



UNIVERSIDAD
DE GRANADA

ipbln



Instituto de Parasitología y Biomedicina López-Neyra (IPBLN)
Consejo Superior de Investigaciones Científicas (CSIC)

Caracterización del papel de la proteína RBP33 en la regulación de la transcripción no productiva en *Trypanosoma brucei*

Universidad de Granada
Programa de doctorado en Bioquímica y Biología molecular

Tesis doctoral

CLAUDIA GÓMEZ LIÑÁN

Granada, 2023

Editor: Universidad de Granada. Tesis Doctorales
Autor: Claudia Gómez Liñan
ISBN: 978-84-1117-853-2
URI : <https://hdl.handle.net/10481/82025>



**UNIVERSIDAD
DE GRANADA**

**Caracterización del papel de la proteína
RBP33 en la regulación de la transcripción
no productiva en *Trypanosoma brucei***

Memoria presentada por la graduada

CLAUDIA GÓMEZ LIÑÁN

para optar al título de

Doctora por la Universidad de Granada

Trabajo realizado bajo la dirección de:

Dr. Antonio M. Estévez García

Investigador Científico Titular

del Consejo Superior de Investigaciones Científicas (CSIC)

Granada, 2023

ÍNDICE

RESUMEN	5
ABREVIATURAS	9
INTRODUCCIÓN	15
1. <i>Trypanosoma brucei</i>	17
1.1. Características morfológicas	17
1.2. Enfermedad del sueño. Etiología, epidemiología y tratamiento	19
1.3. Ciclo de vida	21
1.4. Variación antigénica	23
2. Regulación de la expresión génica en <i>T. brucei</i>	27
2.1. Organización genómica	27
2.2. Regulación de la transcripción	30
2.2.1. Inicio de la transcripción	30
2.2.2. Terminación de la transcripción	33
2.3. Procesamiento, transporte y degradación del RNA mensajero ...	37
2.3.1. Procesamiento: <i>trans-splicing</i> y poliadenilación	37
2.3.2. Transporte del RNA del núcleo al citosol	38
2.3.3. Degradación del RNA	39
3. Transcripción generalizada	41
3.1. Factores implicados en el control de la transcripción generalizada ..	41
3.2. Posibles consecuencias de la transcripción generalizada	44
4. La proteína de unión a RNA RBP33	49
OBJETIVOS	53

MATERIALES Y MÉTODOS	57
1. Materiales	59
1.1. Líneas celulares de <i>Trypanosoma brucei</i>	59
1.2. Oligonucleótidos	59
1.3. Plásmidos	61
1.4. Anticuerpos	63
1.5. Medios y tampones utilizados	63
1.6. Antibióticos	65
1.7. Programas informáticos y bases de datos	65
1.7.1. Programas informáticos	65
1.7.2. Bases de datos	66
2. Métodos	66
2.1. Cultivo de <i>Trypanosoma brucei</i>	66
2.2. Transfección estable de tripanosomas mediante nucleofección ...	67
2.2.1. Tripanosomas sanguíneos	67
2.2.2. Tripanosomas procíclicos	67
2.3. Congelación y descongelación de tripanosomas	68
2.4. Ribointerferencia (RNAi)	68
2.5. Clonaje de vectores	68
2.6. Aislamiento de DNA genómico en <i>T. brucei</i>	68
2.7. Preparación de células competentes de <i>E. coli</i>	69
2.8. Aislamiento de DNA plasmídico de <i>E. coli</i>	69
2.9. Transformación de <i>E. coli</i> por choque térmico	69
2.10. Amplificación de DNA por PCR	69

2.11. Precipitación de DNA	70
2.12. Extracción de RNA de cultivos de <i>T. brucei</i>	70
2.13. RT-PCR cuantitativa	70
2.14. Secuenciación masiva de RNA (RNA-seq)	71
2.15. Citometría de flujo	73
2.16. Ensayo de cromatina accesible por transposasa con secuenciación masiva (ATAC-seq)	73
2.17. Ensayos de vida media	75
2.18. Preparación de lisados de proteínas	76
2.19. Inmunoblot	76
2.20. Purificación por afinidad en tándem (TAP)	77

RESULTADOS79

1. La proteína de <i>Trypanosoma brucei</i> RBP33 es específica de tripanosomátidos	81
2. El silenciamiento de la expresión de la proteína RBP33 conduce a un marcado incremento en los niveles de transcritos antisentido	84
3. Los transcritos regulados por RBP33 derivan principalmente de regiones genómicas silenciadas y de la transcripción antisentido de regiones centrales	87
4. Los transcritos que codifican las VSGs metacíclicas se sobrexpresan en ausencia de RBP33	94
5. Los transcritos regulados por RBP33 están correctamente procesados	96
6. Expresión dependiente de RBP33 de un gen reportero insertado en orientación antisentido dentro de una región genómica silenciada	97
7. Análisis de la accesibilidad de la cromatina en ausencia de RBP33 ..	100

8. El silenciamiento de la expresión de RBP33 por RNAi da como resultado la estabilización de los transcritos que se regulan	105
9. Interacción de RBP33 con otras proteínas	108
10. El exosoma de <i>T. brucei</i> está involucrado en la degradación de los transcritos derivados de cSSRs	109
DISCUSIÓN	113
CONCLUSIONES	135
ANEXOS	139
ANEXO I: FIGURAS SUPLEMENTARIAS	141
ANEXO II: PUBLICACIÓN CIENTÍFICA	145
BIBLIOGRAFÍA	161

RESUMEN

RESUMEN

Durante la última década, gracias a los avances en las tecnologías de secuenciación masiva que han permitido un análisis más profundo de los transcriptomas de varios organismos modelo, se ha determinado que los genomas se transcriben de manera generalizada. Este fenómeno da lugar a una gran cantidad de transcritos de RNA no canónicos y en su mayoría antisentido que se producen por la RNA polimerasa II a partir de casi cualquier contexto genómico. Estas especies de RNA son degradadas por mecanismos de vigilancia que actúan principalmente a nivel nuclear. Sin embargo, la información acerca de las proteínas que controlan el destino de estos transcritos no productivos es muy escasa.

Trypanosoma brucei es un parásito extracelular que pertenece a la clase Kinetoplastea y es el agente causante de la tripanosomiasis humana africana (HAT), también conocida como la enfermedad del sueño. Los tripanosomas son eucariotas unicelulares cuya transcripción por parte de la RNA polimerasa II es constitutiva y tanto el inicio como la terminación de la transcripción ocurren en un número limitado de sitios por cromosoma. En la actualidad se desconoce si existe una transcripción generalizada en organismos donde la regulación transcripcional está prácticamente ausente o qué factores podrían estar implicados en este proceso.

