TcCLB.5081	AGAAGAATAT	GTTTTTT- GA	GGCGTTTGAT	GAGCTAAGGT	ATGCTGTGTT
	1660	1670	1680	1690	1700
TCDM_00289	-----	-----	-----	-----	-----
TcCLB.5103	GGAGGAGTGC	AATGCGACGA	AGAGTGTGCG	GGGATTTGAG	CGTGACGAGC
TcCLB.5081	GGAGGAGTGC	AATGCGACGA	AGAGTGTGCG	GGGATTTGAG	CGTGACGAGC
	1710	1720	1730	1740	1750
TCDM_00289	-----	-----	-----	-----	-----
TcCLB.5103	GCACCATGCA	CGAGCCATGG	TTTGCGCTAG	TGTCCGCTGC	AGATCTTCTT
TcCLB.5081	GCACCATGCA	CGAGCCATGG	TTTGCGCTAG	TGTCCGCTGC	AGATCTTCTT
	1760	-----	-----	-----	-----
TCDM_00289	-----	-----	-----	-----	-----
TcCLB.5103	CCTATTCTAA	AGGCGTAA	-----	-----	-----
TcCLB.5081	CCTATTCTAA	AGGCGTAA	-----	-----	-----

Figura 9. Alineamiento de las secuencias génicas que codifican para las proteínas PAPs de *T. cruzi*. Se muestran las regiones 5' (incluyendo la región del intrón) y 3' de los genes TCDM_00289 (cepa Dm28c); TcCLB.5103 (TcCLB.510317.30, cepa Esmeraldo *like*); TcCLB.5081 (TcCLB.508153.440, cepa non Esmeraldo *like*). En minúscula se muestra la secuencia del intrón anotado en la base de datos del TriTrypDB. En cursiva se indica la secuencia del posible intrón del gen de la cepa Dm28c. En amarillo se destacan las señales de *splicing*. Con letras en azul se muestra la secuencia complementaria al ARN U1 de *T. cruzi* (Djikeng *et al.*, 2001). Se subrayan los codones correspondientes a los aminoácidos de la unión exón-exón. En celeste se indica la citosina insertada en la posición 1612 de gen anotado en la cepa Dm28c. En rojo se indica el codón de *stop* prematuro generado por la inserción de la citosina.

A su vez, el alineamiento de las secuencias del gen que codifica para la proteína PAP de los kinetopóstidos reportados en el TriTrypDB muestra que la posición del intrón y las secuencias consenso necesarias para el *cis-splicing* están conservadas en estos organismos, mientras que los intrones poseen secuencias muy divergentes (ver anexo).

Considerando las diferencias encontradas en el gen que codifica para la proteína PAP de la cepa Dm28c respecto a otros organismos, analizamos las consecuencias a nivel proteico. En la figura 10 se muestra el alineamiento de las proteínas reportadas para los haplotipos Esmeraldo y non-Esmeraldo de CL Brener, las cepas Dm28c y Sylvio de *T. cruzi*, *T. brucei*, *T. Evans*, *L. major*, *Plasmodium falciparum*, *Homo sapiens* y *Drosophila melanogaster*. Las proteínas tienen una alta identidad de secuencia mientras que la traducción del gen reportado para Dm28c, produce una proteína con una inserción de 101 aminoácidos respecto a la secuencia reportada para la cepa CL Brener.

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          160      170      180      190      200
TCDM_00289  ----- AYPFGSCGLG ASVADSDLDL YGESFSNPSTL IPSLILTRST
TcCLB_5081  ----- AYPFGSCGLG ASVADSDLDL
TcCLB_5103  ----- AYPFGSCGLG ASVADSDLDL
Tb927_3.31  ----- AYPFGSCGLS ASVAESDLDV
LmjF_29.26  ----- VYVFGSSGLD AAITGSDIDL
TcIL3000_3  ----- AYPFGSCGLS ASVADSDLDL
TevSTIB805  ----- AYPFGSCGLS AAVAESDLDV
TvY486_030  ----- AYPFGSCGLA ASVADSDLDI
>AAK31791.  PSVVATVGGK IFTFGSYRLG VHTKGAIDDA
>XP_966239  EEEANNISGN LYTFGSYRLG VITPNSDIDC
>AAM29439.  ESAAEKLGGK IYTFGSYRLG VHHKGAIDDA
    
```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          210      220      230      240      250
TCDM_00289  PHFPSLVKLP PLSPSLFLFI FYLSVFEQLE KKKICFSFLF FFIYPCRSFS
TcCLB_5081  -----
TcCLB_5103  -----
Tb927_3.31  -----
LmjF_29.26  -----
TcIL3000_3  -----
TevSTIB805  -----
TvY486_030  -----
>AAK31791.  -----
>XP_966239  -----
>AAM29439.  -----
    
```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          260      270      280      290      300
TCDM_00289  IENARTVLNS FFLHTVHVSVC ISKCRVFSGS VLFCSFKVTA EVFFREFPPL
TcCLB_5081  ----- VLFCSFKVTA EVFFREFPPL
TcCLB_5103  ----- VLICPFKVTA EVFFREFPPL
Tb927_3.31  ----- ALICPSVLTT DIFFDEFPRL
LmjF_29.26  ----- IALCSASVNT SLFFQTFPRS
TcIL3000_3  ----- AVFCPFVTR FIFPHEFPAL
TevSTIB805  ----- ALICPSVLTT DIFFDEFPRL
TvY486_030  ----- VLFCPFVTA AVFFHEFPPL
>AAK31791.  ----- LCVAPRHVER SDFQSFEEK
>XP_966239  ----- IFLAPQLTR EIFFNEFYK
>AAM29439.  ----- LCVAPRNIER TDYFQSFFEY
    
```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          310      320      330      340      350
TCDM_00289  LQS--VVGNV VLVP-ARVPV VKFSYSGTSV DVVFVSAALS APPTTEQLLD
TcCLB_5081  LQP--VVGNV VLVP-ARVPV VKFSYSGTSV DVVFVSAALA APPTTEQLLD
TcCLB_5103  LQP--VVGNV VLVP-ARVPV VKFSYSGTSV DVVFVSAALA APPTTEQLLD
Tb927_3.31  LYN--SVGPV SVVA-ARVPV VKFEYRGTA VVVFVSVGLP QPPTTEQMLD
LmjF_29.26  AEG--RLQNV VVVT SARVPL IKFSYNGVHV DLLFASVDMR TAPDTNELLR
TcIL3000_3  LET--IVGHV IVVA-ARVPV VRFVYQGM SV DVVFI SVALS QPPTSDQLLD
TevSTIB805  LYN--SVGPV SVVA-ARVPV VKFEYRGTA VVVFVSVGLP QPPTTEQMLD
TvY486_030  LQT--VVGHV IVVA-ARVPV VKFAHEGTLV DLVVFVVALA APPTTEQLLD
>AAK31791.  LKHQDGIRNL RAVEDAFVPV IKFEFDGIEI DLV FARLAIQ TISDNLDLRD
>XP_966239  LQQDRNVKKL QALPEYTPPI IQFMYDDVDI D LL LATLPYK TLKDCYYSLD
>AAM29439.  LKKQPEVTEC RSVEEAFVPV IKMNF DGIEI D LL FARLSLK EIPDDFDLRD
    
```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          660      670      680      690      700
TCDM_00289  ----- --FLKGSFCF IVGIKEKVGR RFPQKNLFFR GV-----
TcCLB_5081  ----- --FLKGSFCF IVGIKEKVGR RFPQKNMFFE AFDELRYAVL
TcCLB_5103  ----- --FLKGSFCF IVGIKEKVGR RFPQKNMFFE AFDELRYTVL
Tb927_3.31  ----- --CFLSVNF  LIGVKAKEGG PTPQPSLFTE ACGEFHHAVK
LmjF_29.26  ----- --TVHCSR  AFFFGVRNGN NDVAHSDVEA AFREFEFAVT
TcIL3000_3  ----- --CCSSIYF  LIGVRSKNSG SSLRRCMFDE ATDEFIYALQ
TevSTIB805  ----- --CFLSVNF  LIGVKAKEGG PTPQPSLFTE ACGEFHHAVK
TvY486_030  ----- --IRRCVVF  LIGVREKDR  QKLHQSMFSS ALDEFRYAVQ
>AAK31791.  ----- --YVSMWFL  GIIFRRVENA ESNIDLTYD IQSFTDTVYR
>XP_966239  ----- --FIALIFF  YKNPYENTFN LSYAIRDFID IALNWPQVKV
>AAM29439.  QSQGNQSAVT SAFPCCSMWFI GLEFERSE-- -NLNVDLTES IQNFTEHVMM
    
```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          710      720      730      740      750
TCDM_00289  -----
TcCLB_5081  EECNATKSVR GFERDERTMH EPWFALVSAA DLLPILKA--
TcCLB_5103  EECNATKSVR GFERDERTMH EPWFALVSAA DLLPILKA--
Tb927_3.31  EGCNND SIPW SFQRNDRAMH WPRVKFLEVH QVLHLLQDLK
LmjF_29.26  EGTGKN--- PFEWNRSVML GPRLSFFSIY DPLTADSPCQ ALYSACAEMM
TcIL3000_3  DGCSTENQSR SFHRDERKMH WPQVKLIGTS EISSFVDGL-
TevSTIB805  EGCNND SIPW SFQRNDRAMH WPRVKFLEVH QVLHLLQDLK
TvY486_030  DGCSSTKGSR VFQRDERSMH GPRLTLIDTL EALTFLNK--
>AAK31791.  QANNINMLKE GMKIEATHVK KKQLHHYLP A EILQKKKQS LSDVNRSSGG
>XP_966239  YPNAFKINIM YQKRSQVLEF LDIVTKEENQ KNKQVMENND KDETNNHQNG
>AAM29439.  HGVNIKMLKE GMTIDARHVK RKQLSLYLD S DFLKRERK-S MESHNNFNNT
    
```

Figura 10. Alineamiento de las secuencias de las proteínas PAP. TCDM_00289 (cepa Dm28c); TcCLB.5103, TcCLB.510317.30 (cepa Esmeraldo *like*); TcCLB.5081, TcCLB.508153.440 (cepa non-Esmeraldo *like*); Tb927.3.31, Tb927.3.3160 (*T. brucei*); LmjF.29.26, LmjF.29.2600 (*L. major*); TcIL3000_3, TcIL3000_3_2070.1 (*Trypanosoma congolense*); TevSTIB805, TevSTIB805.3.3230-t26_1 (*Trypanosoma evansi*); TcY486_030, TvY486_0302520 (*Trypanosoma vivax*); >AAK31791., AAK31791.1 (*Homo sapiens*); >XP_966239, XP_966239.1 (*Plasmodium falciparum*); >AAM29439. AAM29439.1 (*Drosophila melanogaster*). En gris se muestra parte del exón 1 y en rosado del exón 2 de la cepa Esmeraldo *like*. En verde, se indica el residuo de isoleucina de la cepa non-Esmeraldo *like*, erróneamente asignado. Con recuadros rojos se muestran los residuos de ácido aspártico que forman la tríada de unión a metales, y en celeste se destacan los aminoácidos que cambian en la cepa Dm28c por la inserción de una citosina en la posición 1612 del gen.

La proteína de *T. cruzi* posee de un dominio nucleotidiltransferasa presente en las proteínas poli(A) polimerasas y uridil terminal transferasas que abarca desde el residuo 6 al 414 (en el haplotipo Esmeraldo *like*). A su vez, los residuos de ácido aspártico de la tríada de unión a metales, necesarios para la función de la proteína, están conservados. En *T. cruzi*, dos de ellos están codificados en el primer exón y el tercero, en el segundo por lo tanto, la inserción de 101 aminoácidos en la proteína de la cepa Dm28c produciría un alejamiento de los mismos que probablemente afectaría su función.

En conclusión, nuestro análisis sugiere que el gen que codifica para la proteína PAP de la cepa Dm28c estaría formado, al igual que el de otras cepas, por dos exones y por lo tanto el proceso de *cis-splicing* también sucedería.

7.2. Análisis de la expresión del gen que codifica para la proteína PAP en los transcriptomas citoplásmico, nuclear y total de epimastigotas de *T. cruzi*.

Los datos obtenidos a partir de los transcriptomas total, nuclear y citoplásmicos de *T. cruzi* pueden esclarecer si el proceso de *cis-splicing* sucede efectivamente en epimastigotas de la cepa Dm28.

Para esto, analizamos los mapeos realizados previamente al genoma de la cepa Esmeraldo *like*. En primer lugar, es importante destacar que, aunque actualmente los genes que codifican para la proteína PAP están anotados en la base de datos del TriTrypDB, para la cepa CL Brener, como un único gen interrumpido por un intrón, el archivo del genoma posee la anotación antigua, en la cual la PAP era codificada dos genes consecutivos (TcCLB.510317.30 + TcCLB.510317.34). Esta observación resalta la importancia del estudio manual de los datos en cada caso particular, al estudiar la expresión de los genes a partir de aproximaciones masivas. El resultado del mapeo de nuestros datos de RNA-Seq al genoma del haplotipo Esmeraldo *like* de la cepa CL Brener de *T. cruzi* en la región que codifica para la proteína PAP se muestra en la figura 11. Se observa que la región que abarca los dos genes (exones) se encuentra cubierta por *reads* y no se evidencian diferencias notorias entre los transcriptomas, nuclear, citoplasmático o total.

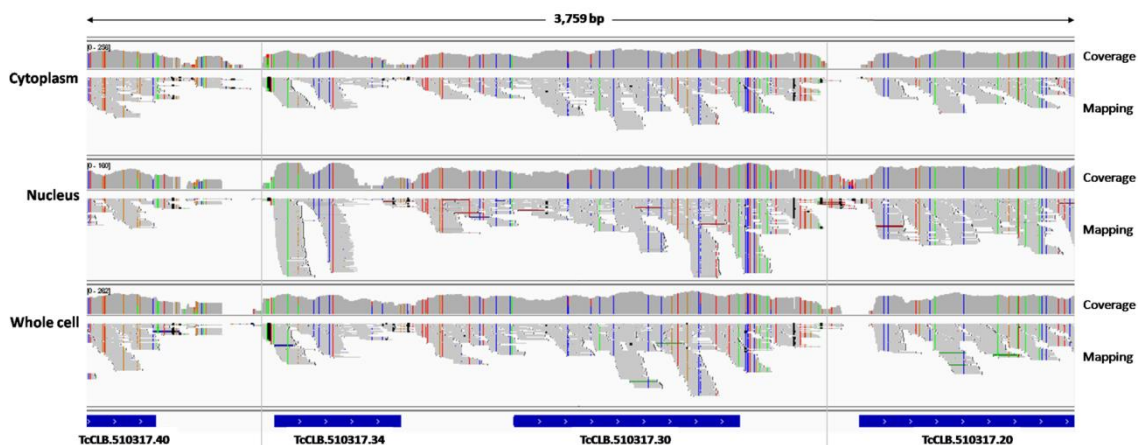


Figura 11. Mapeo de los transcriptomas total, nuclear y citoplásmico de la región que codifica para la proteína poliA polimerasa. Mapeo genómico mostrando los resultados de los transcriptomas citoplasmático, nuclear y total (de arriba a abajo) para los genes TcCLB.510317.30 + TcCLB.510317.34 correspondientes a las regiones codificantes de la poliA polimerasa de la cepa CL Brener Esmeraldo *like* (*release* 4.2). En azul se indican las secuencias codificantes (correspondientes a los exones). En *Mapping*, las líneas en gris indican los *reads* mapeados. En *Coverage* se muestra la cobertura de los genes. Las líneas verticales de colores indican cambios de bases en la secuencia respecto a la referencia.

El análisis de la secuencia de los genes que codifican la proteína PAP, muestra que los genes están anotados de forma tal que, el gen TcCLB.510317.34 corresponde al marco abierto de lectura más largo posible y no a la secuencia del exón. De la misma manera, el gen TcCLB.510317.30 comienza en el ATG que produce el marco abierto de lectura más largo, que tampoco corresponde al exón predicho, extendiéndose corriente arriba del extremo 5' del sitio aceptor de *splicing*, dentro del intrón. Por lo tanto, para realizar un mapeo correcto de este gen, fue necesario corregir las regiones codificantes de la proteína PAP. En la figura 12 se muestra el resultado del mapeo utilizando el genoma que incorpora las modificaciones antes mencionadas de la cepa CL Brener Esmeraldo *like* de *T. cruzi*. El programa CLC Workbench, reporta tanto los *reads* pareados que mapean al genoma, como los *reads* que mapean como *broken pairs*, que corresponden a lecturas pareadas que alinean a regiones que, en el genoma, están más distantes que el tamaño esperado del inserto de la biblioteca. La existencia de *broken pairs* sugiere que el transcripto de esa región es procesado. Sin embargo, este tipo de lecturas no aparece en el transcriptoma nuclear, esto, es llamativo y sugeriría que el mensajero procesado sería rápidamente exportado al citoplasma. Otra característica que debemos señalar del mapeo en el transcriptoma nuclear es que, a diferencia de lo observado en el transcriptoma total y el citoplasmático, la cobertura del gen TcCLB.510317.34 (que codifica el primer exón de la proteína) es más alta que la del TcCLB.510317.30. Esto podría deberse a la competencia entre el *cis-* y el *trans-splicing* del segundo exón (Mair *et al.*, 2000), lo que podría producir la exportación del ARN procesado al citoplasma, incrementando la velocidad y eficiencia del transporte (Luo y Reed 1999; Valencia *et al.*, 2008) y la acumulación del primer exón en el núcleo.

A su vez, es interesante destacar que, en el transcriptoma citoplasmático, se observan lecturas que mapean a la región correspondiente al intrón del gen que codifica la PAP. Esto indica que el mensajero sin procesar se encuentra también en este compartimiento.

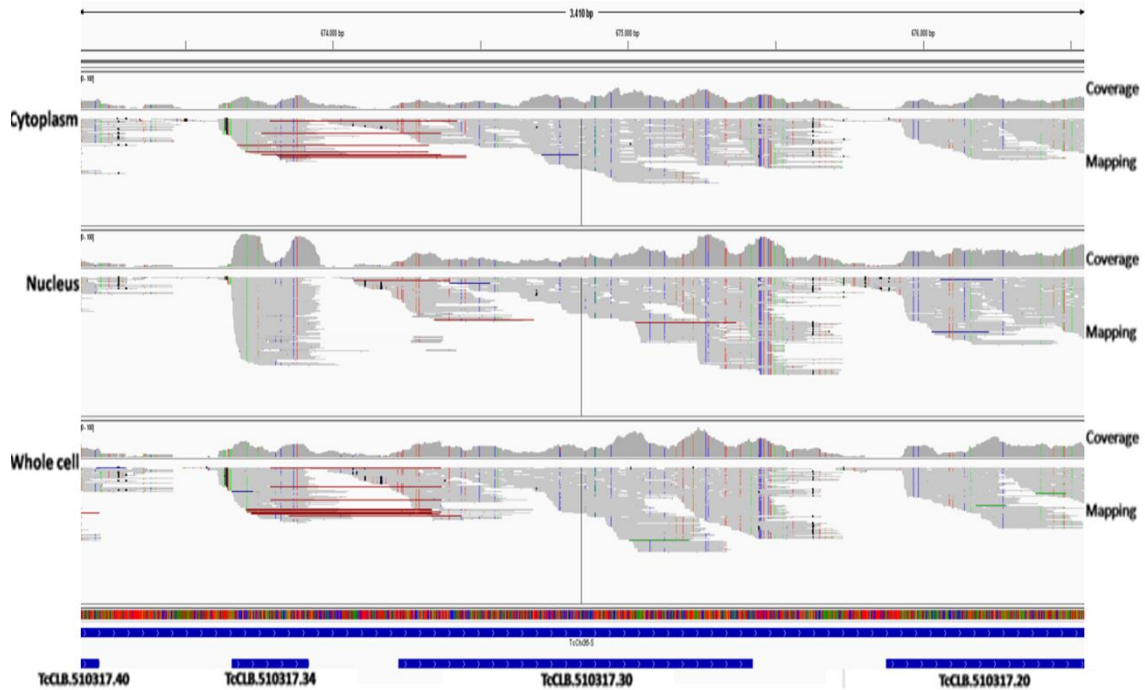


Figura 12. Mapeo de los transcriptomas total, nuclear y citoplásmico de la región que codifica para la proteína poliA polimerasa al genoma corregido de *T. cruzi*. Análisis de RNA-Seq de los transcriptomas citoplasmático, nuclear y total (de arriba a abajo) para los genes TcCLB.510317.30 + TcCLB.510317.34 correspondientes a los exones de la poliA polimerasa de la cepa CL Brener Esmeraldo *like (release 4.2)*. En azul se indican las secuencias codificantes. En *Mapping*, las líneas grises muestran los *reads* pareados mapeados y en rojo los *broken pairs* reportados. En *Coverage* se muestra la cobertura de los genes. Las líneas verticales de colores indican cambios de bases en la secuencia.

Es importante señalar, que en los transcriptomas existen *reads* híbridos, que mapean en ambos exones del gen (provenientes del ARN procesado) (figura13), estos *reads* nos permitieron la verificación del sitio de *splicing*.

