

ARN funcional en *Trypanosoma cruzi*, de virus a parásito

María del Carmen Thomas, Francisco J. Sánchez-Luque, Francisco Macías,
Patricia Carreira, Manuel C. López

Instituto de Parasitología y Biomedicina López Neyra, Consejo Superior de Investigaciones Científicas,
Parque Tecnológico de Ciencias de la Salud, Granada, España

Los retrotransposones son elementos de ADN móviles que se transponen por medio de un ARN intermediario, el cual es copiado e integrado en una nueva posición del genoma. Estos elementos se clasifican en dos grupos, dependiendo de si están flanqueados por regiones largas (*Long Terminal Repeats*, LTR) o si carecen de ellas (non-LTR). Se han descrito dos grupos de retroelementos sin LTR: los LINE o L1 (*Long Interspersed Nucleotide Elements*) que poseen capacidad de codificación y los SINE (*Short Interspersed Nucleotide Elements*) que no codifican proteínas. Todos están flanqueados por secuencias de duplicación directa (*Target Site duplication*, TSD) y portan en su extremo 3' una cola de poli-A.

Si bien durante mucho tiempo se ha calificado a estos elementos como egoístas, ADN basura y ADN parásitos, hoy se sabe que juegan un papel muy importante en la evolución de los genes y de los genomas de un amplio número de organismos. Por ejemplo, estos elementos de ADN dinámicos moldean los genomas que los alojan al generar reorganizaciones, crear y destruir genes y modificar patrones de expresión génica que resultan útiles para el huésped. Por ello, a este proceso se le ha calificado como “domesticación molecular” de los elementos móviles.

El L1Tc es el elemento sin LTR más abundante en el genoma de *Trypanosoma cruzi* (1), un protozoo parásito que causa la enfermedad de Chagas, la cual afecta a 10 millones de personas en el mundo. El L1Tc es activamente transcrito como mensajero poliadenilo (2) y codifica las proteínas que están involucradas en el proceso de movilización del elemento denominado *Target-Primed Reverse Transcription* (TPRT) (3). Así, L1Tc codifica proteínas con actividad de endonucleasa de tipo AP (4,5), 3'fosfatasa y 3' fosfodiesterasa (6), transcriptasa inversa (7), RNasa H (8) y chaperona de ácidos nucleicos (9). Todas estas funciones

de L1Tc le permiten, al no requerir funciones del huésped que lo alberga, ser un elemento autónomo desde el punto de vista de su movilización.

Los primeros 77 nucleótidos de L1Tc (Pr77) corresponden con un promotor interno que genera transcritos abundantes y traducibles (10), aparentemente mediados por la ARN polimerasa II. Los transcritos derivados de Pr77 no son procesados por *trans-splicing* y comienzan en el nucleótido +1 del elemento. Los Pr77 están también presentes en otros retroelementos no autónomos de *T. cruzi*, como NARTc, en retroelementos sin LTR de *Trypanosoma brucei*, como *ingi* y RIME, en los elementos cortos degenerados de diferentes especies de *Leishmania* y asociado a secuencias no relacionadas con retroelementos en diferentes posiciones del genoma de *T. brucei* (10).

En el extremo amino del elemento y en fase con la poliproteína que L1Tc codifica, se ha identificado una secuencia homóloga a la secuencia autocatalítica 2A descrita en picornavirus y otros organismos virales. Esta secuencia, denominada 2AL1Tc, se encuentra presente en todas las copias de L1Tc que posee *T. cruzi* localizada a 60 nucleótidos “corriente abajo” de Pr77. Las secuencias autocatalíticas 2A controlan la traducción *in situ*, de tal modo que, cuando la secuencia 2A es traducida, ejerce autocatálisis escindiendo el polipéptido que codificaba la región que le precedía y permitiendo que el ribosoma continúe la traducción de aquella que se encuentra “corriente abajo”, incluso en ausencia de un codón iniciador. La secuencia 2A de *T. cruzi* ha mostrado ser funcional *in vitro* e *in vivo*, y controla la abundancia de los productos de traducción del elemento L1Tc, los cuales constituyen la maquinaria enzimática necesaria para la movilidad del mismo (11).