Trabajos publicados previamente por nuestro grupo indican que la proteína de unión a RNA RBP33 se une a transcritos procedentes de zonas que normalmente se encuentran silenciadas, lo que sugiere que existe un vínculo entre RBP33 y la terminación de la transcripción en *T. brucei*. Por este motivo, durante esta tesis doctoral se han estudiado distintos aspectos de la regulación de la expresión génica para determinar el papel que la proteína de unión a RNA RBP33 tiene en la regulación de la abundancia de estos RNAs no canónicos en *T. brucei*.

El análisis detallado del transcriptoma de los tripanosomas muestra que el silenciamiento de RBP33 por ribointerferencia provoca una sobreexpresión de aproximadamente el 40% de todos los genes anotados

en el genoma, con una marcada acumulación de transcritos sentido y antisentido que proceden de regiones que no se expresan en condiciones fisiológicas. Además, los resultados sugieren que los transcritos antisentido son potencialmente traducibles, ya que la inserción de un transgén *eYFP* en un locus que no se expresa en condiciones normales, proporciona una señal de fluorescencia detectable en ausencia de RBP33. Por otro lado, la pérdida de la proteína no produce un aumento apreciable en la accesibilidad de la cromatina, por lo que RBP33 no parece actuar a este nivel. Finalmente, los datos muestran que los transcritos regulados por RBP33 son significativamente más estables en tripanosomas cuando se silencia la expresión de RBP33 y que el complejo exosoma es responsable de su degradación.

Estas observaciones nos permiten plantear un modelo donde RBP33 actúa guiando a los transcritos aberrantes para su degradación vía exosoma, amortiguando en consecuencia la transcripción no productiva en los tripanosomas, aunque el mecanismo exacto queda aún por determinar.

ABREVIATURAS

ABREVIATURAS

ATAC-seq. Ensayo de cromatina accesible por transposasa usando secuenciación masiva o *Assay for Transposase-Accessible Chromatin using sequencing*

BES. Sitios de expresión de VSGs de la forma sanguínea o *Bloodstream Expression Sites*

cSSR. Regiones de cambios de cadena convergentes o *convergent Strand Switch-Region*

CUTs. Transcritos crípticos inestables o *Cryptic Unstable Transcripts*

DNA. Ácido desoxirribonucleico o *DesoxirriboNucleic Acid*

dSSR. Regiones de cambios de cadena divergentes o *divergent Strand Switch-Region*

ESAGs. Genes asociados al sitio de expresión o *Expression Site-Associated Genes*

GPI. Glicosilfosfatidilinositol o *GlycosilPhosphate Inositol*

HAT. Tripanosomiasis africana humana o *Human African Trypanosomiasis*

kb. Kilobases

kDa. Kilodalton

lncRNAs. RNA largos no codificantes o *long non-coding RNAs*

MES. Sitios de expresión de VSGs de la forma metacíclica o *Metacyclic Expression Sites*

mRNA. RNA mensajero o *Messenger RNA*

ncRNAs. RNA no codificante o *non-coding RNAs*

NFR. Región libre de nucleosomas o *Nucleosome-Free Region*

NLS. Secuencia de localización nuclear o *Nuclear Localization Sequence*

NMD. *Nonsense-Mediated Decay*

NNS. Complejo Nrd1p-Nab3p-Sen1p o *Nrd1p-Nab3p-Sen1p Complex*

NSD. *Non-Stop Decay*

OMS. Organización Mundial de la Salud

ORF. Marcos abiertos de lectura u *Open Reading Frames*

PAS. Señal de poliadenilación o *Polyadenylation Signal*

pb. Pares de bases

PBS. Solución salina tamponada con fosfato o *Phosphate-Buffered Saline*

PCA. Análisis de componentes principales o *Principal Component Analysis*

PROMPT. *Promoter upstream Transcript*

PTU. Unidades de transcripción policistrónica o *Polycistronic Transcription Units*

RBP. Proteínas de unión a RNA o *RNA Binding Proteins*

RNA. Ácido ribonucleico o *RiboNucleic Acid*

RNAi. Ribointerferencia o *RNA interference*

RPM. Lecturas por millón de lecturas mapeadas o *Reads Per Million mapped reads*

RRM. Motivo de reconocimiento de RNA o *RNA Recognition Motif*

rRNA. RNA ribosomal o *Ribosomal RNA*

RT-PCR. Reverso transcripción y PCR o *Reverse Transcription PCR*

SAS. Sitios aceptores de *splicing* o *Splice-Acceptor Sites*

SDS-PAGE. Electroforesis desnaturalizante en gel de poliacrilamida o *Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis*

SL. *Spliced Leader*

SLiM. Motivo linear corto o *Short Linear Motif*

SLRNA. RNA del *Spliced Leader*

snoRNAs. RNA pequeños nucleolares o *Small Nucleolar RNAs*

snRNAs. RNA pequeños nucleares o *Small Nuclear RNAs*

SSR. Región de cambio de cadena o *Strand Switch-Region*

SUTs. Transcritos estables no anotados o *Stable Unannotated Transcripts*

TAP. Purificación por afinidad en tándem o *Tandem Affinity Purification*

tRNA. RNA de transferencia o *Transfer RNA*

TSS. Sitios de inicio de la transcripción o *Transcription Start Sites*

TTS. Sitios de terminación de la transcripción o *Transcription Termination Sites*

UTR. Región no traducida o *Untranslated Region*

VSG. Glicoproteína de superficie variante o *Variant Surface Glycoprotein*

XUTs. Transcritos inestables sensibles a Xrn1 o *Xrn1-sensitive Unstable Transcripts*

INTRODUCCIÓN

INTRODUCCIÓN

1. *Trypanosoma brucei*

1.1. Características morfológicas

Los kinetoplástidos son un tipo de protistas de vida libre o parásitos donde se encuentra la familia Trypanosomatidae, que engloba exclusivamente parásitos obligados y representa el grupo más diverso en cuanto al número de especies descritas y/o identificadas mediante técnicas moleculares (D'Avila-Levy et al., 2015; Lukeš et al., 2018). Esta familia pertenece a un linaje evolutivo muy divergente de eucariotas dentro de la clase Kinetoplastea. Los kinetoplástidos muestran aspectos biológicos inusuales y aparentemente únicos posiblemente como consecuencia de esta temprana separación (Ghedini et al., 2004). Por ejemplo, la regulación de la expresión génica en estos organismos es muy diferente a la de eucariotas modelo convencionales como levaduras, moscas de la fruta, nematodos y vertebrados, siendo todos ellos miembros del clado Opisthokonta (Clayton, 2019; Keeling & Burki, 2019). De hecho, en los tripanosomátidos se han descubierto por primera vez nuevos fenómenos biológicos relacionados con el mundo del RNA como el *trans-splicing* (Matthews et al., 1994), la edición de RNA (Benne, 1994) o el hecho de que la RNA polimerasa I transcriba genes que codifican proteínas (Günzl et al., 2003).