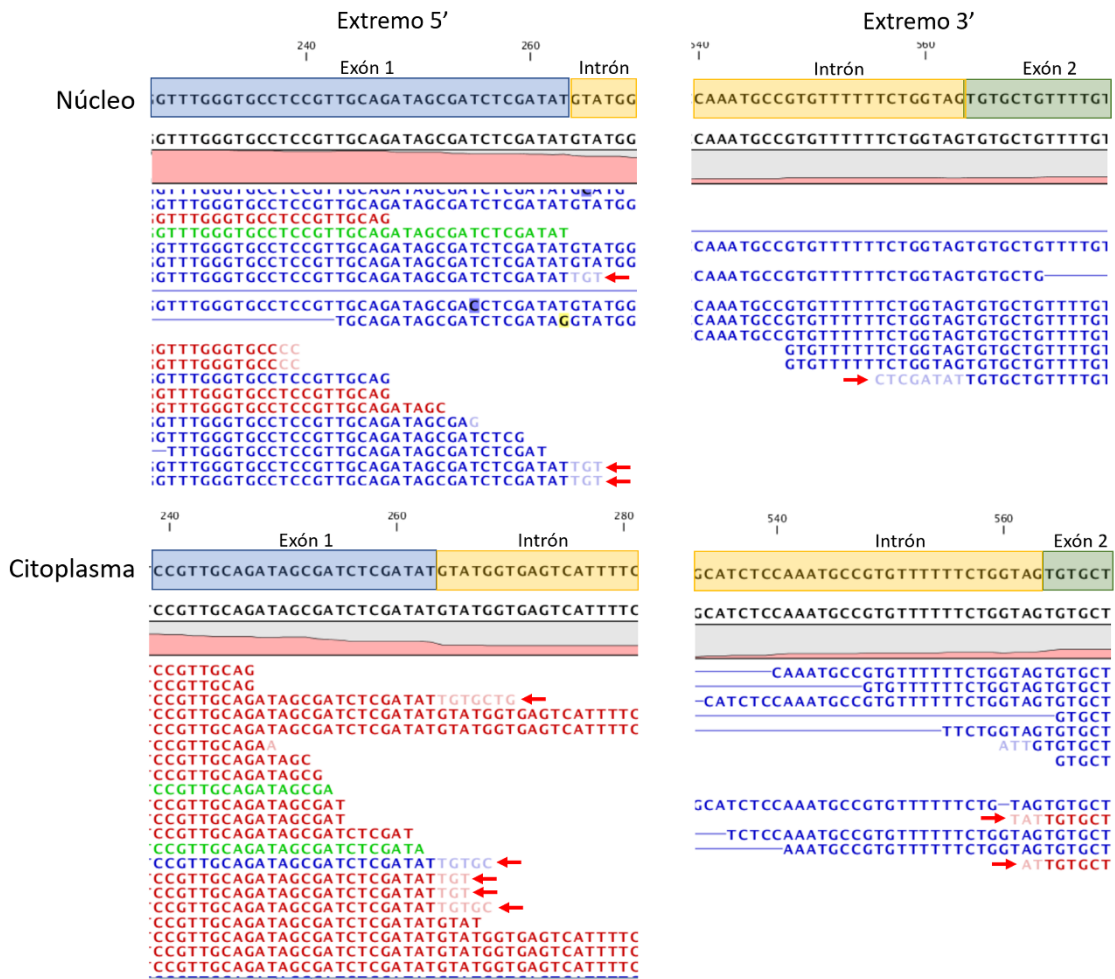


Figura 13. Verificación del sitio de *splicing* en la cepa Dm28c de *T. cruzi*. Se muestra el mapeo de los transcriptomas nuclear y citoplásmico en las regiones correspondientes al sitio de *splicing* del gen que codifica para la poliA polimerasa (TCDM_00289). En el panel superior se muestran los mapeos correspondientes al transcriptoma nuclear, en el inferior, los mapeos correspondientes al transcriptoma citoplásmico. A la izquierda se muestra el mapeo correspondiente a la región 5' del intrón y a la derecha al extremo 3'. En negro se indica la secuencia genómica, con recuadros celestes se indica la secuencia correspondiente al exón 1, en amarillo al intrón y en verde al exón 2. Las lecturas en azul corresponden a *reads* que mapean pareados, las verdes corresponden a *reads* únicos que mapean en sentido directo y las rojas corresponden a *reads* que mapean antisentido. El color más claro en los *reads* indica regiones que no alinean al genoma, con flechas se indican los *reads* híbridos, donde la secuencia más clara corresponde a la secuencia del exón empalmado.

Para comprobar si efectivamente el mensajero inmaduro de PAP está presente en el citoplasma diseñamos un ensayo de qRT-PCR que nos permitió distinguir el ARN mensajero maduro del ARN sin procesar (figura 14). El mismo se basa en la amplificación, por un lado, de una región que incluye una parte del intrón del gen, y por otro, del mensajero maduro utilizando un *primer* que hibrida en el sitio de *splicing*. A su vez, verificamos los resultados de expresión en el RNA-Seq para los transcritos de los genes TcCLB.510317.30 + TcCLB.510317.34 (exones 1 y 2 de PAP), que presentan un *foldchange* N/C de 0,72 con el resultado del qRT-PCR para el mensajero procesado que presenta una relación N/C de 0,66.

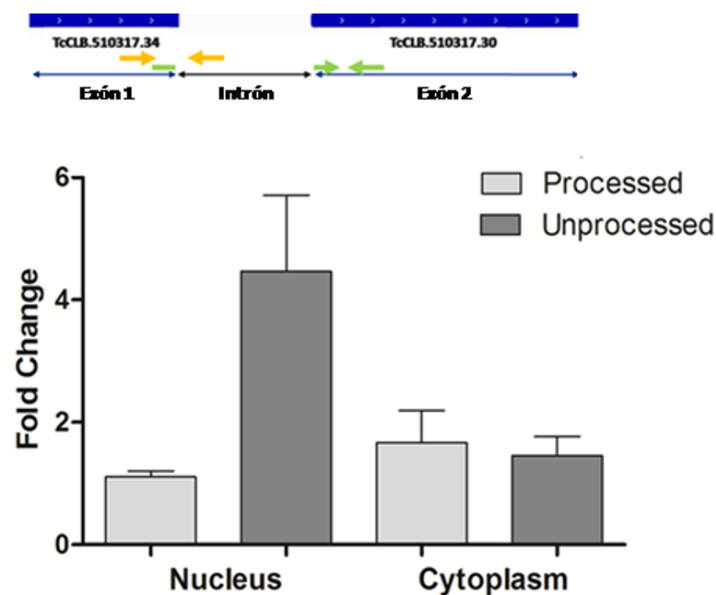


Figura 14. Expresión génica de la poliA polimerasa en *T. cruzi*. PCR cuantitativo del ARN mensajero procesado y no procesado de *pap*. Los datos son relativos al mensajero de *gapdh*. Arriba se muestra un esquema del ensayo de qRT-PCR. En azul se indican las regiones codificantes, el intrón se indica con una flecha negra. Las flechas verdes y amarillas indican los pares de *primers* utilizados en el PCR cuantitativo; las amarillas indican los *primers* utilizados para amplificar el ARN no procesado, las verdes el procesado, el primer directo es complementario a la región de *splicing* de los exones (ver tabla S1, capítulo 5.1).

Los resultados de qRT-PCR muestran que los niveles de ARN maduro e inmaduro de PAP son similares en el citoplasma. Dado que, la contaminación nuclear de la fracción citoplasmática no es superior al 5%, estos resultados confirman los resultados de RNA-

Seq, el transcripto inmaduro, que incluye el intrón, se localiza también en el citoplasma. Cabe destacar, que la presencia de secuencias intergénicas en el citoplasma y su procesamiento en el citoplasma fue previamente reportado (Jager *et al.*, 2007).

Por otro lado, es preciso señalar que los niveles de ARN mensajero inmaduro en el núcleo son más altos que los de mensajero maduro, sin embargo, en el mapeo se observan pocas lecturas en la región del intrón, como comentamos previamente (ver figura 12). Para intentar comprender este resultado, localizamos los *primers* utilizados en el ensayo de PCR en los transcriptomas citoplasmático y nuclear (figura 15).

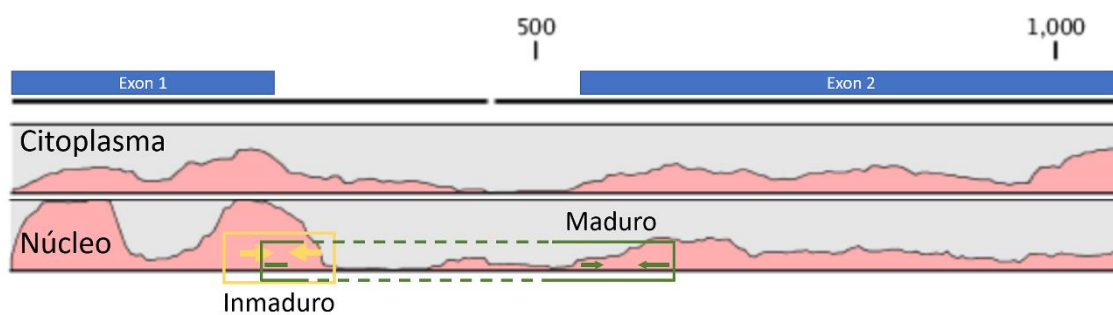


Figura 15. Localización de los *primers* utilizados en el qRT-PCR. En el esquema se muestra la cobertura del mapeo de los transcriptomas citoplásmico y nuclear a la región 5' del gen que codifica para la proteína PAP. En azul se indican las regiones codificantes. En amarillo se muestran los *primers* utilizados para amplificar el ARN inmaduro y en verde los que amplifican el ARN procesado el primer directo es complementario a la región de *splicing* de los exones (ver tabla S1, capítulo 5.1).

Como se observa en la figura 15, la región amplificada del mensajero inmaduro pertenece a una zona de alta cobertura, verificando el resultado del qRT-PCR. Este resultado sugiere que el sitio de *splicing* 5' del intrón se procesaría más lento que el 3', contrariamente a lo que sucede en el *cis-splicing* en eucariotas modernos (Konarska *et al.*, 1985). El mecanismo de *cis-splicing* en *T. cruzi* parece funcionar de forma similar al de *trans-splicing*, en donde el sitio 3' del intrón es procesado primero (Liang *et al.*, 2003) o tal vez sea menos eficiente y por lo tanto más lento. Como comentamos previamente, la competencia entre el *cis-* y el *trans-splicing* fue reportada para el mensajero de PAP

(Mair *et al.*, 2000), lo que nos hace pensar en el *splicing* como mecanismo ancestral, que en estos organismos tal vez, aun tengan roles solapados.

A su vez, este mecanismo no estaría confinado al compartimiento nuclear, tal vez porque todos los mensajeros deben ser procesados antes de ser exportados al citoplasma y si no es lo suficientemente rápido y/o eficiente en el núcleo, puede ser compensado operando también en el citoplasma.

A partir de nuestros datos de secuenciación masiva realizamos también el ensamblado *de novo* de los tres transcriptomas utilizando la herramienta disponible en el programa CLCWorkbench y una búsqueda mediante BLAST local de los transcriptos correspondientes a los mensajeros que codifican para la PAP. Estos resultados nos permitirían comprobar también la presencia del ARN sin procesar en el citoplasma. Se obtuvieron dos *contigs* que alinean con este gen. En la figura 16 se muestra el alineamiento de los ensamblados citoplasmáticos y los genes de las cepas Dm28c, y CL Brener Esmeraldo y non Esmeraldo *like*. Uno de los *contigs* alinea con el exón 1 (Ensamblado 5') y el otro con el exón 2 (Ensamblado 3'). No pudimos encontrar ningún *contig* que incluya la secuencia del intrón en ninguno de los transcriptomas. El ensamblado 1 contiene parte de la secuencia del miniexón, el exón 1 completo y 23 nucleótidos pertenecientes al exón 2. Este resultado confirma los resultados de procesamiento del mensajero de la PAP en la cepa Dm28c. El hecho de no encontrar el mensajero sin procesar en los ensamblados puede deberse a la forma en que el programa realiza los ensamblados *de novo*, ya que estos algoritmos fueron generados para realizar ensamblados de genomas, no de transcriptos que sufren procesamiento.

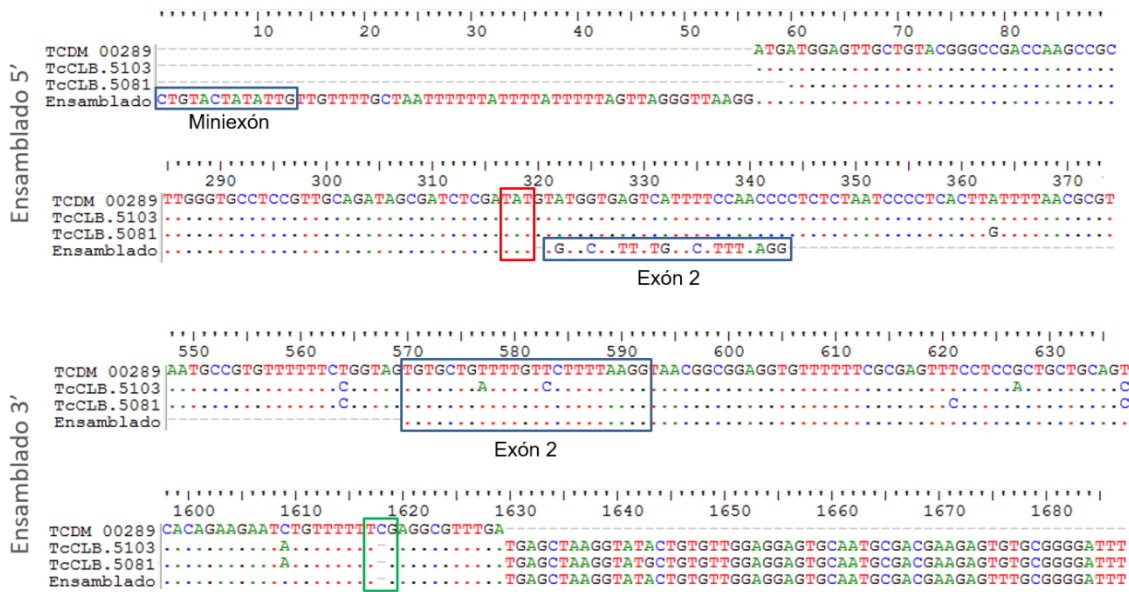


Figura 16. Alineamiento de las secuencias genómicas de la PAP y las secuencias ensambladas a partir de los *reads* del transcriptoma. TCDM_00289 (cepa Dm28c); TcCLB.5103 (TcCLB.510317.30, cepa Esmeraldo *like*); TcCLB.5081 (TcCLB.508153.440, cepa non Esmeraldo *like*). En el ensamblado 5' se destacan la secuencia del miniexón en el extremo 5', la secuencia del exón 2 en el extremo 3' y en rojo, el ultimo codón codificado en el exón 1. En el ensamblado 3' se destaca el inicio del exón 2, coincidente con el fin del ensamblado 5' y en verde la citosina en la posición 1612 de la cepa Dm28c reportada.

Por último, el ensamblado *de novo* nos permitió determinar que la citosina en la posición 1612 del gen, reportada para la cepa Dm28c, causante del corrimiento de marco de lectura, y por la cual se produce un codón de *stop* prematuro, no está presente en nuestros transcriptos, conservándose entonces, el marco abierto de lectura reportado para CL Brener. La discordancia puede ser probablemente debida a un error de secuenciación del genoma.

En conclusión, el gen que codifica para la proteína PAP en Dm28c está compuesto por dos exones interrumpidos por un intrón que posee las secuencias consenso para su procesamiento por *cis-splicing*. A su vez, pudimos determinar que el procesamiento de ARN por *cis-splicing* es funcional en el núcleo de epimastigotas de *T. cruzi* de la cepa Dm28c, pero se exporta también al citoplasma el transcripto inmaduro.

Materiales y Métodos

Análisis de secuencias

Los análisis de BLAST fueron realizados localmente, utilizando el programa Bioedit (Hall, 1999) sobre la base del genoma de *T. cruzi* CL Brener esmeraldo like *release* 5.0.

El alineamiento de las secuencias fue realizado con el programa Bioedit (Hall, 1999) utilizando parámetros por defecto.

Ensamblado *de novo*, CLC Genomics Workbench

Los *reads* pareados de obtenidos a partir de los transcriptomas total, nuclear y citoplasmático de epimastigotas de *T. cruzi* fueron ensamblados utilizando la herramienta de ensamblado *de novo* del programa CLC Genomics Workbench (Qiagen) con los parámetros por defecto del programa.

PCR cuantitativo

Los ensayos de PCR cuantitativo fueron realizados como se indica en los materiales y métodos del artículo presentado en el capítulo 5.1. Los *primers* utilizados se indican en la tabla suplementaria 1 capítulo 5.1.

Anexo

Alineamiento a nivel de nt de las PAPs de Kinetoplastideos y *P. falciparum*. TCDM_00289 (cepa Dm28c); tcruCLBren (arriba), TcCLB.510317.30 (cepa Esmeraldo *like*); tcruCLBren (abajo), TcCLB.508153.440 (cepa non-Esmeraldo *like*); tcruSylvio, STIB805_Chr03 (cepa sylvio); tbruTREU92, Tb927.3.3160 (*T. brucei*); lmajFriedl, LmjF.29.2600 (*L. major*); tvivY486, TvY486_0302520 (*Trypanosoma vivax*); tconIL3000, TcIL3000_3_2070.1 (*Trypanosoma congolense*); >NC_004327, PF3D7_0625600 (*Plasmodium falciparum*). El intrón del gen comienza en la posición 579 del alineamiento y finaliza en 1230. Los dinucleótidos AG y GA necesarios para el *splicing* se muestran en la posición 580 y 1229 del alineamiento.

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          10          20          30          40          50
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl GGCTCCCTCG AACGCTTCAC TTGCCTCGCT CTGCGCCACG GCGAAGTCTC
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          60          70          80          90         100
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl GGTGGCACCG CCAGATCAAG ATGGAGCGCG CGGCGGAGGT GACCCTCATA
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          110         120         130         140         150
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl AGGGATGCGT TTCCTCACTT CAACTTCGCC GCCGACGCAG CGGCGTCTAC
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          160         170         180         190         200
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl AGCGCTGCGG CGAGCCACAG ATGCGTTCCTC GATGGAAAAC ACGGACGAGG
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          210         220         230         240         250
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl TTGTGAGACT TGTTCCGCAA CTACTGGAGA AGATGCTGCA GGCAGAGCCG
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
                260       270       280       290       300
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl TCGGCTGCGG CTGCGTGGAC GCAAGTGTAC GTGTTTGGCT CGAGCGGTCT
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
                310       320       330       340       350
TCDM_00289 -ATGATGGAG TTGCTGTACG GGCCGACCAA GCCGCTTGAG GTTCCCCCGG
tcruCLBren ----ATGGAG TTGCTGTACG GGCCGACCAA GCCGCTTGAG GTTCCCCCGG
tcruCLBren ----ATGGAG TTGCTGTACG GGCCGACCAA GCCGCTTGAG GTTCCCCCGG
tcruSylvio ----ATGGAG TTGCTGTACG GGCCGACCAA GCCGCTTGAG GTTCCCCCGG
tbruTREU92 ----ATGGAG TGGATGTATG GCCCAACGAA ACCACTAGAG GTTCCCCGTC
lmajFriedl AGACGCCGCT ATCACGGGCA GTGACATCGA TTTGTATGGT GAGCCACTTC
tvivY486 ----ATGGAG CTTGTTACAG GCCCCACAAA GCCGCTTGAA GTTCCCCCAC
tconIL3000 -AGTGAAAAG TTGATGTACG GACCAACGAG GCCGCTAGAG GTCCCCGAGC
>NC_004327 ----- -ATGCACACT GTTT-ACGAA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
                360       370       380       390       400
TCDM_00289 CAAAGCCACG TGACTTGGCG GAGTC-CAGA AGGCTGTGCG AGGCGATGAC
tcruCLBren CACAGCCACG TGACTTGGCG GAGTC-CAAA AGGCTGTGCG AGACGATGGC
tcruCLBren CACAGCCACG TGACTTGGCG GAGTC-CAGA AGGCTGTGCG AGACGATGAC
tcruSylvio CAAAGCCACG TGACTTGGCG GAGTC-CAGA AGGCTGTGCG AGACAATGAC
tbruTREU92 CACAAGACCA TGACTACGAC GAGTC-TGAG ATGCTTTACG GCGTGTGGC
lmajFriedl CTTCTTCCCA TCGTTTCTTC TCTCC-CGTT GCGCTTAACT TCTTCCCTTC
tvivY486 CACAGTACCG TGATGAAGTC GAATC-GGCA CGGCTTTACC AGGAGACAGC
tconIL3000 CTCAGGGGTG TGAAGTTGCC GAGTC-AGAC TTAATCGTTA GAGCCACGGA
>NC_004327 AACGAGAACT TGATTGAATG TAGTAATTTT ATGCTGAACA GAAAAGATAG

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
                410       420       430       440       450
TCDM_00289 GACTGTGGTT GCTGACAGAG TGAGGGAGGC GATCGACCTT ATTGAGTTGA
tcruCLBren GACTGTGGTT GCTGACAGAG TGAGGGAGGC GATCGACCTT ATTGAGTTGA
tcruCLBren GACTGTGGTT GCTGACAGAG TGAGGGAGGC GATCGACCTT ATTGAGTTGA
tcruSylvio GACTGTGGTT GCTGACAGAG TGAGGGAGGC GATCGACCTT ATTGAGTTGA
tbruTREU92 CGAATGCCCT TCCATTAGTC CCAATCCGGT TTTAACGCTA GTGGAGTCGA
lmajFriedl CCCCCTCCCT CCCCCTCCCC CTCCCCCGTC TCGCCCGCTC TCTCCTCTCT
tvivY486 AAACGTTTCC ACCATCAGTG TAGCCCCCAC TCTGCAGCTG GTGGAGGACA
tconIL3000 CGAGGTGTCG TGTGGGAGCC CACATGCGGT GCTCGAGTTC GTGGAATCGT
>NC_004327 CCTTGAGGAA TTTTTTGGAG TGAGTGAGAT AAAAGAAGAG AATGAGGAAA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
                460       470       480       490       500
TCDM_00289 CCGCCCA-CC GTCTGGTGGG CCAAATAACC GCAGAGCCGG --AAGAGCGG
tcruCLBren CCGCCCA-CC GTCTGGTGGG CCAAATAACC GCAGAGCCGG --AAGAGCGG
tcruCLBren CTGCCCA-CC GTCTGGTGGG ACAAATAACC GCAGAGCCGG --AAGAGCGG
tcruSylvio CCGCCCA-CC GTCTGGTGGG CCAAATAACC GCAGAGCCGG --AAGAGCGG
tbruTREU92 TGGCGCT-TC GCATTGTGCA GCGGCACTCC CAAAACACTC --AGGAACAA
lmajFriedl TCTCTCTCTC TCTCTCTCTT GCGTCTCTCC TCGGCCACGG CCGTGGGCAG
tvivY486 TTGCCCT-TC GCGTTGTACA GCGGCATATT ACCGAACCCA --CGGAACAA
tconIL3000 TGGCACT-TC GCGTTGTTCA AAAATTTGTA ACTGAGACAG --AGAAGCAA
>NC_004327 CGTATTTAAA TTATAATATA GAATGT-ACA TTAAATAAGA AAATAGAAGA

```



```

                    510          520          530          540          550
TCDM_00289      TGGGTGCG-- AGCCTACCCG TTCGGTTCTT GCGGTTTGGG TGCCCTCCGTT
tcruCLBren      TGGGTGCG-- AGCCTACCCG TTCGGTTCTT GCGGTTTGGG TGCCCTCCGTT
tcruCLBren      TGGGTGCG-- AGCCTATCCG TTCGGTTCCCT GCGGTTTGGG TGCCCTCCGTT
tcruSylvio      TGGGTGCG-- AGCCTACCCG TTCGGTTCTT GCGGTTTGGG TGCCCTCCGTT
tbruTREU92      TGGGCACG-- TGCCTACCCA TTTGGTTCAT GCGGTCTTTC CGCTTCAGTT
lmajFriedl      CGTGCAGGGC AGCGCAGCGG CAGGCGTGTG GGAGAGAGCA GGAGAGAGAG
tvivY486        TGGGTTCG-- TGCTTACCCG TTTGGCTCAT GCGGCCTTGC TGCGTCCGTT
tconIL3000      TGGGCATG-- CGCGTATCCC TTTGGGTCAT GCGGTCTATC TGCTTCTGTT
>NC_004327      TAAGAAGAAT AATGT-TACG TATAGGCGTG AAATAATAAA TGAGGAGAAT

```