La 2AL1Tc no sólo se encuentra presente en el L1Tc de *T. cruzi*, sino que también está presente en el extremo N-terminal de elementos similares

Dominios funcionales en L1Tc

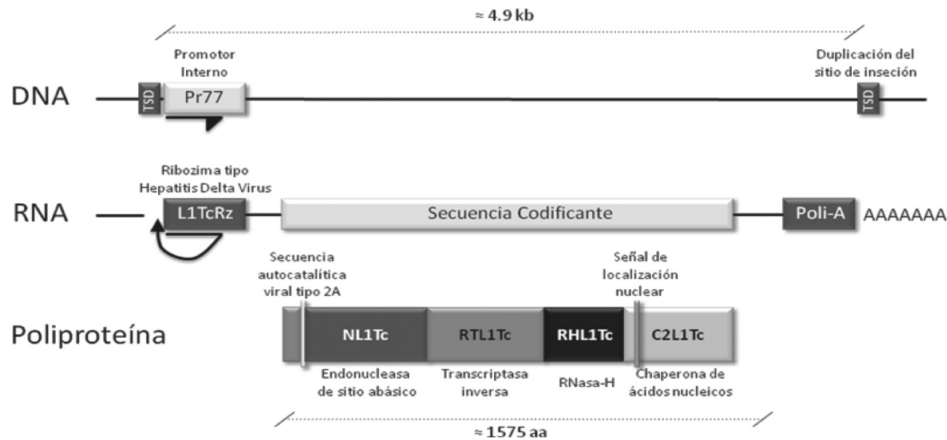


Figura 1. El que los primeros 77 nucleótidos del extremo 5' de L1Tc tengan actividad promotora y exista una actividad de ribozima en el extremo 5' del ARNm de L1Tc, hacen que L1Tc sea el primer retrotransposón en el que se describe una función doble promotor-ribozima a nivel del ADN y ARN. Esta función de ribozima podría permitir al retroelemento escindirse de transcritos policistrónicos con los que pudiera ser cotranscrito.

a los LINE en otros tripanosomátidos, como *T. congolense*, *T. b. gambiense* y *T. vivax*. La amplia distribución del consenso de la secuencia autocatalítica 2A indica que su adquisición ocurrió en un ancestro común y antes de que divergieran las especies (11).

Estos resultados muestran un ejemplo de transmisión horizontal de información génica, en la que un retroelemento aumenta su grado de complejidad, probablemente para mejorar el carácter autónomo que posee. El largo periodo de coexistencia de *T. cruzi* con sus huéspedes vertebrados e invertebrados, junto a las ventajas que la presencia de una secuencia autocatalítica pueda aportar al control de la traducción de su maquinaria enzimática, puede explicar el hecho de que una secuencia viral funcional forme parte de un retroelemento de un eucariota.

La 2A no es la única secuencia viral detectada en L1Tc. Recientemente, y mediante búsqueda manual de estructuras particulares, hemos identificado en L1Tc una conformación parcialmente compatible con la estructura adoptada por las tres "ribozimas" que, con homología a la del virus de la hepatitis D y, hasta la fecha de hoy, han demostrado experimentalmente ser activas. En esta conformación están involucrados los primeros 77 nucleótidos del extremo 5' del ARN de L1Tc. Los ensayos de transcripción *in vitro* han demostrado que la secuencia de ARN de 77 nucleótidos tiene actividad de corte de cotranscripción *in vitro* (12).

Los ensayos de corte cotranscripción llevados a cabo con las secuencias del extremo 5' del ARNm de L1Tc de diferente longitud (+47, +59, +70 y +77), muestran que la mayor actividad autocatalítica de L1TcRz reside en los primeros 77 nucleótidos de L1Tc.