Los tripanosomátidos poseen una gama de huéspedes muy amplia: animales (principalmente insectos y vertebrados), plantas con flores e incluso ciliados (Podlipaev, 1990). Según el tipo de ciclo de vida, generalmente se subdividen en dos grupos no taxonómicos: monoxenos, que se desarrollan en un solo huésped, o heteroxenos, que alternan entre dos siendo uno de ellos el vector.

La mayoría de los géneros de tripanosomátidos son monoxenos y suelen parasitar dípteros y hemípteros (Podlipaev, 1990; Wallace, 1966).

INTRODUCCIÓN

Dentro de los heteroxenos encontramos géneros como *Leishmania* y *Trypanosoma*, que son importantes patógenos de vertebrados, incluidos los humanos. *Trypanosoma* es un género muy diverso donde se incluyen más del 60% de todas las especies descritas en la familia Trypanosomatidae. En concreto, *Trypanosoma brucei* es un parásito extracelular unicelular que se caracteriza por poseer un solo flagelo, que le confiere movilidad continua y una única mitocondria, cuya activación varía según sus necesidades (ver más abajo en el apartado 1.3.). Asociada al cuerpo basal del flagelo, posee una estructura denominada kinetoplasto que consiste en una red formada por DNA circular compacto multicopia y microtúbulos, donde se encuentra el DNA mitocondrial y que le da nombre al orden. El citoesqueleto de los tripanosomas está formado por una red de microtúbulos altamente polarizados que les confiere flexibilidad y facilita que el parásito pueda adoptar una gran variedad de formas durante su ciclo vital (Gull, 1999). Muchos de sus orgánulos y estructuras están presentes en una sola copia y se concentran entre el centro y el extremo posterior de la célula (Figura 1).

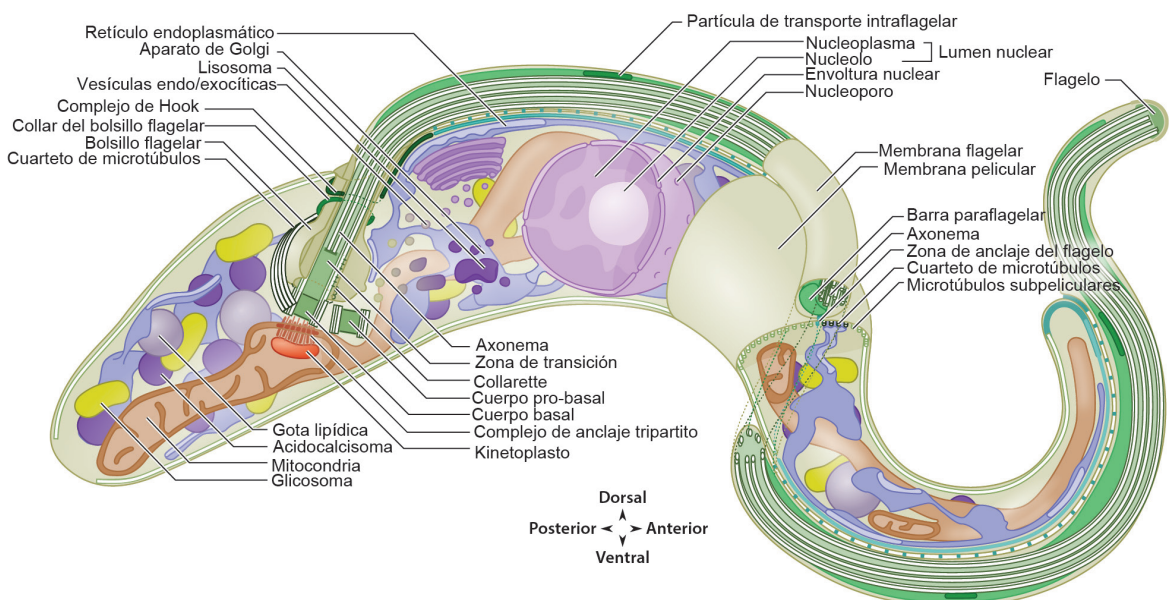


Figura 1. Dibujo esquemático de la morfología de la forma procíclica de *T. brucei*. Imagen adaptada de (Wheeler et al., 2019).

1.2. Enfermedad del sueño. Etiología, epidemiología y tratamiento

T. brucei es el agente etiológico de la tripanosomiasis africana humana (*Human African Trypanosomiasis*, HAT) o enfermedad del sueño, y la nagana en el ganado. Estas enfermedades son endémicas del África subsahariana y son transmitidas por la picadura de la mosca tse-tsé del género *Glossina* (Brun & Blum, 2012). Existen dos subespecies de *T. brucei* que causan esta afección en humanos: *T.b. gambiense* (responsable del 95% de los casos) que prevalece en África occidental y central y *T.b. rhodesiense* (responsable del 5% de los casos) que predomina en África oriental y meridional, ocasionando formas crónicas y agudas de la enfermedad, respectivamente. Por otro lado, *T. brucei brucei* no es capaz de infectar a humanos pero sí a animales domésticos y salvajes causándoles una enfermedad conocida como nagana, que afecta gravemente al sector ganadero (Brun et al., 2010). Los animales también pueden infectarse con *T.b. rhodesiense* y *T.b. gambiense* sin manifestar la enfermedad, pero convirtiéndose así en grandes reservorios.

La HAT sigue siendo una de las principales enfermedades mortales del mundo y afecta a millones de personas en 36 países del África subsahariana. En los últimos cinco años más del 70% de los casos notificados han tenido lugar en la República Democrática del Congo. Angola, el Chad, el Congo, el Gabón, Guinea, Malawi, la República Centroafricana y Sudán del Sur notificaron entre 10 y 100 nuevos casos en 2019, mientras que Camerún, Costa de Marfil, Guinea Ecuatorial, la República Unida de Tanzania, Uganda, Zambia y Zimbabue notificaron entre 1 y 10 nuevos casos ([https://www.who.int/es/news-room/fact-sheets/detail/trypanosomiasis-human-african-\(sleeping-sickness\)](https://www.who.int/es/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness))). En 2009 la Organización Mundial de la Salud (OMS) estimó que había tan solo 9878 nuevos casos de HAT por año. Esta tendencia se ha mantenido hasta el punto de que en los años 2019 y 2020 se han notificado 992 y 663 casos respectivamente, el nivel más bajo desde que se empezaron a recopilar

INTRODUCCIÓN

datos hace 80 años ([https://www.who.int/es/news-room/fact-sheets/detail/trypanosomiasis-human-african-\(sleeping-sickness\)](https://www.who.int/es/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness))). En algunos países la transmisión parece haberse detenido, si bien es cierto que es difícil de determinar debido a factores como la inestabilidad social, la pobreza y la falta de medios. El objetivo de la OMS es detener la transmisión de esta enfermedad para el año 2030, lo que supone encontrar medicamentos seguros y eficaces, así como acabar con los reservorios animales.