```

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                    560          570          580          590          600
TCDM_00289      GCAGATAGCG ATCT-CGATA TGTATGGTGA G----TC--- -----A----
tcruCLBren      GCAGATAGCG ATCT-CGATA TGTATGGTGA G----TC--- -----A----
tcruCLBren      GCAGATAGCG ATCT-CGATA TGTATGGTGA G----TC--- -----A----
tcruSylvio      GCAGATAGCG ATCT-CGATA TGTATGGTGA G----TC--- -----A----
tbruTREU92      GCAGAGAGTG ATCT-CGATG TGTATGGTGA G----TTCTT TCCTTAATCC
lmajFriedl      AGAGGGAGAG AGAGGCAGAA TGGGTGACGA GGAATCTTC GTGTCGTTC
tvivY486        GCTGATAGTG ATCT-CGATA TGTATGGTGA G----TACCT TCT--G----
tconIL3000      GCGGATAGCG ATCT-CGATA TGTATGGTGA G----TATTT CCATAA----
>NC_004327      ATACAAGAGA ATAAAAAATA TGAAGAGGGT GG-----

```

```

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                    610          620          630          640          650
TCDM_00289      ----TTTTCC AACCCCTCTC T---AATCCC CTCACTTATT TTA-----
tcruCLBren      ----TTTTCC AACCCCTCTC T---AATCCC CTCACTTGTT TTA-----
tcruCLBren      ----TTTTCC AACCCCTCTC T---AATCCC CTCACTTATT TTA-----
tcruSylvio      ----TTTTCC AACCCCTCTC T---AATCCC CTCACTTATT TTA-----
tbruTREU92      CTC-CTTTCT ACTCCCTTCT T---CTTTC TCCTCTCTT GTATATGCAT
lmajFriedl      TGCGTATACT GGAGCATCTG CAGACTTCCC ATCTCTCTT CTCTCGCTCT
tvivY486        ----TTTTCC CTGCTGCCCC C---CTCCCC ACAACTCGTC CTC-----G
tconIL3000      ----CTCTCT TTCCTCACCC T---CTCTCC CTCTCTCTT CTGCTAGCGA
>NC_004327      ----TGTTAC AGAGCCCATA T---CATTAA ATTATCCCTC TGAAG-----

```

```

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                    660          670          680          690          700
TCDM_00289      ----- --A-----C GCGTCCACA CCACACTTCC CT-----
tcruCLBren      ----- --A-----C GCGTCCACA TCACACTTCC CT-----
tcruCLBren      ----- --A-----C GCGTCCACA CCACACTTCC CT-----
tcruSylvio      ----- --A-----C GCGTCCACA CCACACTTCC CT-----
tbruTREU92      CTCCCCTCTT CTTCTCTCTT TTCTTTACTA TTCTTTTCTT TTTTGGTGT
lmajFriedl      CCTCACTCTA CTTTCGTCTT CGTTCCTTCC TCCTTTCTCT CTCGGACTCT
tvivY486        TCTCTCCCTT ---CCCATCT CTATTCCACC CCCC GCCCCG CCA-----
tconIL3000      TCTCTCTGTT GTACCACTGG GTACAACCCA CCCCTCTTTT CTC-----
>NC_004327      ----AAGATA TTAAGAGATC AAATGATGTA GTAGAATTAT TAAAGGGTTA

```

```

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                    710          720          730          740          750
TCDM_00289      -----TC TCTT----- -----G TAAAGCTACC -----
tcruCLBren      -----TC TCTT----- -----G TAAAGCTACC -----
tcruCLBren      -----TC TCTT----- -----G TAAAGCTACC -----
tcruSylvio      -----TC TCTT----- -----G TAAAGCTACC -----
tbruTREU92      CTGGGTCCCT TTTTCAATTTA TGTTGGATCA TGTGAACCCT TAAAGGAACG
lmajFriedl      CCGCATGTTT GTCCACGCAT C---GTGCG TGCGAATGAT GATGATCACG
tvivY486        -----CCTC TCTC-CGAGT TCCAGGTCTG TTAAGGACCC AAG-----
tconIL3000      -----CCTC TCTCACGTGT TTTTGGTCCC TGCCGTTCCT ACAT-----
>NC_004327      T----AATTT ATTTGAAGAA GAAAGTGGTT TAAAAAAGAG AGAAA----A

```

```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      760      770      780      790      800
TCDM_00289 -GCCTCTC-- ----- --TCGCCCTC TCTCTTTT-- ----TATT-
tcruCLBren -GCCTCTC-- ----- --TCGCCCTC TCTCTTTT-- ----TATT-
tcruCLBren -GCCTCTC-- ----- --TCGCCCTC TCTCTTTT-- ----TATT-
tcruSylvio -GCCTCTC-- ----- --TCGCCCTC TCTCTTTT-- ----TATT-
tbruTREU92 TGCTTCTGCT TTACGTGTAA TCTTTTCCCT GCTCTTCTCC CGTCTCTT-
lmajFriedl TGGAGCACCA CCTTTCTCCA CATCATGCTC ACAAATCTCG C-CACTCCT-
tvivY486 -GAATATCGG GACAAATAAG -GTTACAATT AGTAGCAT-- ----TTTT-
tconIL3000 TGCCTCTTTT TCCGCGTGTG TGTCTTCCTT ACTCTTTT-- ----TATT-
>NC_004327 TGTTTTAGGT ATGATAAATA AATTATTTCA TGAGTTTCT- ----TATTG

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      810      820      830      840      850
TCDM_00289 TATTTTCTAT TTATCA---- -----
tcruCLBren TATTTTCTAT TTATCA---- -----
tcruCLBren TATTTTCTAT TTATCA---- -----
tcruSylvio TATTTTCTAT TTATCA---- -----
tbruTREU92 CTTTTCTTCC CCTCTCACA TCGTTCCCTG TTCCTCGTTT CACGTGCCGC
lmajFriedl CCTTTCTTTC TCACCTGCTT GCCTACATTG CAAACCGTGT -GTGAACATC
tvivY486 TTTTTGTTTT ATTTTT---- -----
tconIL3000 ATTTCTGTCC TTTCTT---- -----
>NC_004327 AATTATCTAT AAATCAAG-- -----

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      860      870      880      890      900
TCDM_00289 -----GTG TTTGAACAAC
tcruCLBren -----GTG TTTGAACAAC
tcruCLBren -----GTG TTTGAACAAC
tcruSylvio -----GTG TTTGAACAAC
tbruTREU92 CATCGCCTTG TCGACGTCTG GACGCGTTCA TCTTGCCGTG GTTGTCGAAC
lmajFriedl TGTGTACACC CCTCTCATGT TGCACATTCA CGGTTCTGTG ATACTGCCGC
tvivY486 -----ACT CTCATG-ATC
tconIL3000 -----GCG CTGAGGGAAC
>NC_004327 -----GTA TAAGTGAAGA

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      910      920      930      940      950
TCDM_00289 TCGAAAAAAAA AAAAA-----TCTGCT TTTCTT----
tcruCLBren TCAAAAAAAAA AAAAA-----TCTGCT TTTCTT----
tcruCLBren TTAAAAAAAAA AAAAA-----TCTGCT TTTCTT----
tcruSylvio TCAAAAAAAAA AAAAA-----TCTGCT TTTCTT----
tbruTREU92 CTTCGTGGCT GATCTGCGTG CTGTTTCGCT TTCACTGTCT ATACATTTGT
lmajFriedl CATGGTGCAT GCGGGCGCCG TGCCGTTGCT CTCGGCGTTC TCCTCCCTCT
tvivY486 TGC--GCAAC AATAC-----TTCTCT CTTTCC----
tconIL3000 TGCAGGGACC AATCTAAATG ----TTGCT CTTGCTCAAT GTGTCATTCA
>NC_004327 AGAAGCAAAT AATATATCAG GAAATTTATA TACATT----

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      960      970      980      990      1000
TCDM_00289 --T-CCTTTT TTTTT--TT T-ATATTCCG TGCCGGTTCGT T-----TTC
tcruCLBren --T-CCTTTT TTCTT--TT T-ATATTCCG TGCCGGTTCGT T-----TTC
tcruCLBren --T-CCTTTT TTTTT--TT TTATATTCCG TGCCGGTTCGT T-----TTC
tcruSylvio --T-CCTTTT TTTTT--TT --ATATTCCG TGCCGGTTCGT T-----TTC
tbruTREU92 GTTTATTTCC AACATTGACC ATTGCTTCTT CCTTCGTTAA TGATGTGATT
lmajFriedl TCAAGGTTTT GTTCTCGCTT TTTCTTCTG CACGCCTCAC GCCT--TTTT
tvivY486 --TTCTCTTT CGCACCG-TC TTTCTTCTG GGCTCGATGC T-----CC
tconIL3000 TATTCCTTTT CCCTCCG-TC TTTCTTCTT CGCGCGCCGC TGTTGGTTCC
>NC_004327 TGTTCTTAT CGTTTAGGTG TAATAACTCC AAATAGTGAT A-----TA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
            960      970      980      990      1000
TCDM_00289  --T-CCTTTT TTTTT---TT T-ATATTCCG TGCCGGTCGT T-----TTC
tcruCLBren  --T-CCTTTT TTCTT---TT T-ATATTCCG TGCCGGTCGT T-----TTC
tcruCLBren  --T-CCTTTT TTTTT---TT TTATATTCCG TGCCGGTCGT T-----TTC
tcruSylvio  --T-CCTTTT TTTTT---TT --ATATTCCG TGCCGGTCGT T-----TTC
tbruTREU92  GTTTATTTC AACATTGACC ATTGCTTCTT CCTTCGTTAA TGATGTGATT
lmajFriedl  TCAAGGTTT GTTCTCGCTT TTTCTTCTG CACGCCTCAC GCCT--TTTT
tvivY486    --TTCTTTT CGCACCG-TC TTTCTTGCT GGCTCGATGC T-----CC
tconIL3000  TATTCCTTT CCCTCCG-TC TTTCTTCTT CGCGCGCCGC TGTTGGTTCC
>NC_004327  TGGTCTTAT CGTTTAGGTG TAATAACTCC AAATAGTGAT A-----TA

      ....|....| ....|....| ....|....| ....|....| ....|....|
            1010     1020     1030     1040     1050
TCDM_00289  AATTGAAAAC GCT-----C GCAC-----
tcruCLBren  AATTGAAAAC GCT-----C GCAC-----
tcruCLBren  AATTAAAAAC GCT-----C GCAC-----

tcruSylvio  AATTGAAAAC GCT-----C GCAC-----
tbruTREU92  ATCTAAATAT GCATATACAT GTATAATCAT ATATGTACAC GTATACATTT
lmajFriedl  ACCCCGAGGC GTAAAC-TGC GACGGCATGG TGTGCTCTGC GTATTTCCACC
tvivY486    ATCCGGTCAC GAA-----AT CGAC-----
tconIL3000  ACCTGCACGC TTTCTCTAC  CGGG-----
>NC_004327  GATTGTATTT TTTTG---GC ACCC-----

      ....|....| ....|....| ....|....| ....|....| ....|....|
            1060     1070     1080     1090     1100
TCDM_00289  -----T G--TCTTGAA TT-----C
tcruCLBren  -----T G--TCTTGAA TTAT-----T
tcruCLBren  -----T T--TCTTGAA TTTT-----T
tcruSylvio  -----T G--TCTTGAA TT-----C
tbruTREU92  GTGTATATGT GTATATTTTG TTGATGTGAG TCTGTTCGTC GTCGTGGCAC
lmajFriedl  GTGCGCGCCT GTACTGCAAC ATGCCTAATG GTAATACGCT ACGATGGCGG
tvivY486    -----T G-ACGCTTCA TTTACC----T
tconIL3000  -----T GCATCGTTGC TTCGCG----C
>NC_004327  -----C AGAATTTGAC AAGAGAA---A

      ....|....| ....|....| ....|....| ....|....| ....|....|
            1110     1120     1130     1140     1150
TCDM_00289  TTTTTTTTTT CACAC-----T
tcruCLBren  TTTTTTTTTT CACAC-----T
tcruCLBren  TTTTTTTTTT CACAC-----C
tcruSylvio  TTTTTTTTTT CACAC-----T
tbruTREU92  TCGTGACCCG CACATGGTAC TGACGGCAAA CACGGATGCT GATTTTCCAT
lmajFriedl  TGCACCTATG CGCACACGCC GCCCCTTCAC TGC-----TC TCTGCTTTTT
tvivY486    CTACGCCATG CGCAC-----T
tconIL3000  CCCCGTCAA TTCGT-----C
>NC_004327  TATTTTTTAA TGAAT-----

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1160      1170      1180      1190      1200
TCDM_00289 GTCCACTCCG TGTGCATCTC -----
tcruCLBren GTCCACTCCG TGTGCATCTC -----
tcruCLBren GTCCACTCCG TGTGCATCTC -----
tcruSylvio GTCCACTCCG TGTGCATCTC -----
tbruTREU92 TTTTTTTTTG TCAAAACCCC GCAACCATAC CATTTC AACACAATC
lmajFriedl CCTCTTGCTG CTCGCACCTT TCTCTT---- -T ATGAGTTGTG
tvivY486 TGTTG-CACG -CTGCATACT GA----- -T
tconIL3000 TGTTTTTCGCG TCTGAACCCC AC----- -T
>NC_004327 TTTATTTAAA ATTACAACAA GATA----- --GAAATGTA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1210      1220      1230      1240      1250
TCDM_00289 CAAATGCCGT GTTT--TTTC ---TG-GTAG TGTGCTGTTT TGTTCTTTTA
tcruCLBren CAAATGCCGT GTTT--TTTC ---CG-GTAG TGTGCTGTTT TGTTCTTTTA
tcruCLBren CAAATGCCGT GTTT--TTTC ---CG-GTAG TGTGCTGATT TGTCCTTTTA
tcruSylvio CAAATGCCGT GTTT--TTTC ---TG-GTAG TGTGCTGTTT TGTTCTTTTA
tbruTREU92 ACAACGCCGA ATTACGTTGA ATTTGTGAAG CGCGCTGATT TGTCATCTG
lmajFriedl CACTGGTCGT TTGCGCATGC TGCCGCGAAG AATTGCATTG TGCTCCGCCT
tvivY486 CAAATGAATA TCGT--TTAT A--TG-AAAG TGTTCTGTTT TGCCCATTCA
tconIL3000 TCAATGTTGA TTGA--TTGC AAATG-GTAG CGCTGTATTT TGTCGGTTTG
>NC_004327 AAAAAGTTAC AAGCATTACC AGAAACGTAT ACACCTATTA TACAATTTAT

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1260      1270      1280      1290      1300
TCDM_00289 AGGTAACGGC GGAGGTGTTT TTTCGCGAGT TTCCTCCGCT GCTGCAGTCT
tcruCLBren AGGTAACGGC GGAGGTGTTT TTTCGCGAGT TCCCTCCGCT GCTGCAGCCT
tcruCLBren AGGTAACGGC GGAGGTGTTT TTTCGCGAGT TTCCTCCACT GCTGCAGCCT
tcruSylvio AGGTAACGGC GGAGGTGTTT TTTCGCGAGT TTCCTCCGCT GCTGCAGTCT
tbruTREU92 TTTTAACGAC TGATATCTTT TTCGATGAGT TTCCGAGGTT GTTATATAAC
lmajFriedl CTGTGAACAC GAGTTTGTGTTT TTTCAGACGT TCCCAGCCTC GGCTGAGGGG
tvivY486 AAGTTACTGC TGCCGTTTTT TTTCACGAGT TTCCTCTGTT GCTGCAGACG
tconIL3000 TTGTCACCCG TTTCATCTTC TTTCATGAGT TTCCAGCGCT CCTCGAGACG
>NC_004327 GTATGACGAT GTTGACATTG ATTTGTTA-T TAGCAACTTT ACCATATAAA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1310      1320      1330      1340      1350
TCDM_00289 GTCGTGGGGA ACGTGGTTTT GGTTCCT--- GCCAGGGTTC CTGTGGTGAA
tcruCLBren GTCGTGGGGA ACGTGGTTTT GGTTCCT--- GCCAGGGTTC CTGTGGTGAA
tcruCLBren GTCGTGGGGA ACGTGGTTTT GGTTCCT--- GCCAGGGTTC CTGTGGTGAA
tcruSylvio GTCGTGGGGA ACGTGGTTTT GGTTCCT--- GCCAGGGTTC CTGTGGTGAA
tbruTREU92 TCTGTCCGTC CCGTGTCTGT AGTCGCT--- GCCAGAGTTC CCGTTGTTAA
lmajFriedl CGTCTGCAGA ACGTCTGCTG GGTCCACCTCG GCTCGCTCC CTCTGATCAA
tvivY486 GTGGTGGGCC ACGTGATTGT TGTTGCC--- GCACGTGTAC CTGTGCTGAA
tconIL3000 ATTGTTGGCC ATGTGATTGT CGTCGCT--- GCTCGAGTTC CTGTTGTGAG
>NC_004327 ACATT--AAA AGATTGTTAT TATTCTTT-A GATAATGATC ATATTTTGAA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1360      1370      1380      1390      1400
TCDM_00289 G--TTCAGTT ACAGTGGTAC TTCGGTGGAT G--TCGTGTT TGTTTCAGCG
tcruCLBren G--TTCAGTT ACAGTGGTAC TTCGGTGGAT G--TCGTGTT TGTTTCAGCG
tcruCLBren G--TTCAGTT ACAGTGGTAC TTCGGTGGAT G--TCGTGTT TGTTTCAGCG
tcruSylvio G--TTCAGTT ACAGTGGTAC TTCGGTGGAT G--TCGTGTT TGTTTCAGCG
tbruTREU92 A--TTTGAGT ACAGAGGCAC AGCTGTCGAT G--TTGTGTT TGTTTCTGTT
lmajFriedl G--TTTTTCGT ACAACGGGGT TCACGTCGAT C--TCCTGTT TGCCCTCCGTT
tvivY486 G--TTTGCCC ACGAAGGAAC GCTGGTTGAT T--TGGTGTG TGATTCAGTT
tconIL3000 G--TTTGTCT ACCAGGGTAT GTCAGTGGAT G--TTGTGTT TATATCAGTT
>NC_004327 AAATTTAGAT GAAGTAACAG TTCGGTCATT AAACGGTATT AGAGTTGCTG

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1410      1420      1430      1440      1450
TCDM_00289 GCTC--TGTC CGCTCCCCC ACGACAGAGC AATTATTAGA TGATTCCTTC
tcruCLBren GCTC--TGTC CGCTCCCCC ACGACAGAGC AATTATTAGA TGATTCCTTC
tcruCLBren GCTC--TGTC CGCTCCCCC ACGACAGAGC AATTATTAGA TGATTCCTTC
tcruSylvio GCTC--TGTC CGCTCCCCC ACGACAGAGC AATTATTAGA TGATTCCTTC
tbruTREU92 GGAC--TGCC ACAGCCGCTT ACGGAGGAGC AGATGCTGGA TGATTCCTTC
lmajFriedl GATA--TGCG GACGGCACCT GACACGAATG AGCTGCTGCG CGACGACTTT
tvivY486 GCCC--TTGC TGCACCGCCA ACTACTGAGC AATTGCTCGA TGATTCCTTC
tconIL3000 GCGC--TTTC GCAACCACCC ACAAGTGATC AGCTGCTGGA CGACTCATTT
>NC_004327 ACTTAATATT AAATTCTGTA CCAATAAAG ATTATTTTAG AAATACATTA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1460      1470      1480      1490      1500
TCDM_00289 CTTCTTCAGG TCTCCCGTGA CACCCGTCAC AGTGTAATG GTT-TACGCA
tcruCLBren CTTCTTCAGG TCGCCCGCGA CACCCGCCAC AGTGTAATG GTT-TACGCA
tcruCLBren CTTCTTCAGG TCGCCCGTGA CACCCGTCAC AGTGTAATG GTT-TACGCA
tcruSylvio CTTCTTCAGG TCTCCCGTGA CACCCGTCAC AGTGTAATG GTT-TACGCA
tbruTREU92 CTACTGAAGG TGGCTCGAGA AACCCGACCC AGTGCGAACG GAA-TCCGTT
lmajFriedl CTTTCCCTGG TGTCGCTGCC CTGTCGGGCA ACCGTGAACG GGA-TCCGCA
tvivY486 CTATTGCAAG TGGCCAGGGA AACCGGTCAC AGCGTCAATG GTA-TTCGCT
tconIL3000 CTGCTACAGG TTGCCCGTGA CACACGACCG AGCGTTAATG GAA-TTCGCT
>NC_004327 AGATTTATAA AATTATGGGC AAAAAGTAGA GGTATATATA GTAACATATT

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1510      1520      1530      1540      1550
TCDM_00289 CGGTCTT--- GGAACTA--- ---AAGCGGC GTCTACCGGT GGCGTATGAC
tcruCLBren CGGTCTT--- GGAAGTA--- ---AAGCGGC GTCTACCGGT GGCGTATGAC
tcruCLBren CGGTCTT--- GGAAGTA--- ---AAGCGGC GTCTACCGGT GGCGTATGAC
tcruSylvio CGGTCTT--- GGAACTA--- ---AAGCGGC GTCTACCGGT GGCGTATGAC
tbruTREU92 TCACGTT--- TGAGATT--- ---AAGCGGC GACTTCCGGT GCCGTACGAT
lmajFriedl CGATTCT--- GGAGATT--- ---CGTCGCC GCCTTCCCCT TCTACTGGCC
tvivY486 TCACATT--- GGAGATA--- ---AAGCGGC GCTTGCCGTG GACTTATGAC
tconIL3000 TTACGTT--- CGAAGTG--- ---AGACGCC GTCTTCCCGT TCCCATAAG
>NC_004327 AGGTTTCTTA GGAGGTATTT CATGGGCATT ATTAAGTCA AAAATTTGTC