Esta actividad catalítica, denominada L1TcRz, se confirmó cuando la fracción de transcrito sin cortar escindido de gel, se "autocortaba" en una forma dependiente del magnesio. El análisis de la cinética de corte mostró que esta función es muy rápida, dado que una alta proporción del corte ocurría en los primeros dos minutos de transcripción. El punto de corte de la ribozima, determinado mediante ensayos de extensión de iniciador, corresponde con el nucleótido +1 del elemento. La naturaleza del extremo 5' del producto 3' de corte de L1TcRz es hidroxilo.

Los nucleótidos que se localizan "corriente arriba" de L1Tc en diferentes copias de L1Tc en el genoma de *T. cruzi*, así como aquellos que se encuentran tras los 77 nucleótidos de L1Tc, influyen negativamente la actividad de "ribozima" de L1Tc. Este hecho parece deberse a que estas secuencias permiten adoptar conformaciones diferentes a la requerida para la función de "ribozima" de tipo HDV y, por lo tanto, impedir la formación de esta.

En cambio, las secuencias que se localizan tras Pr77 en el retrotransposón no autónomo NARTc, no disminuyen la actividad de "ribozima" L1TcRz. Mediante la búsqueda de secuencias homólogas

a Pr77 que adopten estructuras compatibles con la función de “ribozima” de tipo HDV, se han identificado en *Haliangium ochraceum*, *T. brucei* y *Penicillium chrysogenum* (12 y datos de laboratorio).

Referencias

- Martín F, Marañón C, Olivares M, Alonso C, López MC. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: Homology of the first ORF with the ape family of DNA repair enzymes. *J Mol Biol* 1995;247:49-59.
- Olivares M, Thomas MC, López-Barajas A, Requena JM, García-Pérez JL, Angel S, *et al.* Genomic clustering of the *Trypanosoma cruzi* non-long terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis*. 2000;21:2973-82.
- Thomas MC, Macías F, Alonso C, López MC. The biology and evolution of transposable elements in parasites. *Trends in Parasitology*. 2010;26:350-62.
- Olivares M, Alonso C, López MC. The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. *J Biol Chem*. 1997;272:25224-8.
- Olivares M, García-Pérez JL, Briones P, López MC, Thomas MC. The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. *Biochem Biophys Acta*. 2003;1626:25-32.
- Olivares M, Thomas MC, Alonso C, López MC. The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3'-phosphatase and 3'-phosphodiesterase enzymatic activities. *J Biol Chem*. 1999;274:23883-6.
- García-Pérez JL, González CI, Thomas MC, Olivares M, López MC. Reverse transcriptase activity in a protein encoded by the non-LTR retrotransposon from *T. cruzi*. *Cell and Mol Life Sci*. 2003;60:2692-701.
- Olivares M, García-Pérez JL, Thomas MC, Heras S, López MC. The non-LTR retrotransposon L1Tc from *T. cruzi* codes for a protein with RNaseH activity. *J Biol Chem*. 2002;277:28025-30.
- Heras S, López MC, García-Pérez JL, Martín S, Thomas MC. The C-terminal domain of the L1Tc non-LTR retrotransposon from *Trypanosoma cruzi* contains two C2H2 zinc-finger motifs endowed with nucleic acid chaperone activity. *Mol Cell Biol*. 2005;25:9209-20.
- Heras RS, López MC, Olivares M, Thomas MC. The L1Tc non-LTR retrotransposon of *Trypanosoma cruzi* contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Res*. 2007;35:2199-214.
- Heras S, Thomas MC, García M, de Felipe P, García-Pérez JL, Ryan M, *et al.* L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode. *Cell Mol Life Sci*. 2006;63:1449-60.
- Sánchez-Luque FJ, López MC, Macías F, Alonso C, Thomas MC. Identification of an HDV-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from *Trypanosoma cruzi*. *Nucleic Acids Res*. 2011, en prensa.



Aplicación de ARNi de *Acanthamoeba*

Jacob Lorenzo, Basilio Valladares, Enrique Martínez-Carretero

Instituto de Enfermedades Tropicales y Salud Pública de Canarias, España

Las cepas patógenas pertenecientes al género *Acanthamoeba* son causantes de infecciones graves, tales como una encefalitis fatal y una queratitis amebiana muy dolorosa que puede progresar a ceguera.