La enfermedad del sueño cursa en dos etapas: una etapa temprana conocida como hemolinfática, que tiene lugar tras la picadura de la mosca tse-tsé, durante la cual los parásitos se multiplican y se propagan por el torrente sanguíneo, sistema linfático y órganos sistémicos (Matthews, 2005). En esta fase no suelen detectarse signos clínicos. Posteriormente, los parásitos logran cruzar la barrera hematoencefálica e invaden el sistema nervioso central dando lugar a la segunda etapa o fase meningoencefálica, que deriva en una amplia variedad de trastornos neurológicos llegando incluso al coma (Brun et al., 2010; Kennedy, 2008). Por lo general, esta enfermedad se considera letal sin tratamiento.

La HAT es considerada una enfermedad tropical desatendida y sigue siendo un importante problema de salud pública (Brun et al., 2010). Actualmente existen varios tratamientos disponibles según la subespecie (*T.b. gambiense* o *T. b. rhodesiense*) y la etapa de la enfermedad en la que se encuentre (hemolinfática o meningoencefálica). Durante la etapa temprana los tratamientos administrados son pentamidina y suramina, que resultan ser bastante efectivos, aunque a veces pueden aparecer recidivas. Para la etapa tardía, se utiliza melarsoprol o un tratamiento combinado de nifurtimox y eflornitina, con la gran desventaja de su elevada toxicidad, la aparición de resistencias y su difícil administración, ya se requiere la hospitalización del paciente. Un fármaco eficaz por vía oral libre de efectos secundarios sería de gran valor terapéutico, por lo que la identificación de posibles dianas farmacológicas para desarrollar alternativas terapéuticas es objeto de investigación.

Actualmente hay dos medicamentos orales contra la THA bastante prometedores: acoziborol (AN5568) (Jacobs, Nare, et al., 2011; Jacobs, Plattner, et al., 2011) y fexinidazol (Mesu et al., 2018). También se ha identificado uno nuevo contra la nagana, el compuesto AN11736 (Akama et al., 2018). Tanto acoziborol como AN11736 son benzoxaboroles, siendo su diana la endonucleasa CPSF3 (*Cleavage and Polyadenylation Specificity Factor 3*), implicada en el procesamiento del RNA mensajero (Wall et al., 2018).

1.3. Ciclo de vida

El ciclo de vida de *T. brucei* es complejo, ya que alterna entre un insecto vector (la mosca tsé-tsé) y un hospedador mamífero, por lo que el parásito experimenta un profundo proceso de diferenciación para adaptarse a los diferentes entornos y nutrientes disponibles y conseguir sobrevivir. Los tripomastigotes metacíclicos presentes en las glándulas salivares del insecto ingresan en el torrente sanguíneo del huésped mamífero a través de una picadura, donde se diferencian primero hacia formas proliferativas sanguíneas esbeltas (*slender*) y luego hacia formas rechonchas (*stumpy*) que no se dividen (Figura 2).

La forma esbelta emplea la glucosa en sangre como fuente de energía a través de la glucólisis, por lo que no necesita utilizar su mitocondria y la mantiene parcialmente inactiva. Además, estas células expresan una densa capa externa formada por glicoproteínas de superficie variables (*Variant Surface Glycoproteins*, VSGs), que les confiere la habilidad de evadir la respuesta inmune del hospedador a través de un mecanismo conocido como variación antigénica (Horn, 2014; Vickerman, 1969); el parásito posee numerosas copias de genes VSG que puede intercambiar periódicamente y escapar así de los anticuerpos desarrollados por el hospedador mamífero (Taylor & Rudenko, 2006).

INTRODUCCIÓN

Cuando los tripanosomas en sangre alcanzan una densidad elevada tiene lugar la diferenciación a la forma rechoncha quiescente (Figura 2); estas formas están metabólicamente preadaptadas a la mosca, alargando ligeramente su mitocondria y expresando enzimas del ácido cítrico y de la cadena respiratoria (Priest & Hajduk, 1994). Esta transición se basa en un mecanismo de *quorum sensing* asociado a la densidad celular cuya importancia radica en que una parasitemia descontrolada provocaría la muerte del hospedador y con ello el cese de la transmisión.

La forma rechoncha retorna a la mosca tsé-tsé durante una ingesta de sangre y en su intestino medio ocurre una nueva diferenciación hacia la forma procíclica. En esta transformación están implicadas determinadas proteínas asociadas con la diferenciación (PAD), que pueden detectar la bajada de temperatura desde los 37°C en el mamífero a los 20°C en la mosca y desencadenar el cambio hacia la forma procíclica (Dean et al., 2009). En los tripanosomas procíclicos, el kinetoplasto se desplaza desde la región posterior de la célula hacia una zona más cercana al núcleo (Matthews, 2005), tiene lugar un mayor alargamiento de la mitocondria y una activación de las enzimas que participan en el ciclo de Krebs y de la cadena de transporte de electrones (Flynn & Bowman, 1973). De esta forma, los parásitos pueden metabolizar los aminoácidos (principalmente prolina) en lugar de glucosa para producir ATP (Priest & Hajduk, 1994). Las formas procíclicas también expresan una capa de proteínas en su superficie, en este caso formada por procilinas EP, que reemplazan a las VSGs (Roditi & Liniger, 2002).

Tras su establecimiento en el tracto digestivo, los tripanosomas detienen su división y migran a las glándulas salivares del insecto donde se anclan a la membrana a través del propio flagelo. Posteriormente algunos se transforman en formas metacíclicas libres no proliferativas, que poseen una readquirida capa de VSGs que les sirve como preadaptación a la infección del huésped mamífero, cerrando así el ciclo (Figura 2).