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      1560      1570      1580      1590      1600
TCDM_00289 GTCTTCACGA CGGTACTTAA GTCCGTGAAG TTATGGGCTA TGCGGCGAAA
tcruCLBren GTCTTCACGA CGGTACTTAA GTCCGTGAAG TTATGGGCTA TGCGGCGAAA
tcruCLBren GTCTTCACGA CGGTACTTAA GTCTGTGAAG TTGTGGGCTA TGCGGCGAAA
tcruSylvio GTCTTCACGA CGGTACTTAA GTCCGTGAAG TTATGGGCTA TGCGGCGAAA
tbruTREU92 GTCTTCACTG CCGTCTGAA GACAGTAAAG CTATGGGCGA TGCGTCGGAT
lmajFriedl GCTTACGCCT GCGTGCTGCG CGCGGTCAAG TACTGGGCGG CGCAGCGGCA
tvivY486 GTTTTTAGTA CTGTTCTGAA GGTGTAAAG CTCTGGGCCA TGAAACGGCA
tconIL3000 GTCTTCGCTA CGGTCTAAA AGCTGTGAAG TTTTGGGCCA TGCGTCGACA
>NC_004327 AGTTATACCC AAATTTTAAAT GTTAGTCAAT TAATATGTAA ATTTTTTAAA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      1610      1620      1630      1640      1650
TCDM_00289 GGT----CTA T-----G GAAATCTTTA TTGTTATCCA AGTGGCGTGG
tcruCLBren GGT----CTA T-----G GAAATCTTTA TTGTTATCCA AGTGGCGTGG
tcruCLBren GGT----CTA T-----G GAAATCTTTA TTGTTATCCA AGTGGCGTGG
tcruSylvio GGT----CTA T-----G GAAATCTTTA TTGTTATCCA AGTGGCGTGG
tbruTREU92 GGT----TTA C-----G GCAATGTCTA CACGTATCCG AACGGCGCGG
lmajFriedl GGT----GTA T-----G GGAACCTCTA CACCTTCCCC AACGGCGTGT
tvivY486 GGT----ATA C-----G GGAATTTGTA TACCTATCCG AGTGGTGTGT
tconIL3000 GGT----GTA C-----G GTAACATCTA TACTTATCCG AATGGCGCGG
>NC_004327 GTTTATTCTA TATGGAATTG GAAATATCCA GTACTTTTAC AAAATATTA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      1660      1670      1680      1690      1700
TCDM_00289 TTCTTGCCAT CATGGTGGCT CGGAT---TT GTCAGGTCAT G--CCTGCGT
tcruCLBren TTCTTGCCAT CATGGTGGCT CGGAT---TT GTCAGGTCAT G--CCTGCGT
tcruCLBren TTCTTGCCAT CATGGTGGCT CGGAT---TT GTCAGGTCAT G--CCTGCGT
tcruSylvio TTCTTGCCAT CATGGTGGCT CGGAT---TT GTCAGGTCAT G--CCTGCGT
tbruTREU92 TGCTTGCCAT CATGGTGGCT AGGGT---TT GTCAGGTGCT T--CCATCTT
lmajFriedl GTCTCGCCAT TATGGTGGCC CGTGC---CT GCCAGCTTTG C--CCCGTCG
tvivY486 TCCTTGCAAT TATGGTTCGCG CGTGC---AT GCCAGGTTAT G--CCCGCTT
tconIL3000 TGTTTCGCCAT CATGGTGGCT AGAGT---GT GCCAGGTCTT A--CCCTCTC
>NC_004327 AAAATATAAT ATTGAAGGTT TAAATAATTT TCCAGTTTGG GATCCTGAAA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      1710      1720      1730      1740      1750
TCDM_00289 CTCATCCCAA CGTCCTTTTG CGCTTTTTTT TCTTGTTTTA TACTCAATGG
tcruCLBren CTCATCCCAA CGTCCTTTTG CGCTTTTTTT TCTTGTTTTA TACTCAATGG
tcruCLBren CTCATCCCAA CGTCCTTTTG CGCTTTTTTT TCTTGTTTTA TACTCAATGG
tcruSylvio CTCATCCCAA CGTCCTTTTG CGCTTTTTTT TCTTGTTTTA TACTCAATGG
tbruTREU92 CTCATCCCAG TACTCTTCTT CGATTTTCTT TTCTTTTCTA CACGCAATGG
lmajFriedl CGGACAGTGG GGTCAATTTG CGCTTCTTCT TCTCCCTGTA CGTGTGGTGG
tvivY486 CCCATCCAAA CGTGGTCTG CGGTTTTTCT TCCTTTTTTTA CACACAATGG
tconIL3000 CGTACCCAG TGTTCTTCTG CGGTTCTTCT TCCTTTTTTTA CACTCAGTGG
>NC_004327 AAAATATTAAGATAAATTA CACATTATGC -CTATCATTACACTGCTTT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      1760      1770      1780      1790      1800
TCDM_00289 CTGAGCCGCC ATGATCGTAT TTCTCCAGTC TATATTACCA CATCTTTAGA
tcruCLBren CTGAGCCGCC ATGATCGTAT TTCTCCAGTC TATATTACCA CATCTTTAGA
tcruCLBren CTGAGCCGCC ATGATCGTAT TTCTCCAGTC TATATTACCA CATCTTTAGA
tcruSylvio CTGAGCCGCC ATGATCGTAT TTCTCCAGTC TATATTACTA CATCTTTAGA
tbruTREU92 ATGAGCCGCC ATGATCGTAT TTCACCCGTG TATTTGACTG CGACTCTGGA
lmajFriedl CTCCTGCGTG ACACCCGCAT CGCTCCTGTT TCTATTGTGC CCATGGATGA
tvivY486 ATGAGCCGGC ATGATCGCAT CTCACCTGTA TACATTACCA CCTCCTTACA
tconIL3000 ATGAGTCGGC ATGATCATAT ATCACCCGTG TACTTGACAG CAACGCTCGA
>NC_004327 CCCATGT--- ATGAATTCAA CACACAATGT TACATATTGT ACACGTAATA

```

```

.....|.....|.....|.....|.....|
      1810      1820      1830      1840      1850
ICDM_00289 GTCTCGGGGC CGCATTCCAG GCCTTCCAGA CTCCTGGGAT CCGCG-GCGT
tcruCLBren GTCTCGGGGC CGTATTCCAG GTCTTCCAGA CTCCTGGGAT CCGCG-GCGT
tcruCLBren GTCTCGGGGT CGTATTCCAG GCCTTCCAGA CTCCTGGGAT CCGCG-GCGT
tcruSylvio GTCTCGGGGC CGTATTCCAG GCCTTCCAGA CTCCTGGGAT CCGCG-GCGT
tbruTREU92 GGGTCGCGGT CGCATTCCCG GGCTGCCAGA CTCATGGAAG CCGTC-GAAA
lmajFriedl GAA-CGCGGC C--ATCGCGC GAGTGCCGGG GATGCCGCCG CCGTG-GGAT
tvivY486 AGGTTCGTGGA CGTGTTCAG GCATGCCAGA GCCCTGGGAC CCGGC-ACTT
tconIL3000 AAGCCGCGGT CGCATCCCTG GCTTACCGGA GTCGTGGAAG CCAGC-AAAG
>NC_004327 TTCTAATAGA AGAATTCAAA AGAGCTCATT CTATTATCCA ATATATGGAA

```

```

.....|.....|.....|.....|.....|
      1860      1870      1880      1890      1900
ICDM_00289 GACGC--GTG TCGTGACG-- ATTTGTTGCC TGTTATAAAT CCCG--CATA
tcruCLBren GACGC--GTG TCGTGACG-- ATTTGTTGCC TGTTATAAAT CCCG--CATA
tcruCLBren GACGC--GTG TCGTGACG-- ATTTGTTGCC TGTTATAAAT CCCG--CATA
tcruSylvio GACGC--GTG TCGTGACG-- ATTTGTTGCC TGTTATAAAT CCCG--CATA
tbruTREU92 GATAA--AAT TTGCTGCG-- ATTTGTTTCC CGTGATCAGT CCGG--CGTA
lmajFriedl GCCGTCTGGG ACGCGGCGG- ACCTGTTCCC AGTGCTGAAC CCGG--CACG
tvivY486 GAAAA--GTT TAGCCACG-- ACCTCCTGCC TGTTCTAAAC CCGG--CTTA
tconIL3000 AATGA--GTG TTTTCGTG-- ACCTTTTCCC CGTGATCAGC CCTG--CGTA
>NC_004327 GATGAAAATA ATATCAAATC ATCTATACGT ATGAATAAAT TGTGGCCAAA

```

```

.....|.....|.....|.....|.....|
      1910      1920      1930      1940      1950
ICDM_00289 CCCTTACGTC AACGACGCTC GCAACGTCGG TCGCTGCAGT TTGGAGGTGT
tcruCLBren CCCTTATGTC AACGACGCTC GCAACGTCGG TCGCTGCAGT TTGGAGGTGT
tcruCLBren CCCTTACGTC AACGACGCTC GCAACGTCGG TCGCTGTGGT TTGGAGGTGT
tcruSylvio CCCTTACGTC AACGACGCTC GCAACGTCGG TCGCTGCAGT TTGGAGGTGT
tbruTREU92 TCCATATGTG AACGATGCAA GTAGCGTTGG GCGATGTGGG CTGGAGGCGC
lmajFriedl ACCCACCGTC AACGCCGCCC ACGCGGTGGG CCGCAGTGGG CTGCAGCTCT
tvivY486 TCCCTTTGTG AATGATGCCA ATGCTGTGGG ACGCTGTGGG CTGGAGGTGT
tconIL3000 CCCATACGTG AATGATGCGA GTAACGTTGG CCGCTGTGGA TTGGAGTCAC
>NC_004327 TAATAATGTA AATAATATTA ATAATAGTAG TAATAACGTT TGGATGACCA

```

```

.....|.....|.....|.....|.....|
      1960      1970      1980      1990      2000
ICDM_00289 TTTACGCGGA GTTAACATAC GCTTATCGTC TCCTGTCCAA CCTAGAGACG
tcruCLBren TTTACGCGGA GTTAACATAC GCTTATCGTC TCCTGTCCAA CCTAGAGACG
tcruCLBren TTTACGCGGA GTTAACATAC GCTTATCGTC TCCTGTCCAA TTTAGAGACG
tcruSylvio TTTACGCGGA GTTAACATAC GCTTATCGTC TCCTGTCCAA CCTAGAGACG
tbruTREU92 TTTACAGCGA GATCACACGG GTGCAATGCA TTCTGACGGA GGCGCGGACT
lmajFriedl TTTTGAAGGA GCTCCTCCGC GCTGAGCAGC TGTGTGCCCT AGTCCCAGATG
tvivY486 TTTATTCTGA GGTGACGCGT GCGCATCATC TGCTTTCAAC CACCGAGTCG
tconIL3000 TTTATTCTGA GATCACGCAA GCACAGCGGG TAATTTCCGC GAGCGACACC
>NC_004327 TTTTGAAGGA ATTAG-ATTT ATTCAGTA-C TTATAACAT TTCCTACATA

```

```

.....|.....|.....|.....|.....|
      2010      2020      2030      2040      2050
ICDM_00289 C---CCCTAG AGACGATATG GGAGCCTTAT CACATTTTGG ATGATTATTC
tcruCLBren C---CCCTAG AGGCGATATG GGAGCCTTAT CACATTTTGG ATGATTATTC
tcruCLBren C---CCCTAG AGACGATATG GGAGCCTTAT CACATTTTGG ATGATTATTC
tcruSylvio C---CCCTAG AGACGATATG GGAGCCTTAT CACATTTTGG ATGATTATTC
tbruTREU92 CTCCCCTTGG AAGAGATGTG GGAACCGTAT AGAATAGAGG AAGAGTATTC
lmajFriedl T---CGTACT CGGCACTATG GGCACCGTAC AACATACTCG ATGAGCACCG
tvivY486 TTTCCCCTGG AGGATGTCTG GAAACCATAC TCTATCACTG ATGACTTTTC
tconIL3000 CTGTCACTGG AGAAGCTTTG GAGGCCCTAC TGCATCATGG AAGACTTTGC
>NC_004327 TACAAATTAT GGGTATTAAT GAATTTATAT ATAATAGTTG GAAAGGAT-G

```



```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      2060      2070      2080      2090      2100
TCDM_00289 CACTTTTTTT GTGGTACATG TTACTTGCGA GGAA---GAA AATGAAGAAA
tcruCLBren CACTTTTTTT GTGGTACATG TTACTTGCGA GGAA---GAA AATGAAGAAA
tcruCLBren CACTTTTTTT GTGGTACATG TTACTTGCGA GGAA---GAA AATGAAGAAA
tcruSylvio CACTTTTTTT GTGGTACATG TTACTTGCGA GGAA---GAA AATGAAGAAA
tbbruTREU92 AACCTTCTTT GTTGTGGGTG TAAGTTGTGC GGGC---CAC AGTATGGCGG
lmajFriedl CTCTTTTGTG GCGTGCACA TCTCGAGCGA GCAC---CAG TCACTTGCAG
tvivY486 GTCCTTCTA TTGGTGCAGG TGAGTTGTGA CGGC---AAC AGCGAAGAAG
tconIL3000 TCACTTCTTT CTAGTGGAAG TGAGCAGCCT GGGT---GAT GATATGGCGG
>NC_004327 GATTGAAAGC AAAATACGTC TTTTATTCAA AAAATTAGAA ACTATCAATG

```

```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      2110      2120      2130      2140      2150
TCDM_00289 AATTGGAGGC -GGTCTTTT CGTATGGAGC TCCTACGTGC TGAGTAAACT
tcruCLBren AATTGGAGGC -GGTCTTTT CGTATGGAGC TCCTACGTGC TGAGTAAACT
tcruCLBren AATTGGAGGC -GGTCTTTT CGTATGGAGT TCCTACGTGC TGAGTAAACT
tcruSylvio AATTGGAGGC -GGTCTTTT CGTATGGAGC TCCTACGTGC TGAGTAAACT
tbbruTREU92 AGACTGAGCA -GGCGCTTTC TGTGTGGAGT TCCTATGTGG CCAGCAGGTT
lmajFriedl CGTGCGAGGA -AGCTATCAA TGCATGGAAG GGGTACGTAG AAAGCAAGTT
tvivY486 GTGTGGA AAA -TGACCTAGA AGTCTGGAGT TCTTATGTGG CTAGCAA ACT
tconIL3000 ATGCGGAGCG -TATACTTGG TAGATGGGGT TCATACGTGG GCAGTAA GTT
>NC_004327 AATTAAAAAT TAGACCTTTT CCTAAATCT ATTTCTATAA AAAAGAAAAA

```

```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      2160      2170      2180      2190      2200
TCDM_00289 TCGGATTTTA CTTTATGCAC TGGAGCGTCT TGTGGACGCT CGTCCATATC
tcruCLBren TCGGATTTTA CTTTATGCAC TGGAGCGTAT TGTGGACGCT CGTCCATATC
tcruCLBren TCGGATTTTA CTTTATGCAC TGGAGCGTAT TGTGGACGCT CGTCCATATC
tcruSylvio TCGGATTTTA CTTTATGCAC TGGAGCGTCT TGTGGACGCT CGTCCATATC
tbbruTREU92 ACGAATCCTC ATATACTCCA TAGAGCGTCA TGCACAGGCC CGTCCCTGTC
lmajFriedl GCGTATGTTT GTGTACGCGC TGGAGTGCCT GCGCGAGGTG CGGCCGTTCC
tvivY486 CCGGATTCCT ATCTACGCTG TGGAGTGCAT CGTTGACGCC CGTCCATATC
tconIL3000 GCGCTTCTCT ATTTATTCTG TCGAAAAGGTT TGTTACGCG CGACCCGACC
>NC_004327 TCATATTATT GTTCATCCTT T-----TT TATAG-CACT CATATTCTTT

```

```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      2210      2220      2230      2240      2250
TCDM_00289 CCCAAAAGCT GA-----ATG ATGGTCCACC AAGGTCTGTA CAAAATCCG
tcruCLBren CCCAAAAGCT GA-----ATG ATGGTCCACC AAGGTCTGCA CAAAATCCG
tcruCLBren CCCAAAAGCT GA-----ATG ATGTTCCACC AAGGTCTGTA CAAAACCCG
tcruSylvio CCCAAAAGCT GA-----ATG ATGGTCCACC AAGGTCTGTA CAAAATCCG
tbbruTREU92 CGCGTAAGAT AA-----GGC CAAAAGGA TAATCACGTG GGGGGTCTT
lmajFriedl CACGCCCGGT TGTCGATGAG CCGCCCCGAG CGGTGACTCG GAGAGGGGTG
tvivY486 CTCACAAACT TA-----AGG ACCGCGTTGC CACCAGTCTT TTGGCCCCG
tconIL3000 CACACAAACT AA-----GGG CTAAAATGGA TGGTACTTTT GTGGATA-GA
>NC_004327 TATAAAAACC CAT----ATG AAAATACTTT TAA---CTTA TCATATGCTA

```

```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      2260      2270      2280      2290      2300
TCDM_00289 GCCGTTTTCT -TAAAGGTAG TTGTTTTATC GTGGGGATAA AAGAGAAAGT
tcruCLBren GCCGTTTTCT -TAAAGGTAG TTGTTTTATC GTGGGGATAA AAGAGAAAGT
tcruCLBren GCCGTTTTCT -TAAAGGTAG TTGTTTTATC GTAGGGATAA AAGAGAAAGT
tcruSylvio GCCGTTTTCT -TAAAGGTAG TTGTTTTATC GTGGGGATAA AAGAGAAAGT
tbbruTREU92 CATGTTTTTT ----GTCGT CAATTTCCCT ATAGGTGTTA AGGCAAAGGA
lmajFriedl ACGGTGCACT GCCGCTCTCG CGCATTCTTC TTTGGCGTGC GCAACGGTAA
tvivY486 AACACATTCG GCGGGGTTGT GT-TTTTCTT ATAGGTGTGC GGGAGCGTAA
tconIL3000 AAGGTATGTT GCTCGAGCAT TTATTTTCTT ATAGGTGTTA GGTCAAAGAA
>NC_004327 TTCGTGATTT --TATTGACA TTGCCTTGAA TTGGCCTCAA AAAGTAAAT

```



```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2310      2320      2330      2340      2350
TCDM_00289 TGGTAGGCGC TTTCCACAGA AGAATCTGTT TTTTCGAGGC GTTTGA----
tcruCLBren TGGTAGGCGT TTTCCACAGA AGAATATGTT TTT-TGAGGC GTTTGTATGAG
tcruCLBren TGGTAGGCGC TTTCCACAGA AGAATATGTT TTT-TGAGGC GTTTGTATGAG
tcruSylvio TGGTAGGCGC TTTCCACAGA AGAATCTGTT TTT-CGAGGC GTTTGTATGAG
tbruTREU92 GGGGGGTCCC ACCCCACAGC CGTCTCTGTT TAC-TGAGGC GTGTGGCCGAG
lmajFriedl CAAC-GATGT GGCGCACAGC GATGTGGAA- ----GCCGC ATTCAGAGAG
tvivY486 AGATAGGCAG AAGCTCCACC AGTCCATGTT CTC-CTCTGC TCTTGTATGAG
tconIL3000 TAGTGGGTCC TCTCTCCGCC GATGTATGTT TGA-TGAAGC CACCGATGAG
>NC_004327 ATCCAAACGC ATTTAAAATT AATATCATGT ATCAAAAAAG ATCCCAAGTT

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2360      2370      2380      2390      2400
TCDM_00289 -----
tcruCLBren CTAAGGTATG CTGTGTTGG- AGGAGTGCAA TGCGACGAAG AGTGTGCGGG
tcruCLBren CTAAGGTATA CTGTGTTGG- AGGAGTGCAA TGCGACGAAG AGTGTGCGGG
tcruSylvio CTAAGGTATA CTGTGTTGG- AGGAGTGCAA TGCGACGAAG AGTTTGCGGG
tbruTREU92 TTTCAACCACG CTGTGAAGG- AGGGATGTAA CAATGACTCC ATTCCGTGTT
lmajFriedl TTTGAGTTTG CGGTGACGG- AGGGC----A CAACGGGAAA GAATC-----
tvivY486 TTTAGGTACG CGGTGCAGG- ACGGCTGCAG CAGCACAAAG GGCTCGCGAG
tconIL3000 TTTATTTATG CGCTGCAGG- ATGGCTGTAG TACTGAGAAC CAGTCCCCT
>NC_004327 CTGGAATTCT TAGACATAGT AACAAAAGAA GAAAACCAAA AAAATAAACA

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2410      2420      2430      2440      2450
TCDM_00289 -----
tcruCLBren GATTTGAGCG TGACGAGCGC ACCATGC-AC GAGCCATGGT TTGCGCTAGT
tcruCLBren GATTTGAGCG TGACGAGCGC ACCATGC-AC GAGCCATGGT TTGCGCTAGT
tcruSylvio GATTTGAGCG TGACGAGCGC ACCATGC-AC GAGCCATGGT TTGCGTTAGT
tbruTREU92 CATTTCAGCG GAATAGTCGT GCAATGC-AT TGGCCCCGTG TCAAATTTTT
lmajFriedl CGTTTGAGTG GAATAGGTCA GTCATGC-TA GGGCCACGGC TGTCTTTTTT
tvivY486 TCTTCCAGCG TGACGAGCGC AGCATGC-AC GGCCCGAGAC TGACACTCAT
tconIL3000 CGTTTCACCG AGACGAACGT AAGATGC-AC TGGCCACAAG TGAAGCTGAT
>NC_004327 AGTCATGGAA AATAATGACA AAGATGAGAC TAATCATAAT CAAAATGGTT

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2460      2470      2480      2490      2500
TCDM_00289 -----
tcruCLBren GTC--CGCTG CAGATCTTCT TCCTATTCTA AA----GGCG TAA-----
tcruCLBren GTC--CGCTG CAGATCTTCT TCCTATTCTA AA----GGCG TAA-----
tcruSylvio GTC--CGCTG CAGATCTTCT TCCTATTCTA AA----GACG TAA-----
tbruTREU92 AGAGGTGCAT CAGGTGCTGC ACCTGTTACA AGACCTGAAG TAAAGCTGCG
lmajFriedl CTC--TATCT ACGATCCGCT GACAGCCGAC TCGCCATGTC AGGCAC'TTTA
tvivY486 TGA--CACAT TAGAGGCGCT CACCTTCTC AA----CAAG TAGC--TACC
tconIL3000 TGG--CACAT CAGAGATCTC TTCCTTTGTT GA----CGGT TTGTGATACG
>NC_004327 ATAGAAATAA CCATAAAAT AAGCAATTTG ATGAATACGA CGACCACTTG

```

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
      2510      2520      2530      2540      2550
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 GAGCCGTTAA GTCGACGGGG AACGTAC--- -----
lmajFriedl CAGCGCCTGC GCCGAGATGA TGGGGTCCCTC GGTGTGAAGA GG--CCGCGA
tvivY486 TACTTCTCGT GTCAGCTCGG GCTTCCCTGTA CACTCTCCGC CT--TTATAA
tconIL3000 GGGGTGCTGT GCAACCGCGG TTCTGTTGAC AGCCTTCCAC ACAGTAATGG
>NC_004327 TTTAATTGA- -----