El tratamiento de estas infecciones es, generalmente, empírico y, a menudo, problemático, debido a la existencia de una fase quística en el ciclo de vida de estas amebas que es muy resistente a agresiones físicas y químicas.

Nuestro laboratorio inició estudios de silenciamiento génico basado en ARN de interferencia, en el 2005. Durante los mismos, se desarrollaron y validaron ARNsi frente al dominio catalítico de proteasas de serina de extracelulares y la fosforilasa

de glucógeno del género *Acanthamoeba* enfocados a su empleo terapéutico en el futuro. El silenciamiento de las proteasas causa que *Acanthamoebae* no pueda degradar las células humanas corneales, mientras que el silenciamiento de la fosforilasa de glucógeno que las amebas fueran incapaces de formar quistes maduros.

Recientemente, y tras optimizar el diseño de estos ARNsi, así como la concentración de los mismos para evitar problemas de toxicidad, varias cepas clínicas de *Acanthamoebae* fueron tratadas con una combinación de ambos ARNsi y el efecto en las células fue evaluado mediante microscopía. Esto resultó en que las amebas fueron eliminadas por este tratamiento en menos de 48 horas.

A fin de comprobar los posibles efectos tóxicos de la combinación de ARNsi, tres líneas de células eucariotas (HeLa, macrófagos de ratón y de células de osteosarcoma) fueron tratadas con las mismas moléculas y la citotoxicidad se examinó mediante la medición de la liberación de deshidrogenasa de lactato, resultando que la misma era moderada en la mayoría de los casos.

En la actualidad, estamos trabajando en el diseño de nuevos polímeros de transfección de ARNsi para eliminar la toxicidad indeseada.

El uso futuro de esta combinación de ARNsi se propone como un posible abordaje terapéutico frente a cepas patógenas del género *Acanthamoeba*.

Referencias

1. Lorenzo-Morales J, Martín-Navarro CM, López-Arencibia A, Santana-Morales MA, Afonso-Lehmann

RN, Maciver SK, *et al.* Therapeutic potential of a combination of two gene-specific small interfering RNAs against clinical strains of *Acanthamoeba*. *Antimicrob Agents Chemother.* 2010;54:5151-5.

2. Lorenzo-Morales J, Kliescikova J, Martínez-Carretero E, De Pablos LM, Profotova B, Nohynkova E, *et al.* Glycogen phosphorylase in *Acanthamoeba* spp.: Determining the role of the enzyme during the encystment process using RNA interference. *Eukaryot Cell.* 2008;7509-17.
3. Lorenzo-Morales J, Morcillo-Laiz R, Martín-Navarro CM, López-Vélez R, López-Arencibia A, Arnalich-Montiel F, *et al.* *Acanthamoeba* keratitis due to genotype T11 in a rigid gas permeable contact lens wearer in Spain. *Cont Lens Anterior Eye.* 2011;34:83-6.



Genes *HSP70* de *Leishmania braziliensis*, un modelo para el estudio de regulación posterior a la transcripción

Cesar Ramírez¹, José María Requena², Concepción J. Puerta¹

¹ Laboratorio de Parasitología Molecular, Departamento de Microbiología, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá, D. C., Colombia

² Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Universidad Autónoma de Madrid, Madrid, España

Los tripanosomátidos, incluyendo las leishmanias, constituyen un modelo biológico muy interesante de regulación de la expresión génica. En estos organismos, los genes están arreglados en unidades de transcripción policistrónica independientemente de su función, a diferencia de lo que sucede en las bacterias. Además, el ARNm antes de ser traducido debe ser procesado a monocistrón, evento que ocurre mediante la adición de una secuencia líder o miniexón en el extremo 5' y de la cola de poliadeninas en el extremo 3'. En consecuencia, la regulación génica en estos parásitos opera básicamente a nivel de postranscripción e involucra el procesamiento, transporte, localización y estabilidad del ARNm, así como la eficiencia de la traducción, las modificaciones posteriores a la traducción y la vida media de las proteínas.