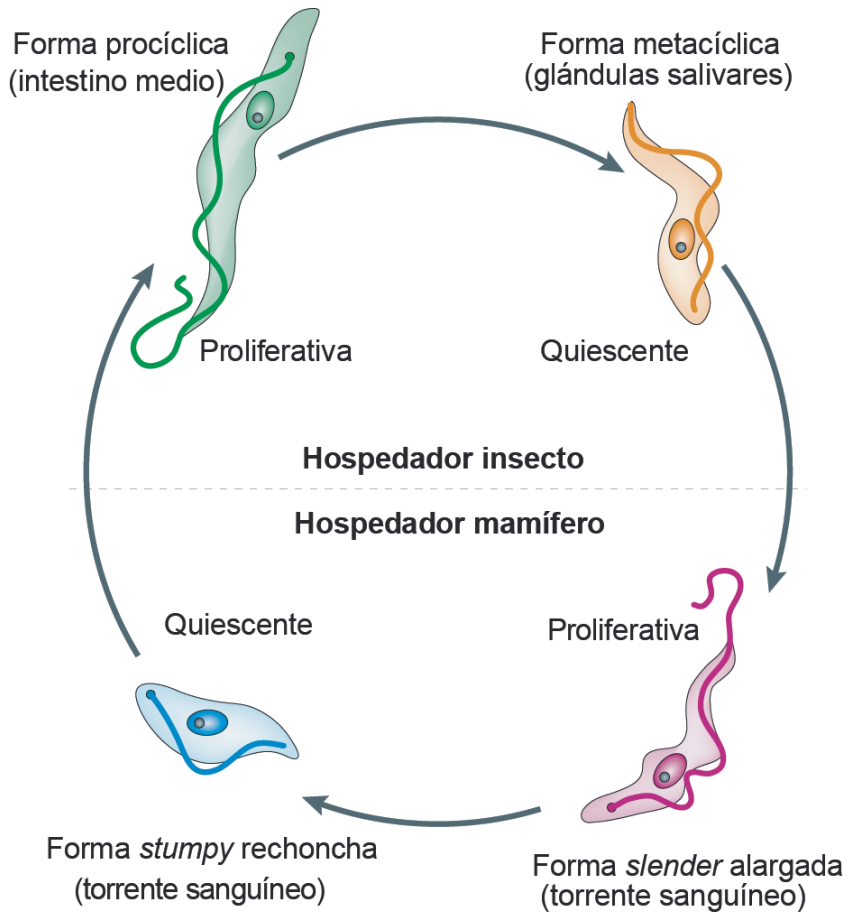


Figura 2. Ciclo de vida de *T. brucei*. El ciclo de este protozoo parásito extracelular alterna entre un hospedador artrópodo (la mosca tse-tsé) y un hospedador mamífero (humanos y otros mamíferos). La infección comienza tras la picadura de la mosca y la forma alargada del parásito (*slender*) prolifera en el torrente sanguíneo del mamífero. La forma quiescente 'rechoncha' (*stumpy*) se transfiere al insecto durante una ingesta de sangre cerrando el ciclo. Imagen adaptada de (Lee et al., 2007).

1.4. Variación antigénica

A diferencia de lo que sucede con otros parásitos intracelulares como *T. cruzi* o *Leishmania*, la infección por *T. brucei* es puramente extracelular. En consecuencia, al no existir ninguna etapa intracelular, el parásito se encuentra expuesto constantemente al sistema inmune del hospedador, por lo que necesita un mecanismo de evasión altamente

INTRODUCCIÓN

sofisticado para lograr una infección exitosa y persistente. De hecho, la enfermedad puede durar meses o incluso años. Este fenómeno se denomina variación antigénica (Horn, 2014; Vickerman, 1969) y consiste en la producción de una densa capa superficial formada por una proteína denominada VSG que sufre un frecuente intercambio, evadiendo la respuesta inmune generada por el huésped mamífero (Cestari & Stuart, 2018) (Figura 3). Esta dinámica conduce a oleadas de parasitemia en sangre y al establecimiento de una infección continua, sin acabar con la vida del hospedador.

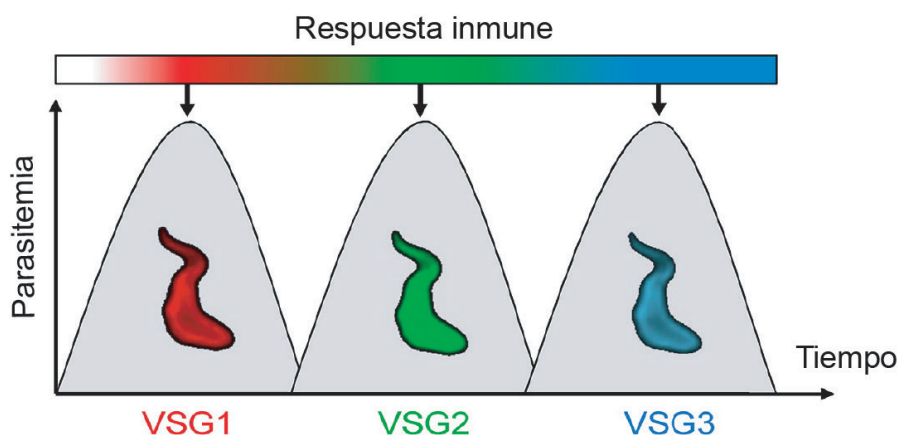


Figura 3. Variación antigénica. El perfil infeccioso se caracteriza por oleadas de parasitemia ocasionadas por parásitos que expresan diferentes VSGs en su superficie (representado con diferentes colores). Imagen adaptada de (Horn, 2014).

Las VSGs son homodímeros anclados a la membrana por glicosilfosfatidilinositol (*GlycosilPhosphate Inositol*, GPI), formando una monocapa de unos 15 nm de espesor que cubre todo el tripanosoma y el flagelo. Son muy abundantes, ya que conforman el 10% del total de la proteína celular y son altamente inmunogénicas. El genoma de *T. brucei* posee alrededor de 2500 genes VSG de los cuales el 80% son pseudogenes (Berriman et al., 2005; Cross et al., 2014). El éxito de la variación antigénica radica en la capacidad de expresar un solo antígeno entre innumerables posibilidades (expresión monoalélica) y la habilidad de cambiar de una

isoforma de antígeno a otra distinta (Duraisingh & Horn, 2016). Sin embargo, a pesar del vasto repertorio genético, sólo una pequeña parte es susceptible de ser expresada en los sitios de expresión (*expression sites*, ES) localizados cerca de los telómeros (subtelómeros) de los cromosomas mega-básicos e intermedios (Figura 4A). El resto están ubicados en conjuntos subtelo méricos haploides silenciosos y en minicromosomas (Figura 4A) (Barry, 1997; Vanhamme & Pays, 1995). A diferencia de los genes que codifican proteínas, que se transcriben por la RNA polimerasa II, los genes *VSG* son transcritos por la RNA polimerasa I (Günzl et al., 2003).