```

```

.....|.....|.....|.....|.....|.....|
          2560      2570      2580      2590      2600
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl CGGTGGTGTG CGCGCTTGTC AGCGCGCATG AGTGGCTCTT CTTTTCGGTG
tvivY486   A-----TCTT CTTTGTGCGA GGTGA-----
tconIL3000 AAGCGATACT CCTACTACTC CGCAGGCGCG ATGAATACAT CTCTTCCCGC
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2610      2620      2630      2640      2650
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl CCGCTTCCAG CTCGTGTGCT GCATGGAGGA ACACCGTTTG TGATCATGGT
tvivY486   -----
tconIL3000 CTGCCTTTAT ATGTGCGGGT GCCCGTGTGA TAACCCGTCA ACACACCATA
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2660      2670      2680      2690      2700
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl TGGGCCCCAG GCGCACGCTT CCTGTTTTTC GTCCTGCGCG TTCCCGCCTG
tvivY486   -----
tconIL3000 TGCTACACTG CAGCTCCCTT TGTTTACCTT GATGGAATGA AATCAATTCT
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2710      2720      2730      2740      2750
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl CCTTTCCAGG CCCTCGTCCC TCCTCCACAC GCCACCGTGA CAGTCCTGCA
tvivY486   -----
tconIL3000 AAA-----
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2760      2770      2780      2790      2800
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl GACACCCTCC AACGTAGCCG AGGGAACAGC AGAGGACGAA CATGAACCAA
tvivY486   -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2810      2820      2830      2840      2850
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl ACAGAAAAAA AAACGGAGGG TCTTAGCGAA AGTATTCAGC TCAGGCGCAC
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2860      2870      2880      2890      2900
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl CTCTCACGTC ATGCAGCTGG CCCTGCTGCC TATTCGGGGC TGTGGGTGGA
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2910      2920      2930      2940      2950
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl GCTGCGCAGC ATGGTATAGC AGGTGCCACC TGACGACCAC TCGAAGAAGA
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          2960      2970      2980      2990      3000
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl GGAAGAGAAA CGCGATCAAA CTCTCTCTCC GTGCCGTTGA ACCCTCTACG
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