Distintos estudios han evidenciado que parte de las funciones antes mencionadas se llevan a cabo gracias a la interacción entre las proteínas de unión al ARNm y las regiones no traducidas de éstos, especialmente la 3' UTR, en donde se localizan la mayoría de las secuencias o motivos de reconocimiento; en esta era posgenómica es de gran importancia la identificación tanto de las

proteínas como de las secuencias blancas con las cuales interactúan (1).

Por otra parte, las proteínas de choque térmico (HSP) y entre ellas las de la familia de 70 kDa (HSP70), se encuentran comprometidas en un conjunto de procesos celulares necesarios para el funcionamiento normal de la célula y la respuesta al estrés ocasionado por choque térmico, entre otros (2). En los tripanosomátidos el estrés térmico se produce al pasar del insecto vector (22 a 28 °C) al huésped mamífero (35 a 37 °C). De hecho, los estudios de proteómica reportan que la transición de promastigote a amastigote se acompaña de un incremento de la HSP70 citoplásmica (3). Aun cuando no es claro si estas proteínas intervienen directamente en el proceso de diferenciación, las HSP70 son de vital importancia para el establecimiento de la infección y desarrollo del parásito en el nuevo ambiente celular. Además, la HSP70, particularmente en *Leishmania*, tiene un efecto antiapoptosis que protege al parásito frente al tratamiento con antimoniales (4).

En *Leishmania* los genes codificantes para las HSP70 típicas citoplásmicas (HSP70A) han sido caracterizados en varias especies, encontrándose

que en la mayoría estos genes se codifican por dos tipos de unidad. Así, el genoma de *Leishmania infantum* contiene 6 genes *HSP70* de 3,8 kb organizados en tándem cabeza-cola, los cuales se localizan en un único *locus* (5). Estos genes son muy conservados en su secuencia, con excepción de la región 3' UTR del sexto gen localizado en el extremo 3' del tándem. Para mayor claridad, los primeros cinco genes del tándem se han denominado *HSP70-I* y, el sexto, *HSP70-II*. Notablemente, se conoce que mientras los transcritos *HSP70-I* se acumulan en respuesta al estrés térmico y se traducen tanto a 26 °C como a 37 °C, los ARNm derivados de los genes *HSP70-II* no se acumulan bajo tratamiento térmico, pero se traducen preferencialmente a 37 °C (6). Es llamativo que los mutantes del parásito carentes del gen *HSP70-II* presenten alteraciones en su fenotipo, crecimiento y virulencia (7).

A diferencia de las especies estudiadas del subgénero *Leishmania*, en *Leishmania Viannia braziliensis* hasta ahora sólo se había reportado la presencia del gen *HSP70-I*. Los análisis de *Southern blot* que usan la región 3' UTR de los genes *HSP70-II* de *L. infantum* como sonda, habían llevado a sugerir la ausencia de estos genes en *L. braziliensis*. Sin embargo, los ensayos de *Northern blot* que usan la región codificante de los genes *HSP70* como sonda, evidencian la presencia de dos transcritos diferentes de tamaños similares. Estos resultados llevaron a plantear que la región 3' UTR de los genes *HSP70-II* podría ser bastante divergente de la región respectiva en especies reportadas del subgénero *Leishmania* (8).

Teniendo en cuenta lo anterior, en este trabajo se estudió la organización genómica y transcripción de los genes *HSP70A* de *L. braziliensis*, así como la determinación de sus regiones 5' y 3' UTR. En primer lugar, los resultados de los análisis de *Southern blot* de ADN genómico digerido parcial o totalmente con diversas endonucleasas y sometido a hibridación con la región intercodificante de los genes *HSP70* (IR-*HSP70*) del parásito, evidenciaron la existencia de, al menos, siete genes *HSP70A* organizados en tándem cabeza-cola.

Este hallazgo complementa la información anotada en el proyecto genoma, en el que se observa la presencia en la hebra antisentido del cromosoma 28, de un gen *HSP70* completo (GeneDB ID: LbrM28_v2.2990) y dos secuencias parciales, LbrM28_v2.2980 (codificante para la región amino-terminal) y LbrM28_v2.2970 (codificante para la región C-terminal), separadas por una brecha de longitud desconocida.