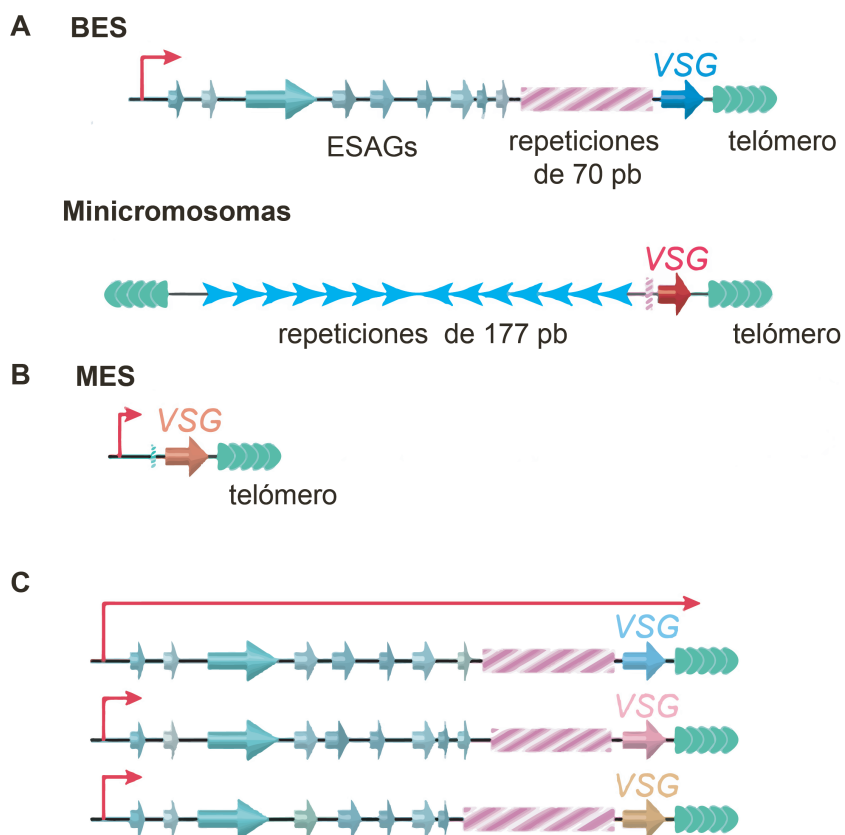


Figura 4. Organización genómica de los genes *VSG*. (A) Estructura de un sitio de expresión típico de sanguíneos (BES) y de un minicromosoma con un gen *VSG* subtelo mérico. (B) Estructura de un sitio de expresión metacíclico (MES). (C) Solo un sitio de expresión es activo en un momento dado. Imagen modificada de (Li & Zhao, 2021).

INTRODUCCIÓN

Los sitios de expresión de la VSG en sanguíneos (*Bloodstream Expression Sites*, BES) son unidades de transcripción policistrónica de 40-60 kb de extensión, que poseen un 90% de identidad de secuencia y una arquitectura conservada (Hertz-Fowler et al., 2008) formada por: un promotor RNA polimerasa I, un conjunto de genes en tándem conocidos como genes asociados al sitio de expresión (*Expression Site-Associated Genes*, ESAGs, la mayoría de los cuales tienen función desconocida) y un gen VSG (Figura 4A). En *T. brucei* solo se transcribe un gen VSG a la vez desde uno de los 15 BES en un momento determinado (Figura 4C) (Borst, 2002), en una estructura extranucleolar conocida como “cuerpo del sitio de expresión” (*expression site body*, ESB) (Navarro & Gull, 2001) que colocaliza con la RNA polimerasa I, garantizado así la expresión monoalélica.

Como se mencionó en el apartado anterior, las formas metacíclicas también están cubiertas por una capa de VSGs (ancladas por GPIs) que se expresan desde uno de los sitios de expresión de la VSG en la forma metacíclica (*Metacyclic Expression Sites*, MES). Esto les confiere una preadaptación a la infección del huésped mamífero (Graham & Barry, 1995). A diferencia de los BES, las MES son unidades de transcripción monocistrónica que contienen un gen VSG y un promotor 5kb aguas arriba (Figura 4B).

La expresión monoalélica de la VSG depende de múltiples factores, aunque el mecanismo aún no se conoce con detalle. El cambio de VSG puede producirse bien por una activación *in situ* o bien por recombinación de DNA. Mediante la activación *in situ*, el ES originalmente activo se silencia y otro ES situado en otro cromosoma que se encontraba silenciado, se activa. En cuanto al cambio por recombinación de DNA, en ocasiones un gen VSG silenciado se duplica y sustituye al VSG activo en el ES o bien puede ocurrir un “*crossover*” o intercambio entre telómeros de un VSG inactivo a uno activo (Myler et al., 1984). Por otro lado, se ha demostrado que la proteína TbRAP1 (Afrin et al., 2020; Pandya et al., 2013; Yang et al.,

2009), la elongación de la transcripción (Kassem et al., 2014; Schulz et al., 2013; Vanhamme & Pays, 1995), la estructura de la cromatina (Günzl et al., 2015; Navarro, 1999) y el complejo VEX1 que se asocia con el ESB (Faria et al., 2019; Glover et al., 2016), contribuyen al silenciamiento telomérico.

El establecimiento de la expresión monoalélica en tripanosomas metacíclicos apenas se conoce. A diferencia de los tripanosomas sanguíneos, la población metacíclica es heterogénea y muestra diferentes variantes de VSG; sin embargo, cada célula individual expresa sólo una MES VSG específica (Tetley et al., 1987; Vigneron et al., 2020). Se ha demostrado que estas formas del parásito transcriben múltiples genes VSG antes de escoger la proteína que se va a expresar en su superficie (Hutchinson et al., 2021). Respecto a los mecanismos implicados en metaciclogénesis en *T. brucei*, se sabe que la sobreexpresión de una proteína de unión a RNA llamada RBP6 desencadena la diferenciación hacia formas metacíclicas (Kolev et al., 2012). Se desconoce cómo los MES se mantienen en silencio en las formas sanguíneas, así como el resto de las proteínas implicadas en su regulación.

En definitiva, la variación antigénica asegura la supervivencia de la población, facilitando la cronificación de la enfermedad y su transmisión, lo cual dificulta el desarrollo de vacunas eficaces contra este parásito.

2. Regulación de la expresión génica en *T. brucei*

2.1. Organización genómica

T. brucei es un organismo diploide con un genoma de ~35 megabases organizado en 11 pares de cromosomas mega-básicos (de 1 a 6 Mb), unos 5 cromosomas intermedios (de 200 a 900 Kb) y aproximadamente 100 minicromosomas (de 50 a 150 Kb) (Berriman et al., 2005; Wickstead et al., 2004). Los cromosomas mega-básicos contienen todos los genes de mantenimiento celular (*housekeeping*), pre-rRNAs, tRNAs, RNAs pequeños

INTRODUCCIÓN

nucleares (snRNAs), RNAs pequeños nucleolares (snoRNAs), el RNA del *spliced leader* (SLRNA) (Berriman et al., 2005) y son diploides excepto en sus extremos (Taylor & Rudenko, 2006); los cromosomas intermedios y los minicromosomas comparten una estructura común que consiste en una zona central palindrómica y una región subtelomérica no repetitiva en ambos extremos, seguido de repeticiones teloméricas (Wickstead et al., 2004). Las regiones subteloméricas de ambos tipos de cromosomas contienen genes VSG y presentan un alto grado de similitud con las zonas subteloméricas de los megacromosomas, lo que permite la recombinación entre ambas, siendo por tanto una reserva de VSGs. Además, alrededor del 20% del genoma de *T. brucei* codifica genes subteloméricos (El-Sayed et al., 2005), la mayoría de los cuales están relacionados con el proceso de variación antigénica.