```

.....|.....|.....|.....|.....|.....|
          3010      3020      3030      3040
TCDM_00289 -----
tcruCLBren -----
tcruCLBren -----
tcruSylvio -----
tbruTREU92 -----
lmajFriedl GCCTCACGGG ATGGCTGTTT CCATCCCTGA AGCGCGCACT CTCCCGCTA
tvivY486 -----
tconIL3000 -----
>NC_004327 -----

```

8. Conclusiones y Perspectivas

Existe un número limitado de estudios genómicos enfocados en el análisis de la compartimentación núcleo-citoplasma a escala ómica. En particular, nuestro trabajo avanza el conocimiento de este fenómeno en un patógeno de importancia mundial, el kinetoplástido *Trypanosoma cruzi*, donde no existían reportes previos a escala genómica que estudiaran la compartimentación núcleo-citoplasma de ARN. Nuestras observaciones indican que existe una distribución diferencial de ARNm donde se destacan que ARNm relacionados a la etapa infectiva del estadio se acumulan preferentemente en el nucleoplasma en un estadio no infectivo. Tomadas en conjunto nuestras investigaciones sientan bases para nuevos experimentos, a continuación, se detallan las conclusiones y perspectivas del presente trabajo.

Desde un punto de vista técnico, se pudo establecer un protocolo de fraccionamiento que permite el aislamiento de núcleos y citoplasma de epimastigotas de *T. cruzi* en cantidad y calidad adecuada (controlada mediante marcadores proteicos y nucleicos) para estudios transcriptómicos. Si bien, las fracciones obtenidas representan la mayor parte del contenido celular, no se puede descartar que la representación de cierta proporción de los ARNs asociados al núcleo, ya sea a nivel de envoltura externa o a nivel de kinetoplasto esté disminuida. Esto, podría afectar los niveles de algunos de los transcritos del citoplasma.

Respecto al diseño experimental, las muestras de ARN obtenidas a partir de cuadruplicados de las fracciones fueron enviadas a secuenciar en tres *pools*. Esta estrategia presenta la ventaja de aumentar la cantidad de ARN obtenido a partir de fracciones de bajo rendimiento y disminuye significativamente los costos. Una de las mayores críticas a esta estrategia es que, si una de las réplicas sufre algún desvío, los resultados resultan afectados. Por esto, en principio, es importante controlar adecuadamente la calidad de las réplicas, por ejemplo, es útil para comprobar que las muestras se comportan como réplicas biológicas, y que ninguna de las réplicas secuenciadas (mezcladas en el *pool*) sufre un desvío que pudiera afectar nuestros resultados. Para esto, realizamos un análisis de componentes principales y correlaciones de los resultados de qRT-PCR de genes seleccionados y pudimos demostrar que las muestras de las diferentes fracciones se agrupan entre ellas y, por lo tanto, es altamente

probable que no existiesen sesgos mayores en nuestros resultados que produzcan falsos positivos.

Otra de las desventajas de esta estrategia es que disminuye la sensibilidad del estudio, por lo que en nuestro análisis no es posible determinar, a nivel global, si cambios pequeños en los niveles de los transcriptos son significativos no pudiendo reportar cambios de genes de baja expresión. De todas formas, esta estrategia ha sido utilizada previamente por varios grupos (Allison *et al.*, 2006; Walrad *et al.*, 2012; Choi *et al.*, 2014; Araújo *et al.*, 2016) entre otros. Respecto a la calidad de los resultados de secuenciación de los transcriptomas particulares de cada fracción, se detectaron cerca de 9.000 de los 10.600 genes anotados en la cepa Esmeraldo *like*, para las tres fracciones subcelulares estudiadas. El transcriptoma citoplasmático correlaciona significativamente con el transcriptoma total. Este hallazgo confirma los resultados obtenidos por Smircich *et al.* (Smircich *et al.*, 2015) que llevaron a proponer que en *T. cruzi*, tanto la transcripción como la maduración actúan de forma generalizada. Aunque la mayoría de los transcriptos son compartidos entre núcleo y citoplasma, existe un grupo de aproximadamente 1200 transcriptos que presenta diferencias importantes en sus niveles relativos. Los genes cuyos transcriptos presentan mayor abundancia nuclear (escasez citoplasmática) están enriquecidos en proteínas de familias multigénicas, tales como las mucinas (Di Noia *et al.*, 1998; Buscaglia *et al.*, 2006), las trans-sialidasas (Schenkman *et al.*, 1994; Nardy *et al.*, 2016), las MASP (Bartholomeu *et al.*, 2009; dos Santos *et al.*, 2012), las DGF (Kawashita *et al.*, 2009; De Pablos y Osuna, 2012) y la familia gp63 (Cuevas *et al.*, 2003; Yao, 2010) que se expresan mayormente durante los estadios infectivos de *T. cruzi* y se asocian a la protección y evasión del sistema inmune (dos Santos *et al.*, 2012; De Pablos *et al.*, 2016) y las RHS, que se expresan en respuesta a ciertos tipos de estrés (Grynberg *et al.*, 2012). Estos transcriptos, son más largos que el promedio de los genes, presentan características estructurales y un mayor contenido GC. A su vez, los transcriptos nucleares son pobremente traducidos en epimastigotas. Por otro lado, el agrupamiento funcional de los genes cuyos transcritos están enriquecidos en el citoplasma se relaciona con procesos metabólicos y con traducción proteica. Además, estos transcriptos son más cortos que el promedio y son altamente traducidos. La distribución diferencial de varios

genes seleccionados fue verificada por qRT-PCR y por FISH se demostró también su relocalización durante el estadio metacíclico.

Estos resultados indican que existe una distribución subcelular diferencial de ARNs que codifican para genes de expresión estadio específicos en *T. cruzi*.

La compartimentación sub-celular de mensajeros podría estar asociada con señales en *cis* en los transcriptos. En este sentido, encontramos que los genes enriquecidos en CET presentan 27 señales de secuencia distintivas. Estos motivos, serían reconocidos por proteínas de unión al ARN para las que pudimos definir 6 ortólogos. Entre ellas, se destaca la PABP1 (da Costa Lima *et al.*, 2010) podría ser responsable de la distribución de abundancia diferencial entre núcleo y citoplasma en epimastigotas de *T. cruzi*. Existen también otras proteínas candidatas, que podrían estar involucradas en la regulación generalizada de este fenómeno, actuando junto con la PABP1 y algunas proteínas que podrían estar involucradas en la regulación de sub-grupos específicos de mensajeros. Por otro lado, los genes enriquecidos en NET presentan 45 señales de secuencia distintivas. Estos motivos podrían ser reconocidos por proteínas de unión al ARN para las cuales se pudo definir 11 ortólogos en *T. cruzi*. Entre ellas, las proteínas TcUBP 1 y 2 tienen un rol en la regulación global de los mensajeros (Di Noia *et al.*, 2000; D'Orso y Frasch, 2001b). Además, la proteína TcCLB.511621.50 (TcRRM1) podría estar involucrada en la retención nuclear de los mensajeros NET en epimastigotas y la proteína de unión al ARN, TcCLB.509581.50 de su expresión en el estadio metacíclico.

La comparación con el transcriptoma poliadenilado de epimastigotas de *T. cruzi* (datos publicados por Smircich *et al.* (Smircich *et al.*, 2015)), nos permitió definir grupos de genes sujetos a una fuerte regulación en pasos particulares de la expresión génica, tales como la regulación del nivel de estado estacionario de mensajeros, la maduración nuclear de ARN o su retención nuclear. Los mensajeros que codifican para proteínas involucradas en la traducción están enriquecidos en el transcriptoma procesado, mientras que mensajeros que codifican para la vía de señalización del AMPc (vía requerida para la infección (Hamedi *et al.*, 2015)) están pobremente representados cuando se lo compara con el transcriptoma nuclear. Además, genes que codifican para proteínas involucradas en patogenicidad están disminuidos en el transcriptoma

procesado. Estos resultados sugieren que en epimastigotas de *T. cruzi* las proteínas altamente expresadas presentan varios niveles de regulación que garantizan su traducción, un primer paso, sería su procesamiento y luego su acumulación en el citoplasma, y presentarían también una fuerte regulación a nivel traduccional (Smircich *et al.*, 2015). Contrariamente, los mensajeros que codifican para proteínas requeridas en otros estadios son regulados también por varios mecanismos, son menos procesados, presentan acumulación nuclear y son menos traducidos que el promedio de los mensajeros.

Por último, determinamos que el proceso de *cis-splicing* sucede en epimastigotas de la cepa Dm28c de *T. cruzi* para el transcripto del gen que codifica para la poliA polimerasa en el núcleo, pero se encuentra transcripto sin procesar también en el citoplasma.

En conjunto, nuestros datos muestran que, en el contexto de una transcripción y procesamiento generalizado, el compartimento nuclear juega un rol en el control de los mecanismos de regulación de la expresión génica en este parásito.

Como perspectivas de este trabajo, planteamos verificar si las proteínas identificadas por TOM-TOM, efectivamente tienen un rol en la compartimentación de mensajeros. Las proteínas serán expresadas en un sistema heterólogo y purificadas para realizar análisis *in vitro* de interacción con los motivos para estudiar la afinidad y especificidad del reconocimiento. Proponemos también, estudiar su función mediante la sobreexpresión en el parásito, de las proteínas fusionadas a un *tag*. Analizaremos el fenotipo que produce (cambios en el crecimiento, en la capacidad de infección *in vitro* e *in vivo*) y estudiaremos las proteínas y ARNs que se asocian a ellas en los parásitos sobreexpresantes o en parásitos que expresen la proteína etiquetada mediante CRISPR-Cas. A su vez, el *knock-out* del gen mediante técnicas clásicas o CRISPR-Cas también permitiría aproximarnos a su función.

Por otro lado, analizaremos la función de las señales *in vivo*, para esto, seleccionaremos algunas señales presentes en las regiones no traducidas de mensajeros diferencialmente localizados, tanto en núcleo como en citoplasma y las mutaremos mediante la técnica CRISPR-Cas. Una vez obtenidos los mutantes, analizaremos si la localización de los mensajeros cambia respecto al *wild type* durante el ciclo de vida del parásito.

Alternativamente, generaremos vectores de expresión en *T. cruzi* y las señales serán clonadas en las regiones no traducidas de un gen reportero, luego transfectaremos parásitos con las construcciones y analizaremos la localización del ARN del gen reportero durante el ciclo de vida de *T. cruzi* comparando los resultados con un control sin las señales.

Respecto al análisis comparativo de los transcriptomas procesado y sin procesar, verificaremos nuestras inferencias acerca del enriquecimiento de ARNs en el transcriptoma procesado de *T. cruzi* mediante qRT-PCR de genes seleccionados en las muestras correspondientes a los diferentes transcriptomas analizados. Para evitar un sesgo en los resultados, debido a los diferentes métodos utilizados, se podrían realizar nuevamente los transcriptomas utilizando el secuenciador IonTorrent para validar los resultados.

Por último, como perspectivas del análisis del procesamiento de ARN mediante *cis-splicing* en *T. cruzi*, nos proponemos realizar un nuevo ensamblado de los datos para la identificación de secuencias procesadas y de este modo, distinguir las diferentes especies de mensajeros en las fracciones. Además, analizaremos la presencia de huellas ribosomales en el intrón de la PAP (utilizando los datos publicados por Smircich *et al.*, 2015) como forma de inferir si el ARN no procesado se asocia a ribosomas, lo que podría sugerir un mecanismo de regulación de la expresión del gen. Los datos de transcriptomas de las fracciones nuclear y citoplásmico serán también utilizados para identificar genes no descritos que puedan sufrir *cis-splicing*. Para esto proponemos la búsqueda de *reads* pareados que alinean a diferentes genes, así como *reads* híbridos, cuyos extremos 3' y 5' alinean con genes diferentes y que además coincidan con la presencia de la secuencia complementaria a U1 en el genoma.

Desde nuestro punto de vista, el trabajo aquí presentado es un aporte novedoso hacia el entendimiento de la biología molecular de *T. cruzi*. Este parásito, que no regula el inicio de la transcripción, posee mecanismos de regulación de la expresión génica mediante los cuales, los mensajeros que no son requeridos en el estadio epimastigota no serían procesados eficientemente y a su vez se acumularían en el núcleo como forma de impedir

su traducción. A su vez, los mensajeros cuyos productos son altamente requeridos, serían procesados y acumulados en el citoplasma, probablemente debido a un aumento de su vida media, garantizándose así su traducción. El procesamiento de los ARNs en este organismo parece bastante primitivo, solapándose los procesos de *cis* y *trans-splicing* y exportándose al citoplasma ARNs no procesados. Aún no se ha encontrado una función clara a este fenómeno, reportado previamente por (Jager *et al.*, 2007), tal vez, sea una forma de regular la expresión proteica, impidiendo la traducción de estos ARNs en el citoplasma y en el momento adecuado, procesándolos para volverlos funcionales. Una segunda opción es que este organismo no posea elementos adecuados que regulen el transporte al citoplasma y por lo tanto exista una fuga de ARNs inmaduros desde el núcleo. Estas suposiciones abren muchas puertas para seguir estudiando estos fenómenos en el futuro.

9. Bibliografía

- Agabian, N. (1990) 'Trans splicing of nuclear pre-mRNAs', *Cell*, 1990/06/29, 61(7), pp. 1157–1160. doi: 0092-8674(90)90674-4 [pii].
- Akef, A., Lee, E. S. and Palazzo, A. F. (no date) 'Splicing promotes the nuclear export of β -globin mRNA by overcoming nuclear retention elements'. doi: 10.1261/rna.051987.115.
- Allison, D. B., Cui, X., Page, G. P. and Sabripour, M. (2006) 'Microarray data analysis: from disarray to consolidation and consensus', *Nature Reviews Genetics*, 7(1), pp. 55–65. doi: 10.1038/nrg1749.
- Alvarez, F., Robello, C. and Vignali, M. (1994) 'Evolution of codon usage and base contents in kinetoplastid protozoans.', *Molecular biology and evolution*, 11(5), pp. 790–802. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7968492> (Accessed: 9 March 2017).
- Alves, L. R., Oliveira, C. and Goldenberg, S. (2015) 'Eukaryotic translation elongation factor-1 alpha is associated with a specific subset of mRNAs in *Trypanosoma cruzi*', *BMC Microbiol*, 15, p. 104. doi: 10.1186/s12866-015-0436-2.
- Alves, L. R., Oliveira, C., Morking, P. A., Kessler, R. L., Martins, S. T., Romagnoli, B. A., Marchini, F. K. and Goldenberg, S. (2014) 'The mRNAs associated to a zinc finger protein from *Trypanosoma cruzi* shift during stress conditions', *RNA Biol*, 11(7), pp. 921–933. doi: 10.4161/rna.29622.
- Antwi, E. B., Haanstra, J. R., Ramasamy, G., Jensen, B., Droll, D., Rojas, F., Minia, I., Terraio, M., Mercé, C., Matthews, K., Myler, P. J., Parsons, M., Clayton, C., Merce, C., Matthews, K., Myler, P. J., Parsons, M. and Clayton, C. (2016) 'Integrative analysis of the *Trypanosoma brucei* gene expression cascade predicts differential regulation of mRNA processing and unusual control of ribosomal protein expression', *BMC Genomics*, 17(1), p. 306. doi: 10.1186/s12864-016-2624-3.
- Araújo, F. S., Coelho, L. M., Silva, L. do C., da Silva Neto, B. R., Parente-Rocha, J. A., Bailão, A. M., de Oliveira, C. M. A., Fernandes, G. da R., Hernández, O., Ochoa, J. G. M., Soares, C. M. de A., Pereira, M. and Pereira, M. (2016) 'Effects of Argentilactone on the Transcriptional Profile, Cell Wall and Oxidative Stress of *Paracoccidioides* spp.', *PLOS Neglected Tropical Diseases*. Edited by J. M. Vinetz. Public Library of Science, 10(1), p. e0004309. doi: 10.1371/journal.pntd.0004309.
- Araújo, P. R., Teixeira, S. M., Araujo, P. R. and Teixeira, S. M. (2011) 'Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review', *Mem Inst Oswaldo Cruz*. 2011/06/10, 106(3), pp. 257–266. doi: S0074-02762011000300002 [pii].
- Archer, S. K., Luu, V. D., de Queiroz, R. A., Brems, S. and Clayton, C. (2009) 'Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle', *PLoS Pathog*. 2009/08/29, 5(8), p. e1000565. doi: 10.1371/journal.ppat.1000565.
- Arner, E., Kindlund, E., Nilsson, D., Farzana, F., Ferella, M., Tammi, M. T. and Andersson, B. (2007) 'Database of *Trypanosoma cruzi* repeated genes: 20,000 additional gene variants', *BMC Genomics*, 8, p. 391. doi: 10.1186/1471-2164-8-391.
- Atayde, V. D., Tschudi, C. and Ullu, E. (2011) 'The emerging world of small silencing RNAs in protozoan parasites.', *Trends in parasitology*, 27(7), pp. 321–7. doi: 10.1016/j.pt.2011.03.002.
- Atwood, J. A., Weatherly, D. B., Minning, T. A., Bundy, B., Cavola, C., Opperdoes, F. R., Orlando, R. and Tarleton, R. L. (2005) 'The *Trypanosoma cruzi* proteome.', *Science (New York, N.Y.)*, 309(5733), pp. 473–6. doi: 10.1126/science.1110289.
- Avila, A. R., Yamada-Ogatta, S. F., da Silva Monteiro, V., Krieger, M. A., Nakamura, C. V., de Souza, W. and Goldenberg, S. (2001) 'Cloning and characterization of the metacyclogenin gene, which is specifically expressed during *Trypanosoma cruzi* metacyclogenesis', *Mol Biochem Parasitol*,

117(2), pp. 169–177. Available at:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11606227.

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., Itzkovitz, S., Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I. and Itzkovitz, S. (2015) 'Nuclear Retention of mRNA in Mammalian Tissues', *Cell Rep*, 13(12), pp. 2653–2662. doi: 10.1016/j.celrep.2015.11.036.

Bailey, T. L. (2011) 'DREME: motif discovery in transcription factor ChIP-seq data.', *Bioinformatics (Oxford, England)*. Oxford University Press, 27(12), pp. 1653–9. doi: 10.1093/bioinformatics/btr261.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. and Noble, W. S. (2009) 'MEME SUITE: tools for motif discovery and searching.', *Nucleic acids research*, 37(Web Server issue), pp. W202–8. doi: 10.1093/nar/gkp335.

Balagopal, V. and Parker, R. (2009) 'Stm1 modulates mRNA decay and Dhh1 function in *Saccharomyces cerevisiae*.' *Genetics*, 181(1), pp. 93–103. doi: 10.1534/genetics.108.092601.

Bandziulis, R. J., Swanson, M. S. and Dreyfuss, G. (1989) 'RNA-binding proteins as developmental regulators', *Genes Dev*, 3(4), pp. 431–437. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2470643.

Bartholomeu, D. C., Cerqueira, G. C., Leao, A. C., daRocha, W. D., Pais, F. S., Macedo, C., Djikeng, A., Teixeira, S. M. and El-Sayed, N. M. (2009) 'Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*', *Nucleic Acids Res.* 2009/04/02, 37(10), pp. 3407–3417. doi: gkp172 [pii]10.1093/nar/gkp172.

Bayer-Santos, E., Gentil, L. G., Cordero, E. M., Correa, P. R., da Silveira, J. F., Corrêa, P. R. C. and da Silveira, J. F. (2012) 'Regulatory elements in the 3' untranslated region of the GP82 glycoprotein are responsible for its stage-specific expression in *Trypanosoma cruzi* metacyclic trypomastigotes', *Acta Trop*, 123(3), pp. 230–233. doi: 10.1016/j.actatropica.2012.03.014.

Bern, C., Kjos, S., Yabsley, M. J. and Montgomery, S. P. (2011) 'Trypanosoma cruzi and Chagas' Disease in the United States', *Clin Microbiol Rev*, 24(4), pp. 655–681. doi: 10.1128/CMR.00005-11.

Berná, L., Chiribao, M. L., Greif, G., Rodriguez, M., Alvarez-Valin, F. and Robello, C. (2017) 'Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*'. PeerJ Inc. doi: 10.7287/PEERJ.PREPRINTS.2782V1.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B.

- G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E. and El-Sayed, N. M. (2005) 'The genome of the African trypanosome *Trypanosoma brucei*', *Science*, 2005/07/16, 309(5733), pp. 416–422. doi: 309/5733/416 [pii]10.1126/science.1112642.
- Blackwell, J. M. and Melville, S. E. (1999) 'Status of protozoan genome analysis: trypanosomatids', *Parasitology*, 118 Suppl, pp. S11–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10466133>.
- Boiani, M., Piacenza, L., Hernández, P., Boiani, L., Cerecetto, H., González, M. and Denicola, A. (2010) 'Mode of action of Nifurtimox and N-oxide-containing heterocycles against *Trypanosoma cruzi*: Is oxidative stress involved?', *Biochemical Pharmacology*, 79(12), pp. 1736–1745. doi: 10.1016/j.bcp.2010.02.009.
- Boucher, N., Wu, Y., Dumas, C., Dube, M., Sereno, D., Breton, M. and Papadopoulou, B. (2002) 'A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element', *J Biol Chem*, 277(22), pp. 19511–19520. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11912202.
- Boulon, S., Westman, B. J., Hutten, S., Boisvert, F.-M. and Lamond, A. I. (2010) 'The nucleolus under stress.', *Molecular cell*. Elsevier, 40(2), pp. 216–27. doi: 10.1016/j.molcel.2010.09.024.
- Branche, C., Ochaya, S., Aslund, L. and Andersson, B. (2006) 'Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi*', *Mol Biochem Parasitol*, 147(1), pp. 30–38. doi: 10.1016/j.molbiopara.2006.01.005.
- Brandão, A. and Jiang, T. (2009) 'The composition of untranslated regions in *Trypanosoma cruzi* genes.', *Parasitology international*. Elsevier Ireland Ltd, 58(3), pp. 215–9. doi: 10.1016/j.parint.2009.06.001.
- Bringaud, F., Biteau, N., Melville, S. E., Hez, S., El-Sayed, N. M., Leech, V., Berriman, M., Hall, N., Donelson, J. E. and Baltz, T. (2002) 'A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*', *Eukaryot Cell*, 1(1), pp. 137–151. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12455980>.
- Buscaglia, C. A., Campo, V. A., Frasch, A. C. and Di Noia, J. M. (2006) 'Trypanosoma cruzi surface mucins: host-dependent coat diversity', *Nat Rev Microbiol*, 4(3), pp. 229–236. doi: 10.1038/nrmicro1351.
- Buscaglia, C. A., Campo, V. A., Di Noia, J. M., Torrecilhas, A. C. T., De Marchi, C. R., Ferguson, M. A. J., Frasch, A. C. C. and Almeida, I. C. (2004) 'The Surface Coat of the Mammal-dwelling Infective Trypomastigote Stage of *Trypanosoma cruzi* Is Formed by Highly Diverse Immunogenic Mucins', *Journal of Biological Chemistry*, 279(16), pp. 15860–15869. doi: 10.1074/jbc.M314051200.
- Camara Mde, L., Bouvier, L. A., Canepa, G. E., Miranda, M. R., Pereira, C. A., Cámara, M. de los M., Bouvier, L. A., Canepa, G. E., Miranda, M. R. and Pereira, C. A. (2013) 'Molecular and functional characterization of a *Trypanosoma cruzi* nuclear adenylate kinase isoform', *PLoS Negl Trop Dis*. Edited by C. A. Buscaglia, 7(2), p. e2044. doi: 10.1371/journal.pntd.0002044.
- Campo, V., Di Noia, J. M., Buscaglia, C. A., Aguero, F., Sanchez, D. O. and Frasch, A. C. (2004) 'Differential accumulation of mutations localized in particular domains of the mucin genes expressed in the vertebrate host stage of *Trypanosoma cruzi*', *Mol Biochem Parasitol*, 133(1), pp. 81–91. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14668015.

- Campos, P. C., Bartholomeu, D. C., DaRocha, W. D., Cerqueira, G. C. and Teixeira, S. M. R. (2008) 'Sequences involved in mRNA processing in *Trypanosoma cruzi*', *Int J Parasitol.* 2008/08/14, 38(12), pp. 1383–1389. doi: S0020-7519(08)00232-4 [pii]10.1016/j.ijpara.2008.07.001.
- Cantey, P. T., Stramer, S. L., Townsend, R. L., Kamel, H., Ofafa, K., Todd, C. W., Currier, M., Hand, S., Varnado, W., Dotson, E., Hall, C., Jett, P. L. and Montgomery, S. P. (2012) 'The United States *Trypanosoma cruzi* Infection Study: evidence for vector-borne transmission of the parasite that causes Chagas disease among United States blood donors', *Transfusion*, 52(9), pp. 1922–1930. doi: 10.1111/j.1537-2995.2012.03581.x.
- Cassola, A. (2011) 'RNA Granules Living a Post-transcriptional Life: the Trypanosomes' Case', *Curr Chem Biol*, 5(2), pp. 108–117. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21949551>.
- Cassola, A. and Frasch, A. C. (2009) 'An RNA recognition motif mediates the nucleocytoplasmic transport of a trypanosome RNA-binding protein', *J Biol Chem.* 2009/10/06, 284(50), pp. 35015–35028. doi: M109.031633 [pii]10.1074/jbc.M109.031633.
- Cassola, A., De Gaudenzi, J. G. and Frasch, A. C. (2007) 'Recruitment of mRNAs to cytoplasmic ribonucleoprotein granules in trypanosomes', *Mol Microbiol.* 2007/07/20, 65(3), pp. 655–670. doi: MMI5833 [pii]10.1111/j.1365-2958.2007.05833.x.
- Cavalier-Smith, T. (2010) 'Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree', *Biology Letters*, 6(3), pp. 342–345. doi: 10.1098/rsbl.2009.0948.
- Cazzulo, J. J. (1994) 'Intermediate metabolism in *Trypanosoma cruzi*', *J Bioenerg Biomembr*, 26(2), pp. 157–165. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8056782>.
- Chagas, C. (1909) 'Nova tripanozomíaze humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi*, agente etiológico da nova entidade mórbida do homem.', *Mem. Inst. Oswaldo Cruz*, 1, pp. 159–219.
- Chávez, S., Beilharz, T., Rondón, A. G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J. Q., Lithgow, T. and Aguilera, A. (2000) 'A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*.', *The EMBO journal.* European Molecular Biology Organization, 19(21), pp. 5824–34. doi: 10.1093/emboj/19.21.5824.
- Choi, Y.-J., Aliota, M. T., Mayhew, G. F., Erickson, S. M. and Christensen, B. M. (2014) 'Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions.', *PLoS neglected tropical diseases.* Public Library of Science, 8(5), p. e2905. doi: 10.1371/journal.pntd.0002905.
- Clayton, C. (2013) 'The regulation of trypanosome gene expression by RNA-binding proteins', *PLoS Pathog*, 9(11), p. e1003680. doi: 10.1371/journal.ppat.1003680.
- Clayton, C. E. (2002) 'Life without transcriptional control? From fly to man and back again', *EMBO J*, 21(8), pp. 1881–1888. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11953307.
- Coller, J. M., Tucker, M., Sheth, U., Valencia-Sanchez, M. A. and Parker, R. (2001) 'The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes', *RNA.* 2002/01/10, 7(12), pp. 1717–1727. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11780629.
- Coller, J. and Parker, R. (2005) 'General translational repression by activators of mRNA

- decapping', *Cell*, 2005/09/24, 122(6), pp. 875–886. doi: S0092-8674(05)00706-3 [pii]10.1016/j.cell.2005.07.012.
- Coller, J. and Wickens, M. (2007) 'Tethered function assays: an adaptable approach to study RNA regulatory proteins.', *Methods in enzymology*, 429(7), pp. 299–321. doi: 10.1016/S0076-6879(07)29014-7.
- Conesa, A. and Gotz, S. (2008) 'Blast2GO: A comprehensive suite for functional analysis in plant genomics', *Int J Plant Genomics*, 2008, p. 619832. doi: 10.1155/2008/619832.
- da Costa Lima, T. D., Moura, D. M. N., Reis, C. R. S., Vasconcelos, J. R. C., Ellis, L., Carrington, M., Figueiredo, R. C. B. Q. and de Melo Neto, O. P. (2010) 'Functional characterization of three Leishmania poly(A) binding protein homologues with distinct binding properties to RNA and protein partners', *Eukaryotic Cell*, 9(10), pp. 1484–1494. doi: 10.1128/EC.00148-10.