Asimismo, los resultados de la separación de los cromosomas mediante electroforesis en campo pulsado seguida de transferencia e hibridación con la sonda IR-*HSP70*, indican que los genes *HSP70A* se ubican en un cromosoma de 1,5 Mb, tamaño que concuerda con el estimado para el cromosoma 28. Tras ensayos de RT-PCR con un oligo d(T) para la síntesis del ADNc, seguido de amplificación con cebadores correspondientes a las secuencias del gen miniexón y a una secuencia corriente abajo del ATG iniciador del gen *HSP70* para obtener la región 5' UTR, y los cebadores correspondientes a una región "corriente arriba" del codón de parada del gen *HSP70* y el oligo d(T), para el caso de la región 3' UTR, se determinaron las secuencias UTR de los genes *HSP70A* de *L. braziliensis*.

De especial interés es que se encontraron dos clases de regiones 3' UTR, correspondientes a los genes *HSP70-I* y *II*, en el que la 3' UTR-I de 936 pb corresponde a la entrada LbrM28_v2.2990 y, la 3' UTR-II de 932 pb, a la entrada LbrM28_v2.2970. Los ensayos de *Southern blot* con la región 3' UTR-II como sonda, sugieren la presencia de un solo gen *HSP70-II*, precedido por seis copias de los genes *HSP70-I*. Esta organización es similar a la reportada en las otras especies de *Leishmania*, tanto en el número como en el orden de los genes. Esta estricta conservación de la organización genómica en ambos subgéneros del parásito, presupone un papel o impacto funcional de la misma.

Por otra parte, se compararon las secuencias entre las regiones UTR de *L. braziliensis* con las secuencias equivalentes en los proyectos genomas de *L. major* y *L. infantum*. Mientras que las regiones 5' UTR y 3' UTR-I de *L. braziliensis* tienen identidades entre 71 y 81 % con su contraparte en *L. major* y *L. infantum*, la región 3' UTR-II tan sólo conserva entre 55 y 57 % de identidad, presenta algunos motivos exclusivos de especie y carece de algunas secuencias comunes a las otras especies. Al realizar los ensayos de *Northern blot* con la sonda IR-*HSP70*, se observó la presencia de dos fragmentos de hibridación de tamaño similar. El uso de sondas específicas para las regiones 3' UTR I y II permitió elucidar que los transcritos de mayor tamaño corresponden al ARNm del gen *HSP70-II* y, los de menor tamaño, a los de los genes *HSP70-I*. De especial interés es que ninguno de los transcritos mostró acumulación dependiente de la temperatura.

Sin embargo, las diferencias encontradas en la secuencia y el contenido de GC entre las regiones 3' UTR I y II, sugieren que la expresión de estos transcritos se regula de forma diferencial. Así,

como una primera aproximación al estudio de los factores proteicos que pudieran estar actuando en *trans* con las regiones UTR de los genes *HSP70A* de *L. braziliensis*, se realizaron ensayos de *pull-down* usando como cebo las regiones UTR de los genes *HSP70* unidas a perlas magnéticas en presencia de extractos proteicos del parásito (obtenidos en condiciones normales y de choque térmico), con posterior identificación mediante análisis de electroforesis en dos dimensiones y MALDI TOF/TOF, usando las regiones UTR de los genes codificantes para la α -tubulina como control no relevante y perlas sin ARN como control negativo.

Tras los análisis efectuados, se identificaron las proteínas subunidad 1 del factor de replicación A (Rpa1) y glicerol-3-fosfatodeshidrogenasa (G3PDH), como de interacción exclusiva con la región 3' UTR-II de los genes *HSP70* de *L. braziliensis* y la β -tubulina como de interacción exclusiva con la 5' UTR de estos mismos genes del parásito. Asociada a la 3' UTR-II de la *HSP70* a 26 °C y a la 3' UTR de la α -tubulina a ambas temperaturas, se identificó la hexocinasa. Finalmente, asociada a ambas regiones 3' UTR de la *HSP70* (a todas las temperaturas) y la α -tubulina (sólo a 26 °C), se encontró la proteína Rpb38 (Tc38).