Una de las características más interesantes y únicas del parásito es su particular organización genética. Los genes se organizan en unidades de transcripción policistrónica (*Polycistronic Transcription Units*, PTUs) que pueden contener hasta cientos de genes en tándem, que se transcriben desde la misma cadena de manera constitutiva por la RNA polimerasa II. Aunque el orden de los genes está parcialmente conservado a nivel evolutivo (Jackson et al., 2008), los genes que forman parte de las unidades transcripcionales no están relacionados con respecto a la expresión o función de la proteína que codifican. La transcripción tiene lugar desde unos pocos sitios dentro del cromosoma, es decir, al contrario que en otros eucariotas, no hay un promotor para cada gen. Las PTUs están separadas por regiones de cambios de cadena (*Strand Switch-Region*, SSR) que pueden ser divergentes (*divergent Strand Switch-Region*, dSSR) o convergentes (*convergent Strand-Switch Region* cSSR) (Figura 5). Los mRNAs monocistrónicos se generan mediante reacciones acopladas de *trans-splicing* (Michaeli, 2011), que consiste en la incorporación del mini-exón *spliced leader* (SL) en el extremo 5' y la poliadenilación del extremo 3' del transcrito (Clayton & Michaeli, 2011). El SL está hipermetilado en su extremo 5' y su incorporación conlleva por tanto a la adición de una caperuza (conocida como *cap4*) al extremo 5'

de todos los mRNAs del parásito (Perry et al., 1987), lo que permite su traducción en el citoplasma y a la vez los protege de las maquinarias de degradación.

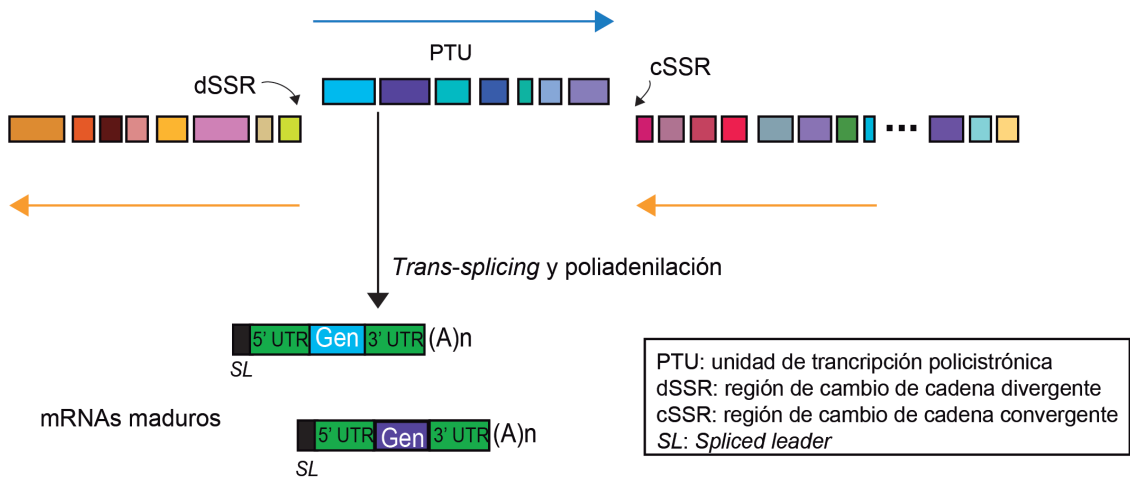


Figura 5. Transcripción policistrónica y procesamiento del mRNA en tripanosomátidos.

Las dSSRs se asocian con sitios de inicio de la transcripción, mientras que las cSSRs se asocian con sitios de terminación. El mRNA maduro contiene el mini-exón SL en su extremo 5' y una cola de poli(A) en su extremo 3'. Basado en (Siegel et al., 2009).

Los tripanosomátidos poseen tres RNA polimerasas; la RNA polimerasa III transcribe los RNA de transferencia (tRNA), los U-RNA (U2, U3 y U4) y los RNA pequeños no codificantes. La RNA polimerasa II se encarga de la transcripción de los RNA mensajeros (mRNAs) y del SLRNA. Por último, la RNA polimerasa I, además de transcribir los RNA ribosomales (rRNA) como en el resto de eucariotas, en tripanosomátidos tiene la peculiaridad de transcribir los genes VSG en las formas sanguíneas y la prociclina en las formas procíclicas (Clayton et al., 1990; Günzl et al., 2003; Rudenko et al., 1989).

INTRODUCCIÓN

2.2. Regulación de la transcripción

2.2.1. Inicio de la transcripción

En eucariotas un paso importante en la regulación de la expresión génica ocurre a nivel de inicio de la transcripción, debido a la presencia de promotores bien definidos y con la participación de numerosos factores de transcripción.

En tripanosomas encontramos un escenario bastante diferente debido a la singular organización en tándem de sus genes, donde el inicio de la transcripción no parece tener un papel muy relevante en el control de la expresión génica (Clayton, 2002) y se asume que la transcripción por parte de la RNA polimerasa II es constitutiva (Martínez-Calvillo et al., 2003). De hecho, los factores de transcripción basales de la RNA polimerasa II conocidos en estos organismos son extremadamente divergentes en cuanto a su secuencia aminoacídica por lo que, para su identificación, se suele recurrir a estudios funcionales y estructurales (Günzl et al., 2020). Entre ellos se incluyen la proteína de unión a la caja TATA (*TATA box Binding Protein*, TBP) (Ruan et al., 2004), TFIIB (Schimanski, 2006), la subunidad pequeña TFIIA y SNAPc (Das et al., 2008), TFIIH (Lecordier et al., 2007), el complejo mediador (Lee et al., 2010), el complejo PAF (Ouna et al., 2012) y dos proteínas específicas de kinetoplastidos, similares a TFIIIF (Schimanski et al., 2005; Srivastava et al., 2018). Estos factores son necesarios para la transcripción del precursor de 142 nt del *SLRNA in vitro* y seguramente también sean necesarios para la transcripción policistónica de genes que codifican proteínas, aunque en la mayoría de los casos este aspecto no se ha demostrado. Es muy probable que los tripanosomátidos posean factores de transcripción adicionales, aunque su identificación bioinformática resulte difícil, bien porque sean muy diferentes, o bien porque sean específicos de estos organismos que divergieron muy pronto en la evolución (Günzl et al., 2020).