- Coughlin, B. C., Teixeira, S. M., Kirchhoff, L. V and Donelson, J. E. (2000) 'Amastin mRNA abundance in Trypanosoma cruzi is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein', *J Biol Chem*, 275(16), pp. 12051–12060. Available at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10766837.
- Cuevas, I. C., Cazzulo, J. J. and Sanchez, D. O. (2003) 'gp63 homologues in Trypanosoma cruzi: surface antigens with metalloprotease activity and a possible role in host cell infection', *Infect Immun*. 2003/09/23, 71(10), pp. 5739–5749. Available at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14500495.
- Cuevas, I. C., Frasch, A. C. and D'Orso, I. (2005) 'Insights into a CRM1-mediated RNA-nuclear export pathway in Trypanosoma cruzi', *Mol Biochem Parasitol*, 139(1), pp. 15–24. doi: 10.1016/j.molbiopara.2004.11.002.
- D'Orso, I. and Frasch, A. C. (2001a) 'Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in Trypanosoma cruzi by interaction with specific RNA-binding proteins', *J Biol Chem*, 276(19), pp. 15783–15793. Available at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11278796.
- D'Orso, I. and Frasch, A. C. (2001b) 'TcUBP-1, a developmentally regulated U-rich RNA-binding protein involved in selective mRNA destabilization in trypanosomes', *J Biol Chem*, 276(37), pp. 34801–34809. Available at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11435421.
- D'Orso, I. and Frasch, A. C. C. C. (2002) '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', *J Biol Chem*, 277(52), pp. 50520–50528. doi: 10.1074/jbc.M209092200.
- D'Orso, I., De Gaudenzi, J. G. and Frasch, A. C. (2003) 'RNA-binding proteins and mRNA turnover in trypanosomes', *Trends Parasitol*, 19(4), pp. 151–155. Available at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12689640.
- Dallagiovanna, B., Correa, A., Probst, C. M., Holetz, F., Smircich, P., de Aguiar, A. M., Mansur, F., da Silva, C. V, Mortara, R. A., Garat, B., Buck, G. A., Goldenberg, S. and Krieger, M. A. (2008) 'Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein

- from *Trypanosoma cruzi*', *J Biol Chem.* 2007/12/07, 283(13), pp. 8266–8273. doi: M703097200 [pii]10.1074/jbc.M703097200.
- Dallagiovanna, B., Perez, L., Sotelo-Silveira, J., Smircich, P., Duhagon, M. A. and Garat, B. (2005) 'Trypanosoma cruzi: Molecular characterization of TcPUF6, a Pumilio protein', *Exp Parasitol*, 109(4), pp. 260–264. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15755425.
- Das, A., Bellofatto, V., Rosenfeld, J., Carrington, M., Romero-Zaliz, R., del Val, C. and Estévez, A. M. (2015) *High throughput sequencing analysis of Trypanosoma brucei DRBD3/PTB1-bound mRNAs*, *Molecular and Biochemical Parasitology*. doi: 10.1016/j.molbiopara.2015.02.003.
- Dean, S., Sunter, J. D. and Wheeler, R. J. (2017) 'TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource.', *Trends in parasitology*. Elsevier, 33(2), pp. 80–82. doi: 10.1016/j.pt.2016.10.009.
- Dennis Jr., G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. and Lempicki, R. A. (2003) 'DAVID: Database for Annotation, Visualization, and Integrated Discovery', *Genome Biol*, 4(5), p. P3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12734009>.
- Díaz de Toranzo, E. G., Castro, J. A., Franke de Cazzulo, B. M. and Cazzulo, J. J. (1988) 'Interaction of benzimidazole reactive metabolites with nuclear and kinetoplastic DNA, proteins and lipids from *Trypanosoma cruzi*.' *Experientia*, 44(10), pp. 880–1. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3053234> (Accessed: 18 May 2017).
- Djikeng, A., Ferreira, L., D'Angelo, M., Dolezal, P., Lamb, T., Murta, S., Triggs, V., Ulbert, S., Villarino, A., Renzi, S., Ullu, E. and Tschudi, C. (2001) 'Characterization of a candidate *Trypanosoma brucei* U1 small nuclear RNA gene.', *Molecular and biochemical parasitology*, 113(1), pp. 109–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11254959> (Accessed: 2 April 2017).
- Docampo, R. and Moreno, S. N. (2011) 'Acidocalcisomes', *Cell Calcium*, 50(2), pp. 113–119. doi: 10.1016/j.ceca.2011.05.012.
- Docampo, R., de Souza, W., Miranda, K., Rohloff, P. and Moreno, S. N. (2005) 'Acidocalcisomes - conserved from bacteria to man', *Nat Rev Microbiol*, 3(3), pp. 251–261. doi: 10.1038/nrmicro1097.
- Drechsler, H. and McAinsh, A. D. (2012) 'Exotic mitotic mechanisms', *Open Biol*, 2(12), p. 120140. doi: 10.1098/rsob.120140.
- Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R. and Clayton, C. (2013) 'Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein', *PLoS Pathog*, 9(4), p. e1003286. doi: 10.1371/journal.ppat.1003286.
- Duhagon, M. A., Dallagiovanna, B., Ciganda, M., Ruyechan, W., Williams, N. and Garat, B. (2003) 'A novel type of single-stranded nucleic acid binding protein recognizing a highly frequent motif in the intergenic regions of *Trypanosoma cruzi*', *Biochem Biophys Res Commun*, 309(1), pp. 183–188. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12943680.
- Duhagon, M. A., Dallagiovanna, B. and Garat, B. (2001) 'Unusual features of poly[dT-dG].[dC-dA] stretches in CDS-flanking regions of *Trypanosoma cruzi* genome', *Biochem Biophys Res Commun*, 287(1), pp. 98–103.

Duhagon, M. A., Pastro, L., Sotelo-Silveira, J. R., Perez-Diaz, L., Maugeri, D., Nardelli, S. C., Schenkman, S., Williams, N., Dallagiovanna, B. and Garat, B. (2009) 'The Trypanosoma cruzi nucleic acid binding protein Tc38 presents changes in the intramitochondrial distribution during the cell cycle', *BMC Microbiol.* 2009/02/13, 9, p. 34. doi: 1471-2180-9-34 [pii]10.1186/1471-2180-9-34.

Duhagon, M. A., Smircich, P., Forteza, D., Naya, H., Williams, N. and Garat, B. (2011) 'Comparative genomic analysis of dinucleotide repeats in Trityps', *Gene.* 2011/08/10, 487(1), pp. 29–37. doi: S0378-1119(11)00358-1 [pii]10.1016/j.gene.2011.07.022.

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. and Andersson, B. (2005) 'The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease', *Science.* 2005/07/16, 309(5733), pp. 409–415. doi: 309/5733/409 [pii]10.1126/science.1112631.

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D. and Hall, N. (2005) 'Comparative genomics of trypanosomatid parasitic protozoa', *Science.* 2005/07/16, 309(5733), pp. 404–409. doi: 309/5733/404 [pii]10.1126/science.1112181.

Erben, E., Chakraborty, C. and Clayton, C. (2014) 'The CAF1-NOT complex of trypanosomes', *Frontiers in genetics*, 4, p. 299. doi: 10.3389/fgene.2013.00299.

Erben, E. D., Fadda, A., Lueong, S., Rg, J., Hoheisel, D. and Clayton, C. (2014) 'A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in Trypanosoma brucei', *PLoS Pathog*, 10(6). doi: 10.1371/journal.ppat.1004178.

Estevez, A. M. (2008) 'The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes', *Nucleic Acids Res.* 2008/07/10, 36(14), pp. 4573–4586. doi: gkn406 [pii]10.1093/nar/gkn406.

Fadda, A., Ryten, M., Droll, D., Rojas, F., Farber, V., Haanstra, J. R., Merce, C., Bakker, B. M., Matthews, K. and Clayton, C. (2014) 'Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels', *Mol Microbiol*, 94(2), pp. 307–326. doi: 10.1111/mmi.12764.

Fiebig, M., Gluenz, E., Carrington, M. and Kelly, S. (2014) 'SLaP mapper: a webserver for identifying and quantifying spliced-leader addition and polyadenylation site usage in kinetoplastid genomes', *Mol Biochem Parasitol*, 196(2), pp. 71–74. doi: 10.1016/j.molbiopara.2014.07.012.

Franzen, O., Arner, E., Ferella, M., Nilsson, D., Respuela, P., Carninci, P., Hayashizaki, Y., Aslund, L.,

- Andersson, B. and Daub, C. O. (2011) 'The short non-coding transcriptome of the protozoan parasite *Trypanosoma cruzi*', *PLoS Negl Trop Dis*, 5(8), p. e1283. doi: 10.1371/journal.pntd.0001283.
- Franzen, O., Ochaya, S., Sherwood, E., Lewis, M. D., Llewellyn, M. S., Miles, M. A. and Andersson, B. (2011) 'Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener', *PLoS Negl Trop Dis*, 5(3), p. e984. doi: 10.1371/journal.pntd.0000984.
- Franzén, O., Talavera-López, C., Ochaya, S., Butler, C. E., Messenger, L. A., Lewis, M. D., Llewellyn, M. S., Marinkelle, C. J., Tyler, K. M., Miles, M. A. and Andersson, B. (2012) 'Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*', *BMC Genomics*, 13(1), p. 531. doi: 10.1186/1471-2164-13-531.
- Gale, M., Carter, V. and Parsons, M. (1994) 'Cell cycle-specific induction of an 89 kDa serine/threonine protein kinase activity in *Trypanosoma brucei*.' *Journal of cell science*, pp. 1825–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7527048> (Accessed: 9 March 2017).
- García-Silva, M. R., Cura das Neves, R. F., Cabrera-Cabrera, F., Sanguinetti, J., Medeiros, L. C., Robello, C., Naya, H., Fernandez-Calero, T., Souto-Padron, T., de Souza, W. and Cayota, A. (2014) 'Extracellular vesicles shed by *Trypanosoma cruzi* are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells', *Parasitology Research*. Springer Berlin Heidelberg, 113(1), pp. 285–304. doi: 10.1007/s00436-013-3655-1.
- García-Silva, M. R., Frugier, M., Tosar, J. P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S. and Cayota, A. (2010) 'A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules', *Mol Biochem Parasitol*. 2010/02/17, 171(2), pp. 64–73. doi: S0166-6851(10)00035-6 [pii]10.1016/j.molbiopara.2010.02.003.
- De Gaudenzi, J., Frasch, A. C. and Clayton, C. (2005) 'RNA-binding domain proteins in Kinetoplastids: a comparative analysis', *Eukaryot Cell*. 2005/12/13, 4(12), pp. 2106–2114. doi: 4/12/2106 [pii]10.1128/EC.4.12.2106-2114.2005.
- De Gaudenzi, J., Frasch, A. C. and Clayton, C. (2005) 'RNA-Binding Domain Proteins in Kinetoplastids: a Comparative Analysis', *EUKARYOTIC CELL*, 4(12), pp. 2106–2114. doi: 10.1128/EC.4.12.2106–2114.2005.
- De Gaudenzi, J. G., Carmona, S. J., Agüero, F., Frasch, A. C., Agüero, F. and Frasch, A. C. (2013) 'Genome-wide analysis of 3'-untranslated regions supports the existence of post-transcriptional regulons controlling gene expression in trypanosomes', *PeerJ*. 2013/08/02. PeerJ, Inc, 1, p. e118. doi: 10.7717/peerj.118118 [pii].
- De Gaudenzi, J. G., Jager, A. V., Izcovich, R., Campo, V. A., Jäger, A. V., Izcovich, R., Campo, V. A., Jäger, A. V., Izcovich, R., Campo, V. A., Jäger, A. V., Izcovich, R. and Campo, V. A. (2016) '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*. Blackwell Publishing Inc., 63(4), pp. 440–452. doi: 10.1111/jeu.12288.
- De Gaudenzi, J. G., Noe, G., Campo, V. A., Frasch, A. C. and Cassola, A. (2011) 'Gene expression regulation in trypanosomatids', *Essays Biochem*. 2011/10/26, 51, pp. 31–46. doi: bse0510031 [pii]10.1042/bse0510031.
- Gerber, A. P., Herschlag, D. and Brown, P. O. (2004) 'Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast', *PLoS Biol*. 2004/03/17, 2(3), p. E79. doi: 10.1371/journal.pbio.0020079.
- de Godoy, L. M., Marchini, F. K., Pavoni, D. P., Rampazzo Rde, C., Probst, C. M., Goldenberg, S.

- and Krieger, M. A. (2012) 'Quantitative proteomics of *Trypanosoma cruzi* during metacyclogenesis', *Proteomics*. 2012/07/05, 12(17), pp. 2694–2703. doi: 10.1002/pmic.201200078.
- Gomez, M. L., Erijman, L., Arauzo, S., Torres, H. N. and Tellez-Inon, M. T. (1989) 'Protein kinase C in *Trypanosoma cruzi* epimastigote forms: partial purification and characterization', *Mol Biochem Parasitol*. 1989/09/01, 36(2), pp. 101–108. doi: 0166-6851(89)90182-5 [pii].
- Gondran, P., Amiot, F., Weil, D. and Dautry, F. (1999) 'Accumulation of mature mRNA in the nuclear fraction of mammalian cells.', *FEBS letters*, 458(3), pp. 324–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10570933> (Accessed: 28 June 2017).
- Greif, G., Ponce de Leon, M., Lamolle, G., Rodriguez, M., Piñeyro, D., Tavares-Marques, L. M., Reyna-Bello, A., Robello, C. and Alvarez-Valin, F. (2013) 'Transcriptome analysis of the bloodstream stage from the parasite *Trypanosoma vivax*', *BMC Genomics*, 14(1), p. 149. doi: 10.1186/1471-2164-14-149.
- Grisard, E. C., Teixeira, S. M. R., de Almeida, L. G. P., Stoco, P. H., Gerber, A. L., Talavera-Lopez, C., Lima, O. C., Andersson, B. and de Vasconcelos, A. T. R. (2014) 'Trypanosoma cruzi Clone Dm28c Draft Genome Sequence', *Genome Announcements*, 2(1), pp. e01114-13-e01114-13. doi: 10.1128/genomeA.01114-13.
- Grynberg, P., Passos-Silva, D. G., Mourao Mde, M., Hirata Jr., R., Macedo, A. M., Machado, C. R., Bartholomeu, D. C. and Franco, G. R. (2012) 'Trypanosoma cruzi gene expression in response to gamma radiation', *PLoS One*. 2012/01/17, 7(1), p. e29596. doi: 10.1371/journal.pone.0029596PONE-D-11-12240 [pii].
- Gualdrón-Lopez, M., Brennand, A., Hannaert, V., Quinones, W., Cáceres, A. J., Bringaud, F., Concepcion, J. L. and Michels, P. A. (2012) 'When, how and why glycolysis became compartmentalised in the Kinetoplastea. A new look at an ancient organelle', *Int J Parasitol*, 42(1), pp. 1–20. doi: 10.1016/j.ijpara.2011.10.007.
- Guerra-Giraldez, C., Quijada, L. and Clayton, C. E. (2002) 'Compartmentation of enzymes in a microbody, the glycosome, is essential in *Trypanosoma brucei*', *J Cell Sci*, 115(Pt 13), pp. 2651–2658. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12077356>.
- Guerra-Slompo, E. P., Probst, C. M., Pavoni, D. P., Goldenberg, S., Krieger, M. A. and Dallagiovanna, B. (2012) 'Molecular characterization of the *Trypanosoma cruzi* specific RNA binding protein TcRBP40 and its associated mRNAs', *Biochem Biophys Res Commun*. 2012/03/20, 420(2), pp. 302–307. doi: S0006-291X(12)00428-7 [pii]10.1016/j.bbrc.2012.02.154.
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. T. L., Noble, W. S., Maniatis, T., Goodbourn, S., Fischer, J., Pawson, T., Nash, P., Tompa, M., Li, N., Bailey, T. T. L., Church, G., Moor, B., Eskin, E., Favorov, A., Frith, M., Fu, Y., Kent, W., Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W., Lenhard, B., Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., Schacherer, F., Henikoff, S., Henikoff, J., Pietrokovski, S., Hughes, J., Estep, P., Tavazoie, S., Church, G., Wang, T., Stormo, G., Schones, D., Sumazin, P., Zhang, M., Roepcke, S., Grossmann, S., Rahmann, S., Vingron, M., Thijs, G., Marchal, K., Lescot, M., Rombauts, S., Moor, B., Rouze, P., Moreau, Y., Aerts, S., Loo, P., Thijs, G., Moreau, Y., Moor, B., Choi, I., Kwon, J., Kim, S., Sandelin, A., Wasserman, W., Staden, R., Tatusov, R., Altschul, S., Koonin, E., Bailey, T. T. L., Gribskov, M., Bailey, T. T. L., Elkan, C., Grundy, W., Bailey, T. T. L., Elkan, C., Baker, M., Hanley, J., McNeil, B., Agresti, A., Crooks, G., Hon, G., Chandonia, J. and Brenner, S. (2007) 'Quantifying similarity between motifs.', *Genome biology*. BioMed Central, 8(2), p. R24. doi: 10.1186/gb-2007-8-2-r24.
- Haag, J., O'hUigin, C. and Overath, P. (1998) 'The molecular phylogeny of trypanosomes:

- evidence for an early divergence of the *Salivaria*', *Mol Biochem Parasitol*, 91(1), pp. 37–49. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9574924> (Accessed: 13 February 2017).
- Haile, S., Estevez, A. M. and Clayton, C. (2003) 'A role for the exosome in the in vivo degradation of unstable mRNAs', *RNA*, 2003/11/19, 9(12), pp. 1491–1501. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14624005.
- Haile, S. and Papadopoulou, B. (2007) 'Developmental regulation of gene expression in trypanosomatid parasitic protozoa', *Curr Opin Microbiol*, 2008/01/08, 10(6), pp. 569–577. doi: 10.1016/j.mib.2007.10.001.
- Halbeisen, R. E., Galgano, A., Scherrer, T. and Gerber, A. P. (2008) 'Post-transcriptional gene regulation: from genome-wide studies to principles', *Cell Mol Life Sci*, 2007/11/29, 65(5), pp. 798–813. doi: 10.1007/s00018-007-7447-6.
- Hall, T. A. (1999) 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.', *Nucleic Acids Symposium Series*, 41, pp. 95–98.
- Hall, B. S., Bot, C. and Wilkinson, S. R. (2011) 'Nifurtimox Activation by Trypanosomal Type I Nitroreductases Generates Cytotoxic Nitrile Metabolites', *Journal of Biological Chemistry*, 286(15), pp. 13088–13095. doi: 10.1074/jbc.M111.230847.
- Hamed, A., Botelho, L., Britto, C., Fragoso, S. P., Umaki, A. C., Goldenberg, S., Bottu, G. and Salmon, D. (2015) 'In vitro metacyclogenesis of *Trypanosoma cruzi* induced by starvation correlates with a transient adenylyl cyclase stimulation as well as with a constitutive upregulation of adenylyl cyclase expression', *Mol Biochem Parasitol*, 200(1–2), pp. 9–18. doi: 10.1016/j.molbiopara.2015.04.002.
- Hartmann, C., Benz, C., Brems, S., Ellis, L., Luu, V.-D., Stewart, M., D'Orso, I., Busold, C., Fellenberg, K., Frasch, A. C. C., Carrington, M., Hoheisel, J. and Clayton, C. E. (2007) 'Small Trypanosome RNA-Binding Proteins TbUBP1 and TbUBP2 Influence Expression of F-Box Protein mRNAs in Bloodstream Trypanosomes', *Eukaryotic Cell*, 6(11), pp. 1964–1978. doi: 10.1128/EC.00279-07.
- Hartmann, C. and Clayton, C. (2008) 'Regulation of a transmembrane protein gene family by the small RNA-binding proteins TbUBP1 and TbUBP2', *Mol Biochem Parasitol*, 2007/11/21, 157(1), pp. 112–115. doi: 10.1016/j.molbiopara.2007.09.009.
- Hendriks, E. F. and Matthews, K. R. (2005) 'Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein', *Mol Microbiol*, 57(3), pp. 706–716. doi: 10.1111/j.1365-2958.2005.04679.x.
- Hendriks, E. F., Robinson, D. R., Hinkins, M. and Matthews, K. R. (2001) 'A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form', *EMBO J*, 20(23), pp. 6700–6711. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11726506.
- Herold, A., Teixeira, L. and Izaurralde, E. (2003) 'Genome-wide analysis of nuclear mRNA export pathways in *Drosophila*', *EMBO J*, 22(10), pp. 2472–2483. doi: 10.1093/emboj/cdg233.
- Hershberg, R. and Petrov, D. A. (2008) 'Selection on Codon Bias', *Annual Review of Genetics*, 42(1), pp. 287–299. doi: 10.1146/annurev.genet.42.110807.091442.
- Hessle, V., Bjork, P., Sokolowski, M., Gonzalez de Valdivia, E., Silverstein, R., Artemenko, K., Tyagi, A., Maddalo, G., Ilag, L., Helbig, R., Zubarev, R. A. and Visa, N. (2009) 'The exosome associates cotranscriptionally with the nascent pre-mRNP through interactions with heterogeneous nuclear

- ribonucleoproteins', *Mol Biol Cell*. 2009/06/06, 20(15), pp. 3459–3470. doi: E09-01-0079 [pii]10.1091/mbc.E09-01-0079.
- Holetz, F. B., Correa, A., Avila, A. R., Nakamura, C. V., Krieger, M. A. and Goldenberg, S. (2007) 'Evidence of P-body-like structures in *Trypanosoma cruzi*', *Biochem Biophys Res Commun*. 2007/04/03, 356(4), pp. 1062–1067. doi: S0006-291X(07)00589-X [pii]10.1016/j.bbrc.2007.03.104.
- Horn, D. (2008) 'Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids', *BMC Genomics*, 9(1), p. 2. doi: 10.1186/1471-2164-9-2.
- Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D. and Savioli, L. (2007) 'Control of neglected tropical diseases', *N Engl J Med*. 2007/09/07, 357(10), pp. 1018–1027. doi: 357/10/1018 [pii]10.1056/NEJMra064142.
- Hotez, P., Raff, S., Fenwick, A., Richards Jr., F. and Molyneux, D. H. (2007) 'Recent progress in integrated neglected tropical disease control', *Trends Parasitol*. 2007/10/24, 23(11), pp. 511–514. doi: S1471-4922(07)00249-8 [pii]10.1016/j.pt.2007.08.015.
- Hotz, H. R., Hartmann, C., Huober, K., Hug, M. and Clayton, C. (1997) 'Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation', *Nucleic Acids Res*, 25(15), pp. 3017–3026. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9224601.
- Houston-Ludlam, G. A., Belew, A. T. and El-Sayed, N. M. (2016) 'Comparative Transcriptome Profiling of Human Foreskin Fibroblasts Infected with the Sylvio and Y Strains of *Trypanosoma cruzi*', *PLoS One*, 11(8), p. e0159197. doi: 10.1371/journal.pone.0159197.
- Ignatovich, A. V., Takagi, Y., Liu, Y., Nagata, K. and Ho, C. K. (2015) 'The messenger RNA decapping and recapping pathway in *Trypanosoma*.' *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(22), pp. 6967–72. doi: 10.1073/pnas.1424909112.
- Inoue, A. H., Serpeloni, M., Hiraiwa, P. M., Yamada-Ogatta, S. F., Muniz, J. R., Motta, M. C., Vidal, N. M., Goldenberg, S. and Avila, A. R. (2014) 'Identification of a novel nucleocytoplasmic shuttling RNA helicase of trypanosomes', *PLoS One*, 9(10), p. e109521. doi: 10.1371/journal.pone.0109521.
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B. and Myler, P. J. (2005) 'The genome of the kinetoplastid parasite, *Leishmania major*', *Science*, 309(5733), pp. 436–442. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16020728.

- Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J. M., Dacks, J. B., Acosta-Serrano, A., Field, M. C., Ginger, M. L. and Berriman, M. (2016) 'Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism', *Curr Biol*, 26(2), pp. 161–172. doi: 10.1016/j.cub.2015.11.055.
- Jager, A. V., De Gaudenzi, J. G., Cassola, A., D'Orso, I. and Frasch, A. C. (2007) 'mRNA maturation by two-step trans-splicing/polyadenylation processing in trypanosomes', *Proc Natl Acad Sci U S A*. 2007/02/03, 104(7), pp. 2035–2042. doi: 0611125104 [pii]10.1073/pnas.0611125104.
- Jensen, B. C., Ramasamy, G., Vasconcelos, E. J., Ingolia, N. T., Myler, P. J. and Parsons, M. (2014) 'Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*', *BMC Genomics*, 15, p. 911. doi: 10.1186/1471-2164-15-911.
- Jimeno, S. and Aguilera, A. (2010) 'The THO complex as a key mRNP biogenesis factor in development and cell differentiation', *J Biol*, 9(1), p. 6. doi: 10.1186/jbiol217.
- Jimeno, S., Rondón, A. G., Luna, R. and Aguilera, A. (2002) 'The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability.', *The EMBO journal*. European Molecular Biology Organization, 21(13), pp. 3526–35. doi: 10.1093/emboj/cdf335.
- Jonas, S. and Izaurralde, E. (2015) 'Towards a molecular understanding of microRNA-mediated gene silencing', *Nature Reviews Genetics*. Nature Research, 16(7), pp. 421–433. doi: 10.1038/nrg3965.
- Jung, H., Gkogkas, C. G., Sonenberg, N. and Holt, C. E. (2014) 'Remote control of gene function by local translation', *Cell*, 157(1), pp. 26–40. doi: 10.1016/j.cell.2014.03.005.
- Kawashita, S. Y., da Silva, C. V., Mortara, R. A., Burleigh, B. A. and Briones, M. R. (2009) 'Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins', *Mol Biochem Parasitol*. 2009/04/28, 165(1), pp. 19–31. doi: S0166-6851(09)00003-6 [pii]10.1016/j.molbiopara.2008.12.010.
- Keene, J. D. (2007) 'RNA regulons: coordination of post-transcriptional events', *Nat Rev Genet*. 2007/06/19, 8(7), pp. 533–543. doi: nrg2111 [pii]10.1038/nrg2111.
- Kohler, A. and Hurt, E. (2007) 'Exporting RNA from the nucleus to the cytoplasm', *Nat Rev Mol Cell Biol*. 2007/09/06, 8(10), pp. 761–773. doi: nrm2255 [pii]10.1038/nrm2255.
- Kolev, N. G., Franklin, J. B., Carmi, S., Shi, H., Michaeli, S. and Tschudi, C. (2010) 'The Transcriptome of the Human Pathogen *Trypanosoma brucei* at Single-Nucleotide Resolution', *PLoS Pathogens*. Edited by S. M. Beverley. Public Library of Science, 6(9), p. e1001090. doi: 10.1371/journal.ppat.1001090.
- Konarska, M. M., Grabowski, P. J., Padgett, R. A. and Sharp, P. A. (1985) 'Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors', *Nature*, 313(6003), pp. 552–557. doi: 10.1038/313552a0.
- Kramer, S. (2012) 'Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids', *Mol Biochem Parasitol*. 2011/10/25, 181(2), pp. 61–72. doi: S0166-6851(11)00225-8 [pii]10.1016/j.molbiopara.2011.10.002.
- Kramer, S. (2014) 'RNA in development: how ribonucleoprotein granules regulate the life cycles of pathogenic protozoa', *Wiley Interdiscip Rev RNA*, 5(2), pp. 263–284. doi: 10.1002/wrna.1207.
- Kramer, S. and Carrington, M. (2011) 'Trans-acting proteins regulating mRNA maturation,

stability and translation in trypanosomatids', *Trends Parasitol.* 2010/07/09, 27(1), pp. 23–30. doi: S1471-4922(10)00121-2 [pii]10.1016/j.pt.2010.06.011.

Kramer, S., Piper, S., Estevez, A. and Carrington, M. (2016) 'Polycistronic trypanosome mRNAs are a target for the exosome', *Mol Biochem Parasitol*, 205(1–2), pp. 1–5. doi: 10.1016/j.molbiopara.2016.02.009.

Kramer, S., Queiroz, R., Ellis, L., Hoheisel, J. D., Clayton, C. and Carrington, M. (2010) 'The RNA helicase DHH1 is central to the correct expression of many developmentally regulated mRNAs in trypanosomes', *J Cell Sci.* 2010/02/04, 123(Pt 5), pp. 699–711. doi: jcs.058511 [pii]10.1242/jcs.058511.

Kramer, S., Queiroz, R., Ellis, L., Webb, H., Hoheisel, J. D., Clayton, C. and Carrington, M. (2008) 'Heat shock causes a decrease in polysomes and the appearance of stress granules in trypanosomes independently of eIF2(alpha) phosphorylation at Thr169', *J Cell Sci.* 2008/08/21, 121(Pt 18), pp. 3002–3014. doi: jcs.031823 [pii]10.1242/jcs.031823.

LeBowitz, J. H., Smith, H. Q., Rusche, L. and Beverley, S. M. (1993) 'Coupling of poly(A) site selection and trans-splicing in Leishmania', *Genes Dev.* 1993/06/01, 7(6), pp. 996–1007. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8504937.

Levy, G. V., Bañuelos, C. P., Níttolo, A. G., Ortiz, G. E., Mendiondo, N., Moretti, G., Tekiel, V. S. and Sánchez, D. O. (2015) 'Depletion of the SR-Related Protein TbRRM1 Leads to Cell Cycle Arrest and Apoptosis-Like Death in Trypanosoma brucei.', *PLoS one.* Public Library of Science, 10(8), p. e0136070. doi: 10.1371/journal.pone.0136070.

Li, Y., Shah-Simpson, S., Okrah, K., Belew, A. T., Choi, J., Caradonna, K. L., Padmanabhan, P., Ndegwa, D. M., Temanni, M. R., Corrada Bravo, H., El-Sayed, N. M. and Burleigh, B. A. (2016) 'Transcriptome Remodeling in Trypanosoma cruzi and Human Cells during Intracellular Infection', *PLoS Pathog.* Public Library of Science, 12(4), p. e1005511. doi: 10.1371/journal.ppat.1005511.

Li, Z. H., De Gaudenzi, J. G., Alvarez, V. E., Mendiondo, N., Wang, H., Kissinger, J. C., Frasch, A. C. and Docampo, R. (2012) 'A 43-nucleotide U-rich element in 3'-untranslated region of large number of Trypanosoma cruzi transcripts is important for mRNA abundance in intracellular amastigotes', *J Biol Chem.* 2012/04/14, 287(23), pp. 19058–19069. doi: M111.338699 [pii]10.1074/jbc.M111.338699.

Liang, X. H., Haritan, A., Uliel, S. and Michaeli, S. (2003) 'trans and cis splicing in trypanosomatids: mechanism, factors, and regulation', *Eukaryot Cell.* 2003/10/14, 2(5), pp. 830–840. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14555465.

Liao, J. Y., Ma, L. M., Guo, Y. H., Zhang, Y. C., Zhou, H., Shao, P., Chen, Y. Q. and Qu, L. H. (2010) 'Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers', *PLoS One.* 2010/05/26, 5(5), p. e10563. doi: 10.1371/journal.pone.0010563.

Linder, J. C. and Staehelin, L. A. (1977) 'Plasma membrane specializations in a trypanosomatid flagellate', *J Ultrastruct Res*, 60(2), pp. 246–262. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/886643>.

Lueong, S., Merce, C., Fischer, B., Hoheisel, J. D. and Erben, E. D. (2016) 'Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome', *Molecular Microbiology*, 100(3), pp. 457–471. doi: 10.1111/mmi.13328.

- Luo, M. J. and Reed, R. (1999) 'Splicing is required for rapid and efficient mRNA export in metazoans.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(26), pp. 14937–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10611316> (Accessed: 12 May 2017).