Referencias

1. Requena JM. Lights and shadows on gene organization and regulation of gene expression in *Leishmania*. *Front Biosci*. 2011;16:2069-85.
2. Requena JM, Soto M, Quijada L, Alonso C. Antigenicidad e inmunogenicidad de las proteínas de choque térmico. *Ars Pharmacia*. 1997;38:191-208.
3. McNicoll F, Drummelsmith J, Müller M, Madore E, Boilard N, Ouellette M, *et al*. A combined proteomic and transcriptomic approach to the study of stage differentiation in *Leishmania infantum*. *Proteomics*. 2006;6:3567-81.
4. Raina P, Kaur S. Chronic heat-shock treatment driven differentiation induces apoptosis in *Leishmania donovani*. *Mol Cell Biochem*. 2006;289:83-90.
5. Quijada L, Soto M, Alonso C, Requena JM. Analysis of post-transcriptional regulation operating on transcription products of the tandemly linked *Leishmania infantum* hsp70 genes. *J Biol Chem*. 1997;272:4493-9.
6. Folgueira C, Quijada L, Soto M, Abanades DR, Alonso C, Requena JM. The translational efficiencies of the two *Leishmania infantum* HSP70 mRNAs, differing in their 3'-untranslated regions, are affected by shifts in the temperature of growth through different mechanisms. *J Biol Chem*. 2005;280:35172-83.
7. Folgueira C, Carrión J, Moreno J, Saugar JM, Cañavate C, Requena JM. Effects of the disruption of the HSP70-II gene on the growth, morphology, and virulence of *Leishmania infantum* promastigotes. *Int Microbiol*. 2008;11:81-9.
8. Folgueira C, Cañavate C, Chicharro C, Requena JM. Genomic organization and expression of the HSP70 locus in New and Old World *Leishmania* species. *Parasitology*. 2006;134:1-9.



Using single molecule sequencing to decipher the non-coding transcriptome

G. St. Laurent^{1,2,3}, M. Tackett³, D. Shtokalo³, S. Nechkin³, T. Eremina³, J. Finch³, Y. Vyatkin³, Y. Saava¹, R. Reenan¹, T. McCaffrey⁴, P. Kapranov³

¹ Brown University, Providence, RI, USA

² Universidad de Antioquia, Medellín, Colombia

³ The St. Laurent Institute, Cambridge, MA, USA

⁴ The Georges Washington University Medical Center, Washington, D.C., USA

Surprises revealed by the sequencing of the human genome (1,2) have fueled a decade long debate about the existence and functionality of ncRNA in complex organisms. Based on the vast differences in complexity between humans and model organisms such as the worm and the fly, most observers had predicted that humans would possess far more genes than these lower organisms, with estimates often in the range of 100,000 genes. Instead fewer than 25,000 human genes exist, contained within a tiny fraction of 2% of the overall genome, leaving the spectacular

complexity of the human organism to be explained by some yet to be discovered mechanisms.

The number of coding genes, and their even smaller footprint in the genome, has driven speculation that the non-coding genome could play a decisive role in mammalian complexity. In fact, the Mattick group made a striking observation that the complexity of organisms correlates monotonically with the percentage of their genomes occupied by non-coding DNA, for all organisms sequenced to date. The discovery of pervasive transcription (3) and large numbers of non-coding RNAs [collectively

known as “dark matter” RNAs (4)] have opened the possibility of widespread function of ncRNAs in human physiology (5).

The transcriptome’s breathtaking complexity parallels the emerging theories that it functions as a multi-layer, dynamic, computational matrix (6). Many non-coding RNAs interact with proteins (7), DNA (8), and other RNAs (9), to acquire, process, and distribute high density information throughout the cell. Although such evidence has accumulated to support the hypothesis that ncRNA molecules fulfill the role of “informational currency” in the cell (10), other investigators have concluded that ncRNAs largely represent transcriptional noise, random degradation products, and artifacts of previous RNA detection methods (11). The evolving controversy of the existence, function and significance of dark matter transcription now occupies a central position in contemporary Biology.

Consistent with the potential importance of dark matter transcripts, we have developed single molecule RNA-seq methodologies to accurately capture RNAs transcribed from non-exonic regions, and achieve an unbiased, reproducible, and quantitative analysis of exonic and non-exonic RNA species. Our approach adapts the Helicos single molecule sequencing technology to generate large numbers of unbiased sequence reads from all informative RNA classes in the transcriptome. These techniques permit measurements of gene expression, ncRNA expression, alternative splicing, transcriptional start sites (CAGE), RNA editing, and transcription termination, with unparalleled accuracy and better reproducibility than can be achieved on other platforms (12,13). Here we present examples of these techniques in several model organisms, for the purpose of elucidating the mechanisms of non-coding transcriptome processing and function.

The study of RNA editing exemplifies the technical challenges inherent in mapping processing events in the transcriptome. Many confounding events can cloud the accurate identification of ADAR editing sites in the transcriptome, resulting in high percentages of false positives and false negatives in transcriptome wide datasets. We have created a 3 billion aligned read dataset in *Drosophila*, designed to use the superior depth and resolution provided by these techniques to achieve the first comprehensive, genome wide map of ADAR editing for a complex organism. A review of our results to date demonstrates editing sites in exons, introns, and intergenic regions, with distinct distributions and conservation patterns.

In human cells, we have recently demonstrated that the majority of the mass of the cellular pool of RNA comes from “dark matter” (non-exonic) regions of the genome (14). Extending these studies, we have obtained 2 billion aligned reads from cellular RNA of mouse lung stimulated with LPS for six different time points (n=7). We have chosen the inflammation cascade as the model system since it sets in motion a series of intricate responses to the perturbation of invading pathogens, with many interlocking feedback loops involved in the precise orchestration of onset and resolution. We identified thousands of transcriptional responses to inflammation, including splicing changes, editing changes, and consistent expression changes in the non-coding “dark matter” portions of the genome. We asked a question whether this mass of “dark matter” RNA harbors within itself signals necessary to regulate exons of protein-coding transcripts in trans. Interestingly, we found a reciprocal connection: whereas most of the exonic transcripts increase in expression during early time points of the inflammation cascade, thousands of intronic regions showed decreases in expression. To probe a possible connection between up-regulated exons and down-regulated intronic transcripts, our bioinformatics strategy searched for short sequence motifs (seeds) in such intronic transcripts, and compared the distribution of these seeds in up-regulated, down-regulated, and unchanged exons across the entire inflammation transcriptome. The results demonstrate a statistically significant increase in incidence of motifs shared by down-regulated introns and up-regulated exons. These and other results suggest that intronic RNAs contain within them sequences that could regulate exons of other transcripts.

References

1. Lander ES, *et al.* Science. 2001;291:1304-51.
2. Venter JC, *et al.* Nature. 2001;409:860-921.
3. Carninci P, *et al.* Science. 2005;309:1559-63.
4. Johnson JM, *et al.* Trends Genet. 2005;21:93-102.
5. St. Laurent G, *et al.* Trends Neurol. 2007;30:612-21.
6. St. Laurent G, *et al.* Neurosci Lett. 2009;466:89-98.
7. Khalil AM, *et al.* Proc Natl Acad Sci USA. 2009;106:11667-72.
8. Mattick JS, *et al.* Bioessays. 2009;31:51-9.
9. Flynt AS, Lai EC. Nat Rev Genet. 2008;9:831-42.
10. Mattick JS. Bioessays. 2003;25:930-9.
11. van Bakel H, *et al.* PLOS. 2010;8:e1000371.
12. Raz T, *et al.* PLoS One. 2011;6:e19287.
13. Mamanova L, *et al.* Nat Methods. 2010;7:130-2.
14. Kapranov P, *et al.* BMC Biology. 2010;8:149.