- Madison-Antenucci, S., Grams, J. and Hajduk, S. L. (2002) 'Editing machines: the complexities of trypanosome RNA editing', *Cell*, 108(4), pp. 435–438. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11909515.
- Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H. O., Bishop, J. R., Falcone, F. H., Gavrilescu, C., Montgomery, J. L., Santori, M. I., Stern, L. S., Wang, Z., Ullu, E. and Tschudi, C. (2000) 'A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA', *RNA*, 6(2), pp. 163–169. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10688355.
- Makin, L. and Gluenz, E. (2015) 'cAMP signalling in trypanosomatids: role in pathogenesis and as a drug target', *Trends Parasitol*, 31(8), pp. 373–379. doi: 10.1016/j.pt.2015.04.014.
- Mallick, B., Ghosh, Z. and Chakrabarti, J. (2008) 'MicroRNA switches in *Trypanosoma brucei*', *Biochemical and Biophysical Research Communications*, 372(3), pp. 459–463. doi: 10.1016/j.bbrc.2008.05.084.
- Manful, T., Cristodero, M. and Clayton, C. (2009) *DRBD1 is the Trypanosoma brucei homologue of the Spliceosome-Associated Protein 49*, *Molecular and Biochemical Parasitology*. doi: 10.1016/j.molbiopara.2009.03.003.
- Manger, I. D. and Boothroyd, J. C. (1998) 'Identification of a nuclear protein in *Trypanosoma brucei* with homology to RNA-binding proteins from cis-splicing systems', *Mol Biochem Parasitol*, 97(1–2), pp. 1–11. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9879882.
- Martinez-Calvillo, S., Vizuet-de-Rueda, J. C., Florencio-Martinez, L. E., Manning-Cela, R. G. and Figueroa-Angulo, E. E. (2010) 'Gene expression in trypanosomatid parasites', *J Biomed Biotechnol*. 2010/02/20, 2010, p. 525241. doi: 10.1155/2010/525241.
- Matthews, K. R. (2005) 'The developmental cell biology of *Trypanosoma brucei*', *J Cell Sci*, 118(Pt 2), pp. 283–290. doi: 10.1242/jcs.01649.
- Matthews, K. R., Tschudi, C. and Ullu, E. (1994) 'A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes', *Genes Dev*. 1994/02/15, 8(4), pp. 491–501. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7907303.
- Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S. and Matthews, K. (2006) 'Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in *Trypanosoma brucei*: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements', *Nucleic Acids Res*, 34(18), pp. 5312–5324. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17012283.
- Minning, T. A., Weatherly, D. B., Atwood 3rd, J., Orlando, R. and Tarleton, R. L. (2009) 'The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*', *BMC Genomics*. 2009/08/12, 10, p. 370. doi: 1471-2164-10-370 [pii]10.1186/1471-2164-10-370.

Morillo, C. A., Marin-Neto, J. A., Avezum, A., Sosa-Estani, S., Rassi, A., Rosas, F., Villena, E., Quiroz, R., Bonilla, R., Britto, C., Guhl, F., Velazquez, E., Bonilla, L., Meeks, B., Rao-Melacini, P., Pogue, J., Mattos, A., Lazdins, J., Rassi, A., Connolly, S. J. and Yusuf, S. (2015) 'Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy', *New England Journal of Medicine*. Massachusetts Medical Society, 373(14), pp. 1295–1306. doi: 10.1056/NEJMoa1507574.

Morking, P. A., Dallagiovanna, B. M., Foti, L., Garat, B., Picchi, G. F., Umaki, A. C., Probst, C. M., Krieger, M. A., Goldenberg, S. and Fragoso, S. P. (2004) 'TcZFP1: a CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro', *Biochem Biophys Res Commun*, 319(1), pp. 169–177. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15158457.

Morking, P. A., Rampazzo Rde, C., Walrad, P., Probst, C. M., Soares, M. J., Gradia, D. F., Pavoni, D. P., Krieger, M. A., Matthews, K., Goldenberg, S., Fragoso, S. P. and Dallagiovanna, B. (2012) 'The zinc finger protein TcZFP2 binds target mRNAs enriched during *Trypanosoma cruzi* metacyclogenesis', *Mem Inst Oswaldo Cruz*. 2012/09/20, 107(6), pp. 790–799. doi: S0074-02762012000600014 [pii].

Naguleswaran, A., Gunasekera, K., Schimanski, B., Heller, M., Hemphill, A., Ochsenreiter, T. and Roditi, I. (2015) 'Trypanosoma brucei RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure.', *mBio*. American Society for Microbiology, 6(2), p. e00114. doi: 10.1128/mBio.00114-15.

Nardelli, S. C., Avila, A. R., Freund, A., Motta, M. C., Manhães, L., de Jesus, T. C. L., Schenkman, S., Fragoso, S. P., Krieger, M. A., Goldenberg, S. and Dallagiovanna, B. (2007) 'Small-subunit rRNA processome proteins are translationally regulated during differentiation of *Trypanosoma cruzi*.' *Eukaryotic cell*, 6(2), pp. 337–45. doi: 10.1128/EC.00279-06.

Nardy, A. F., Freire-de-Lima, C. G., Perez, A. R. and Morrot, A. (2016) 'Role of *Trypanosoma cruzi* Trans-sialidase on the Escape from Host Immune Surveillance', *Front Microbiol*, 7, p. 348. doi: 10.3389/fmicb.2016.00348.

Nazer, E. and Sanchez, D. O. (2011) 'Nucleolar accumulation of RNA binding proteins induced by Actinomycin D is functional in *Trypanosoma cruzi* and *Leishmania mexicana* but not in *T. brucei*', *PLoS One*, 6(8), p. e24184. doi: 10.1371/journal.pone.0024184.

Nazer, E., Verdun, R. E. and Sanchez, D. O. (2011) 'Nucleolar localization of RNA binding proteins induced by actinomycin D and heat shock in *Trypanosoma cruzi*', *PLoS One*, 6(5), p. e19920. doi: 10.1371/journal.pone.0019920.

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I. and Ochsenreiter, T. (2010) 'Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*', *PLoS Pathog*. 2010/08/12, 6(8). doi: 10.1371/journal.ppat.1001037.

Di Noia, J. M., D'Orso, I., Aslund, L., Sanchez, D. O. and Frasch, A. C. (1998) 'The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions', *J Biol Chem*, 273(18), p. 10843–50. Available at: <http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/referer?http://www.jbc.org/cgi/content/full/273/18/10843>.

Di Noia, J. M., D'Orso, I., Sanchez, D. O. and Frasch, A. C. (2000) 'AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency', *J Biol Chem*, 275(14), pp. 10218–10227. Available at: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list>

_uids=10744707.

O'Connell, D. (2007) 'Neglected Diseases', *Nature*. Nature Publishing Group, 449(7159), pp. 157–157. doi: 10.1038/449157a.

Obado, S. O., Brillantes, M., Uryu, K., Zhang, W., Ketaren, N. E., Chait, B. T., Field, M. C. and Rout, M. P. (2016) 'Interactome Mapping Reveals the Evolutionary History of the Nuclear Pore Complex', *PLoS Biol*, 14(2). doi: 10.1371/journal.pbio.1002365.

Obado, S. O., Field, M. C. and Rout, M. P. (2017) 'Comparative interactomics provides evidence for functional specialization of the nuclear pore complex', *Nucleus*, pp. 00–00. doi: 10.1080/19491034.2017.1313936.

Opperdoes, F. R. and Borst, P. (1977) 'Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome', *FEBS Lett*, 80(2), pp. 360–364. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/142663>.

Ouellette, M. and Papadopoulou, B. (2009) 'Coordinated gene expression by post-transcriptional regulons in African trypanosomes', *J Biol*, 8(11), p. 100. doi: 10.1186/jbiol203.

De Pablos, L. M., Diaz Lozano, I. M., Jercic, M. I., Quinzada, M., Gimenez, M. J., Calabuig, E., Espino, A. M., Schijman, A. G., Zulantay, I., Apt, W. and Osuna, A. (2016) 'The C-terminal region of *Trypanosoma cruzi* MASPs is antigenic and secreted via exovesicles', *Sci Rep*, 6, p. 27293. doi: 10.1038/srep27293.

De Pablos, L. M. and Osuna, A. (2012) 'Multigene families in *Trypanosoma cruzi* and their role in infectivity', *Infect Immun*. 2012/03/21. doi: IAI.06225-11 [pii]10.1128/IAI.06225-11.

Padilla-Mejía, N. E., Florencio-Martínez, L. E., Figueroa-Angulo, E. E., Manning-Cela, R. G., Hernández-Rivas, R., Myler, P. J. and Martínez-Calvillo, S. (2009) 'Gene organization and sequence analyses of transfer RNA genes in Trypanosomatid parasites', *BMC Genomics*, 10(1), p. 232. doi: 10.1186/1471-2164-10-232.

Palazzo, A. F. and Akef, A. (2012) 'Nuclear export as a key arbiter of "mRNA identity" in eukaryotes', *Biochim Biophys Acta*, 1819(6), pp. 566–577. doi: 10.1016/j.bbagr.2011.12.012.

Palazzo, A. F. and Akef, A. (2012) 'Nuclear export as a key arbiter of "mRNA identity" in eukaryotes', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(6), pp. 566–577. doi: 10.1016/j.bbagr.2011.12.012.

Parker, R. and Sheth, U. (2007) 'P bodies and the control of mRNA translation and degradation', *Mol Cell*. 2007/03/14, 25(5), pp. 635–646. doi: S1097-2765(07)00111-6 [pii]10.1016/j.molcel.2007.02.011.

Parsons, M., Stuart, K. and Smiley, B. L. (1991) 'Trypanosoma brucei: analysis of codon usage and nucleotide composition of nuclear genes.', *Experimental parasitology*, 73(1), pp. 101–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2055297> (Accessed: 9 March 2017).

Pastro, L., Smircich, P., Di Paolo, A., Becco, L., Duhagon, M. A., Sotelo-Silveira, J. and Garat, B. (2017) 'Nuclear Compartmentalization Contributes to Stage-Specific Gene Expression Control in *Trypanosoma cruzi*', *Frontiers in Cell and Developmental Biology*. Frontiers, 5, p. 8. doi: 10.3389/fcell.2017.00008.

Paterou, A., Walrad, P., Craddy, P., Fenn, K. and Matthews, K. (2006) 'Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development', *J Biol Chem*. 2006/10/18, 281(51), pp. 39002–39013. doi: M604280200 [pii]10.1074/jbc.M604280200.

Perez-Diaz, L., Correa, A., Moretao, M. P., Goldenberg, S., Dallagiovanna, B. and Garat, B. (2012) 'The overexpression of the trypanosomatid-exclusive TcRBP19 RNA-binding protein affects cellular infection by *Trypanosoma cruzi*', *Mem Inst Oswaldo Cruz*. 2013/01/09, 107(8), pp. 1076–1079. doi: S0074-02762012000800021 [pii].

Perez-Diaz, L., Duhagon, M. A., Smircich, P., Sotelo-Silveira, J., Robello, C., Krieger, M. A., Goldenberg, S., Williams, N., Dallagiovanna, B. and Garat, B. (2007) 'Trypanosoma cruzi: molecular characterization of an RNA binding protein differentially expressed in the parasite life cycle', *Exp Parasitol*. 2007/05/04, 117(1), pp. 99–105. doi: S0014-4894(07)00084-7 [pii]10.1016/j.exppara.2007.03.010.

Perez-Diaz, L., Pastro, L., Smircich, P., Dallagiovanna, B. and Garat, B. (2013) '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', *Biochem Biophys Res Commun*. 2013/06/08, 436(2), pp. 295–299. doi: S0006-291X(13)00903-0 [pii]10.1016/j.bbrc.2013.05.096.

Pérez-Díaz, L., Silva, T. C. and Teixeira, S. M. R. (2017) 'Involvement of an RNA binding protein containing Alba domain in the stage-specific regulation of beta-amastin expression in *Trypanosoma cruzi*', *Molecular and Biochemical Parasitology*, 211, pp. 1–8. doi: 10.1016/j.molbiopara.2016.12.005.

Polak, A. and Richle, R. (1978) 'Mode of action of the 2-nitroimidazole derivative benznidazole.', *Annals of tropical medicine and parasitology*, 72(1), pp. 45–54. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/418744> (Accessed: 18 May 2017).

Portal, D., Espinosa, J. M., Lobo, G. S., Kadener, S., Pereira, C. A., De La Mata, M., Tang, Z., Lin, R. J., Kornblihtt, A. R., Baralle, F. E., Flawia, M. M. and Torres, H. N. (2003) 'An early ancestor in the evolution of splicing: a *Trypanosoma cruzi* serine-arginine-rich protein (TcSR) is functional in cis-splicing', *Mol Biochem Parasitol*, 127(1), pp. 37–46. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12615334.

Portal, D., Lobo, G. S., Kadener, S., Prasad, J., Espinosa, J. M., Pereira, C. A., Tang, Z., Lin, R. J., Manley, J. L., Kornblihtt, A. R., Flawia, M. M. and Torres, H. N. (2003) 'Trypanosoma cruzi TcSRPK, the first protozoan member of the SRPK family, is biochemically and functionally conserved with metazoan SR protein-specific kinases', *Mol Biochem Parasitol*. 2003/03/05, 127(1), pp. 9–21. doi: S0166685102002992 [pii].

Porto-Carreiro, I., Attias, M., Miranda, K., De Souza, W. and Cunha-e-Silva, N. (2000) 'Trypanosoma cruzi epimastigote endocytic pathway: cargo enters the cytosome and passes through an early endosomal network before storage in reservosomes', *Eur J Cell Biol*, 79(11), pp. 858–869. doi: 10.1078/0171-9335-00112.

Qiagen (no date) *CLC Genomics Workbench*. Available at: <https://www.qiagenbioinformatics.com/>.

Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J. D. and Clayton, C. (2009) 'Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons', *BMC Genomics*. 2009/10/28, 10, p. 495. doi: 1471-2164-10-495 [pii]10.1186/1471-2164-10-495.

Rassi, A., Rassi, A. and Marcondes de Rezende, J. (2012) 'American Trypanosomiasis (Chagas Disease)', *Infectious Disease Clinics of North America*, 26(2), pp. 275–291. doi: 10.1016/j.idc.2012.03.002.

Rastrojo, A., Carrasco-Ramiro, F., Martín, D., Crespillo, A., Reguera, R. M., Aguado, B. and Requena, J. M. (2013) 'The transcriptome of *Leishmania major* in the axenic promastigote stage:

transcript annotation and relative expression levels by RNA-seq', *BMC Genomics*, 14(1), p. 223. doi: 10.1186/1471-2164-14-223.

Ray, D., Weirauch, H. K. K. B. C. M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Rakesh, S., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. O. F., Lei, E. P., Fraser, A. G., Blencowe, B. J., Morris, Q. D. and Hughes, T. R. (2013) 'A compendium of RNA-binding motifs for decoding gene regulation.', *Nature*, 499(7457), pp. 172–7. doi: 10.1038/nature12311.

Reifur, L., Garcia-Silva, M. R., Poubel, S. B., Alves, L. R., Arauco, P., Buiar, D. K., Goldenberg, S., Cayota, A. and Dallagiovanna, B. (2012) 'Distinct subcellular localization of tRNA-derived fragments in the infective metacyclic forms of *Trypanosoma cruzi*', *Mem Inst Oswaldo Cruz*. 2012/09/20, 107(6), pp. 816–819. doi: S0074-02762012000600018 [pii].

Remme, J. H. F., Feenstra, P., Lever, P. R., Medici, A. C., Morel, C. M., Noma, M., Ramaiah, K. D., Richards, F., Seketeli, A., Schmunis, G., van Brakel, W. H. and Vassall, A. (2006) 'Tropical Diseases Targeted for Elimination: Chagas Disease, Lymphatic Filariasis, Onchocerciasis, and Leprosy', in Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A., and Musgrove, P. (eds) *Disease Control Priorities in Developing Countries*. 2nd edn. Washington (DC). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21250324> (Accessed: 13 February 2017).

Roberts, T. G., Sturm, N. R., Yee, B. K., Yu, M. C., Hartshorne, T., Agabian, N. and Campbell, D. A. (1998) 'Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans', *Mol Cell Biol*. 1998/07/22, 18(8), pp. 4409–4417. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9671450.

Rout, M. P. and Field, M. C. (2001) 'Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component', *J Biol Chem*. 2001/07/31, 276(41), pp. 38261–38271. doi: 10.1074/jbc.M104024200M104024200 [pii].

Ruvalcaba-Trejo, L. I. and Sturm, N. R. (2011) 'The *Trypanosoma cruzi* Sylvio X10 strain maxicircle sequence: the third musketeer', *BMC Genomics*. 2011/01/26, 12, p. 58. doi: 1471-2164-12-58 [pii]10.1186/1471-2164-12-58.

Saas, J., Ziegelbauer, K., von Haeseler, A., Fast, B. and Boshart, M. (2000) 'A developmentally regulated aconitase related to iron-regulatory protein-1 is localized in the cytoplasm and in the mitochondrion of *Trypanosoma brucei*', *J Biol Chem*. 2000/01/25, 275(4), pp. 2745–2755. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10644738.

dos Santos, S. L., Freitas, L. M., Lobo, F. P., Rodrigues-Luiz, G. F., Mendes, T. A., Oliveira, A. C., Andrade, L. O., Chiari, E., Gazzinelli, R. T., Teixeira, S. M., Fujiwara, R. T. and Bartholomeu, D. C. (2012) 'The MASP family of *Trypanosoma cruzi*: changes in gene expression and antigenic profile during the acute phase of experimental infection', *PLoS Negl Trop Dis*, 6(8), p. e1779. doi: 10.1371/journal.pntd.0001779.

Schenkman, S., Eichinger, D., Pereira, M. E. and Nussenzweig, V. (1994) 'Structural and functional properties of *Trypanosoma trans-sialidase*', *Annu Rev Microbiol*. 1994/01/01, 48, pp. 499–523. doi: 10.1146/annurev.mi.48.100194.002435.

- Schenkman, S., Pascoalino Bdos, S. and Nardelli, S. C. (2011) 'Nuclear structure of *Trypanosoma cruzi*', *Adv Parasitol*, 75, pp. 251–283. doi: 10.1016/B978-0-12-385863-4.00012-5.
- Schwanhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) 'Global quantification of mammalian gene expression control', *Nature*, 473(7347), pp. 337–342. doi: 10.1038/nature10098.
- Serpeloni, M., Vidal, N. M., Goldenberg, S., Avila, A. R. and Hoffmann, F. G. (2011) 'Comparative genomics of proteins involved in RNA nucleocytoplasmic export', *BMC Evol Biol*, 11, p. 7. doi: 10.1186/1471-2148-11-7.
- Shapiro, T. A. and Englund, P. T. (1995) 'The structure and replication of kinetoplast DNA', *Annu Rev Microbiol*, 49(1), pp. 117–143. doi: 10.1146/annurev.mi.49.100195.001001.
- Shaw, J. M., Feagin, J. E., Stuart, K. and Simpson, L. (1988) 'Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons', *Cell*, 53(3), pp. 401–411. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2452696.
- Siegel, T. N., Gunasekera, K., Cross, G. A. and Ochsenreiter, T. (2011) 'Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing', *Trends Parasitol*, 27(10), pp. 434–441. doi: 10.1016/j.pt.2011.05.006.
- Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S. and Cross, G. A. (2010) 'Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites', *Nucleic Acids Res*. 2010/04/14, 38(15), pp. 4946–4957. doi: gkq237 [pii]10.1093/nar/gkq237.
- da Silva Augusto, L., Moretti, N. S., Ramos, T. C., de Jesus, T. C., Zhang, M., Castilho, B. A. and Schenkman, S. (2015) 'A membrane-bound eIF2 alpha kinase located in endosomes is regulated by heme and controls differentiation and ROS levels in *Trypanosoma cruzi*', *PLoS Pathog*, 11(2), p. e1004618. doi: 10.1371/journal.ppat.1004618.
- Simpson, L., Sbicego, S. and Aphasizhev, R. (2003) 'Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business', *RNA*, 9(3), pp. 265–276. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12591999>.
- Simpson, L., Thiemann, O. H., Savill, N. J., Alfonzo, J. D. and Maslov, D. A. (2000) 'Evolution of RNA editing in trypanosome mitochondria', *Proc Natl Acad Sci U S A*. National Academy of Sciences, 97(13), pp. 6986–6993. doi: 10.1073/PNAS.97.13.6986.
- Smircich, P., Eastman, G., Bispo, S., Duhagon, M. A., Guerra-Slompo, E. P., Garat, B., Goldenberg, S., Munroe, D. J., Dallagiovanna, B., Holetz, F. and Sotelo-Silveira, J. R. (2015) 'Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*', *BMC Genomics*, 16, p. 443. doi: 10.1186/s12864-015-1563-8.
- Smith, D. F. and Parsons, M. (1996) *Molecular biology of parasitic protozoa*. 1st edn, *Frontiers in molecular biology*; 13. 1st edn. Oxford ; New York: IRL Press at Oxford University Press.
- Solnestam, B. W., Stranneheim, H., Hallman, J., Kaller, M., Lundberg, E., Lundeberg, J. and Akan, P. (2012) 'Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs', *BMC Genomics*. 2012/11/01, 13, p. 574. doi: 1471-2164-13-574 [pii]10.1186/1471-2164-13-574.
- de Souza, W. (2009) 'Structural organization of *Trypanosoma cruzi*', *Mem Inst Oswaldo Cruz*. 2009/09/24, 104 Suppl, pp. 89–100. doi: S0074-02762009000900014 [pii].

- De Souza, W. (2002) 'Basic cell biology of *Trypanosoma cruzi*', *Curr Pharm Des*, 8(4), pp. 269–285. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11860366.
- de Souza, W., Attias, M. and Rodrigues, J. C. F. (2009) 'Particularities of mitochondrial structure in parasitic protists (Apicomplexa and Kinetoplastida)', *The International Journal of Biochemistry & Cell Biology*, 41(10), pp. 2069–2080. doi: 10.1016/j.biocel.2009.04.007.
- De Souza, W. and Meyer, H. (1974) 'On the fine structure of the nucleus in *Trypanosoma cruzi* in tissue culture forms. Spindle fibers in the dividing nucleus', *J Protozool*. 1974/02/01, 21(1), pp. 48–52. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4594241.
- Strambio-De-Castillia, C., Niepel, M. and Rout, M. P. (2010) 'The nuclear pore complex: bridging nuclear transport and gene regulation', *Nat Rev Mol Cell Biol*, 11(7), pp. 490–501. doi: 10.1038/nrm2928.
- Strom, A. C. and Weis, K. (2001) 'Importin-beta-like nuclear transport receptors', *Genome Biol*, 2(6), p. REVIEWS3008. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11423015>.
- Suganuma, K., Yamasaki, S., Asada, M., Kawazu, S. and Inoue, N. (2012) 'The epimastigote stage-specific gene expression of CESP is tightly regulated by its 3' UTR', *Mol Biochem Parasitol*. 2012/10/13, 186(1), pp. 77–80. doi: S0166-6851(12)00243-5 [pii]10.1016/j.molbiopara.2012.09.010.
- Tagoe, D. N., Kalejaiye, T. D. and de Koning, H. P. (2015) 'The ever unfolding story of cAMP signaling in trypanosomatids: vive la difference!', *Front Pharmacol*, 6, p. 185. doi: 10.3389/fphar.2015.00185.
- Talavera-López, C. N. (2016) 'TRYPANOSOMA CRUZI GENOME PLASTICITY AND EVOLUTION'.
- Tan, T. H. P., Pach, R., Crausaz, A., Ivens, A. and Schneider, A. (2002) 'tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import.', *Molecular and cellular biology*, 22(11), pp. 3707–17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11997507> (Accessed: 23 February 2017).
- Trask, H. W., Cowper-Sal-lari, R., Sartor, M. A., Gui, J., Heath, C. V, Renuka, J., Higgins, A. J., Andrews, P., Korc, M., Moore, J. H. and Tomlinson, C. R. (2009) 'Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs', *RNA*, 15(10), pp. 1917–1928. doi: 10.1261/rna.1677409.
- Ullu, E., Matthews, K. R. and Tschudi, C. (1993) 'Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts', *Mol Cell Biol*. 1993/01/01, 13(1), pp. 720–725. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8417363.
- Ulrich, P. N., Jimenez, V., Park, M., Martins, V. P., Atwood 3rd, J., Moles, K., Collins, D., Rohloff, P., Tarleton, R., Moreno, S. N., Orlando, R. and Docampo, R. (2011) 'Identification of contractile vacuole proteins in *Trypanosoma cruzi*', *PLoS One*, 6(3), p. e18013. doi: 10.1371/journal.pone.0018013.
- Valencia, P., Dias, A. P. and Reed, R. (2008) 'Splicing promotes rapid and efficient mRNA export in mammalian cells.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(9), pp. 3386–91. doi: 10.1073/pnas.0800250105.

- Vasquez, J. J., Hon, C. C., Vanselow, J. T., Schlosser, A. and Siegel, T. N. (2014) 'Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages', *Nucleic Acids Res*, 42(6), pp. 3623–3637. doi: 10.1093/nar/gkt1386.
- Vazquez, M., Atorrasagasti, C., Bercovich, N., Volcovich, R. and Levin, M. J. (2003) 'Unique features of the *Trypanosoma cruzi* U2AF35 splicing factor', *Mol Biochem Parasitol*, 128(1), pp. 77–81. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12706799.
- Vickerman, K. (1970) 'Functional aspects of the cytology of trypanosomes', *Trans R Soc Trop Med Hyg*, 64(1), pp. 180–181. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/5442056>.
- Vlasova-St Louis, I. and Bohjanen, P. R. (2011) 'Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1.', *Current opinion in genetics & development*, 21(4), pp. 444–51. doi: 10.1016/j.gde.2011.03.002.
- Walrad, P. B., Capewell, P., Fenn, K. and Matthews, K. R. (2012) 'The post-transcriptional trans-acting regulator, TbZFP3, co-ordinates transmission-stage enriched mRNAs in *Trypanosoma brucei*.' *Nucleic acids research*. Oxford University Press, 40(7), pp. 2869–83. doi: 10.1093/nar/gkr1106.
- Wang, Y., Opperman, L., Wickens, M. and Hall, T. M. (2009) 'Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein', *Proc Natl Acad Sci U S A*. 2009/11/11, 106(48), pp. 20186–20191. doi: 0812076106 [pii]10.1073/pnas.0812076106.
- Weil, D., Boutain, S., Audibert, A. and Dautry, F. (2000) 'Mature mRNAs accumulated in the nucleus are neither the molecules in transit to the cytoplasm nor constitute a stockpile for gene expression.', *RNA (New York, N.Y.)*, 6(7), pp. 962–75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10917593> (Accessed: 28 June 2017).
- Wen, Y. Z., Zheng, L. L., Qu, L. H., Ayala, F. J. and Lun, Z. R. (2012) 'Pseudogenes are not pseudo any more', *RNA Biol*, 9(1), pp. 27–32. doi: 10.4161/rna.9.1.18277.
- Wente, S. R. and Rout, M. P. (2010) 'The Nuclear Pore Complex and Nuclear Transport', *Cold Spring Harbor Perspectives in Biology*, 2(10), pp. a000562–a000562. doi: 10.1101/cshperspect.a000562.
- Westenberger, S. J., Cerqueira, G. C., El-Sayed, N. M., Zingales, B., Campbell, D. A. and Sturm, N. R. (2006) 'Trypanosoma cruzi mitochondrial maxicircles display species- and strain-specific variation and a conserved element in the non-coding region', *BMC Genomics*. 2006/03/24, 7, p. 60. doi: 1471-2164-7-60 [pii]10.1186/1471-2164-7-60.
- Weston, D., La Flamme, A. C. and Van Voorhis, W. C. (1999) 'Expression of *Trypanosoma cruzi* surface antigen FL-160 is controlled by elements in the 3' untranslated, the 3' intergenic, and the coding regions', *Mol Biochem Parasitol*, 102(1), pp. 53–66. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10477176.
- Wharton, R. P. and Aggarwal, A. K. (2006) 'mRNA regulation by Puf domain proteins', *Sci STKE*. 2006/09/28, 2006(354), p. pe37. doi: stke.3542006pe37 [pii]10.1126/stke.3542006pe37.
- WHO (2009) *Global_chagas_2009*. Available at: http://gamapserver.who.int/mapLibrary/Files/Maps/Global_chagas_2009.png (Accessed: 21 February 2017).
- WHO (2016) 'Chagas disease (American trypanosomiasis)', *World Health Organ Tech Rep*. World

Health Organization, (340). Available at: <http://www.who.int/mediacentre/factsheets/fs340/en/> (Accessed: 21 February 2017).

Wickens, M., Bernstein, D. S., Kimble, J. and Parker, R. (2002) 'A PUF family portrait: 3'UTR regulation as a way of life', *Trends Genet.* 2002/02/23, 18(3), pp. 150–157. doi: S0168952501026166 [pii].

Wittkopp, P. J. and Kalay, G. (2011) 'Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence', *Nat Rev Genet.* 13(1), pp. 59–69. doi: 10.1038/nrg3095.

Xu, P., Wen, L., Benegal, G., Wang, X. and Buck, G. A. (2001) 'Identification of a spliced leader RNA binding protein from *Trypanosoma cruzi*', *Mol Biochem Parasitol.* 112(1), pp. 39–49. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11166385.

Yao, C. (2010) 'Major surface protease of trypanosomatids: one size fits all?', *Infect Immun.* 78(1), pp. 22–31. doi: 10.1128/IAI.00776-09.

Yoshida, N. (2008) 'Trypanosoma cruzi infection by oral route', *Parasitology International.* 57(2), pp. 105–109. doi: 10.1016/j.parint.2007.12.008.

Zeiner, G. M., Sturm, N. R. and Campbell, D. A. (2003) 'Exportin 1 mediates nuclear export of the kinetoplastid spliced leader RNA', *Eukaryot Cell.* 2(2), pp. 222–230. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12684371>.

Zhang, K., Shi, Z.-M., Chang, Y.-N., Hu, Z.-M., Qi, H.-X. and Hong, W. (2014) 'The ways of action of long non-coding RNAs in cytoplasm and nucleus', *Gene.* 547(1), pp. 1–9. doi: 10.1016/j.gene.2014.06.043.

Zingales, B., Andrade, S. G., Briones, M. R., Campbell, D. A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A. M., Machado, C. R., Miles, M. A., Romanha, A. J., Sturm, N. R., Tibayrenc, M. and Schijman, A. G. (2009) 'A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI', *Mem Inst Oswaldo Cruz.* 2009/12/23, 104(7), pp. 1051–1054. doi: S0074-02762009000700021 [pii].