Las unidades de transcripción policistrónica están separadas por zonas de cambios de cadena, pudiendo ser divergentes (dSSRs) o convergentes (cSSRs) (Figura 5). Estudios de *run-on* han demostrado que las dSSRs se asocian con los sitios de inicio de la transcripción (*Transcription Start Sites*, TSS) de la RNA polimerasa II, mientras que los sitios de terminación de la transcripción (*Transcription Termination Sites*, TTS) se ubican en las cSSRs o en las proximidades de los genes transcritos por diferentes polimerasas (Clayton, 2019; Martínez-Calvillo et al., 2010; Martínez-Calvillo et al., 2003; Siegel et al., 2009), ya que son sitios de parada y reinicio de la transcripción. Sin embargo, el mecanismo exacto de cómo se inicia o se detiene la transcripción sigue sin conocerse (Martínez-Calvillo et al., 2010). Recientemente se han identificado numerosos factores enriquecidos en las TSS que podrían actuar como potenciales reguladores de la cromatina (Staneva et al., 2021). En concreto, se ha demostrado que el complejo regulador asociado al promotor (*SET27Promoter-Associated Regulatory Complex*, SPARC) formado por 6 proteínas, asegura un correcto inicio de la transcripción por parte de la RNA polimerasa II (Staneva et al., 2022).

Los TSS poseen una serie de características definidas: carecen de marcos abiertos de lectura (*Open Reading Frames*, ORFs), son zonas de cromatina abierta (Wedel et al., 2017) y están enriquecidos en marcas epigenéticas esenciales como H4K10ac, H3K4ac, H2AZ, H2BV, el factor de bromodominio BDF3 (Siegel et al., 2009; Wright et al., 2010) e híbridos RNA-DNA (Briggs et al., 2018). Aunque varios resultados indican que la iniciación por parte de la polimerasa II puede favorecerse simplemente por una estructura abierta de la cromatina (LeBowitz et al., 1990; Marchetti, 1998; McAndrew et al., 1998), recientemente se ha visto que en estas zonas existen promotores específicos de secuencia ricos en GT capaces de iniciar la transcripción (Cordon-Obras et al., 2022; Wedel et al., 2017). Se desconoce cómo estas secuencias son reconocidas por factores remodeladores o si promueven la apertura de la cromatina por sí mismas.

INTRODUCCIÓN

El *SLRNA* es el único transcrito de polimerasa II que posee un promotor bien caracterizado en *T. brucei* (Das et al., 2005). Este promotor carece de caja TATA, sin embargo, se han identificado varios factores de transcripción ortólogos que interactúan específicamente con este promotor (Schimanski, 2006; Schimanski et al., 2005). Alrededor de 100 copias del gen *SLRNA* se encuentran dispuestas en tándem (de Lange et al., 1983) debido a la elevada síntesis requerida para el *trans-splicing*, tanto de los genes que codifican proteínas de mantenimiento celular, como de los genes *VSG* y *EP* que son muy abundantes; de hecho, el ES activo colocaliza con el locus *SLRNA* en el núcleo (Faria et al., 2021). En las formas sanguíneas cada gen *SLRNA* necesita producir alrededor de 140 mRNAs por hora, en contraste con los genes que codifican proteínas que probablemente se transcriben sólo una o dos veces por hora (Haanstra et al., 2008). También encontramos picos de nucleosomas al principio y al final de los genes *SLRNA* (Maree & Patterton, 2014).

Como se ha indicado anteriormente, en *T. brucei* además de los rDNA, tanto los genes VSGs como los de la prociclina se transcriben por la RNA polimerasa I (Günzl et al., 2003) y sus promotores han sido caracterizados (Palenchar & Bellofatto, 2006). Mientras que el promotor de la prociclina y del rDNA son muy similares y contienen tres elementos reguladores, el de la VSG requiere otras secuencias reguladoras adicionales, lo que sugiere la participación de factores de transcripción específicos. Actualmente los mecanismos de regulación de estos promotores son desconocidos, pero se sabe que la localización cromosómica y los efectos epigenéticos juegan un papel importante en la variación antigénica.

Una característica muy importante de la transcripción mediada por RNA polimerasa I es que puede ser hasta 10 veces más potente que la transcripción mediada por la RNA polimerasa II (Biebinger et al., 1996). Esto garantiza una alta producción de transcrito a partir de un solo gen *VSG*, dando lugar al 10% del total de mensajeros en la forma sanguínea del parásito (Clayton, 2002). Además, al contrario de lo que sucede en otros organismos, la RNA polimerasa I puede generar transcritos traducibles

debido a que el procesamiento de RNA y el inicio de la transcripción están desacoplados en tripanosomátidos; el *trans-splicing* asegura la incorporación del mini-exón SL y de la caperuza hipermetilada en el extremo 5' de los VSG mRNAs.

2.2.2. Terminación de la transcripción

El mecanismo de terminación de la transcripción ha sido ampliamente estudiado en la levadura *Saccharomyces cerevisiae*, siendo un proceso clave que no solo delimita las unidades de transcripción, sino que se necesita para mantener la expresión adecuada de los genes vecinos así como el reciclaje de la propia polimerasa. Tiene lugar cuando la polimerasa y el RNA naciente se liberan del DNA molde. Sin embargo, a pesar de que en los últimos años ha cobrado una gran importancia biológica, sigue siendo uno de los mecanismos menos comprendidos en la expresión génica.

Desde hace cierto tiempo se sabe que la terminación de la transcripción por parte de la RNA polimerasa II de los genes que codifican proteínas, está relacionada funcionalmente con la escisión y la poliadenilación del extremo 3' del transcrito naciente. En esta reacción en dos pasos, la transcripción de la señal de poliadenilación AAUAAA (*Polyadenylation Signal*, PAS) desencadena la escisión endonucleolítica del transcrito naciente, generando un producto de corte aguas arriba que se poliadenila inmediatamente (Colgan & Manley, 1997). El producto de escisión restante que queda aguas abajo es muy inestable y se degrada rápidamente (Manley et al., 1982). El mecanismo de terminación dependiente de PAS ha sido muy debatido en torno a dos modelos predominantes que no son mutuamente excluyentes: el modelo alostérico (Logan et al., 1987) y el modelo de torpedo (Kuehner et al., 2011; Luo et al., 2006; Richard & Manley, 2009).

INTRODUCCIÓN

El modelo alostérico postula que cuando la polimerasa se encuentra con el PAS, sufre un cambio conformacional, por ejemplo, por el reclutamiento de un factor de elongación negativo o la liberación de un factor de antiterminación. Tales cambios desestabilizarían el complejo de elongación y, por tanto, favorecerían la terminación. Por otro lado, el modelo de torpedo (Connelly & Manley, 1988) propone que el fosfato expuesto en el extremo 5' tras la escisión en el PAS, proporciona un punto de entrada para la exonucleasa de RNA 5'-3' (Rat1 en levaduras y Xrn2 en mamíferos). La exonucleasa rastrea y degrada el RNA naciente aguas abajo del sitio de corte y, cuando colisiona con la polimerasa, la desaloja del molde de DNA.

Debido a la existencia de datos que respaldan ambos modelos probablemente la terminación de la transcripción refleja una combinación de ambos mecanismos, dando lugar a un modelo híbrido tal y como se muestra en la Figura 6 (Luo et al., 2006).

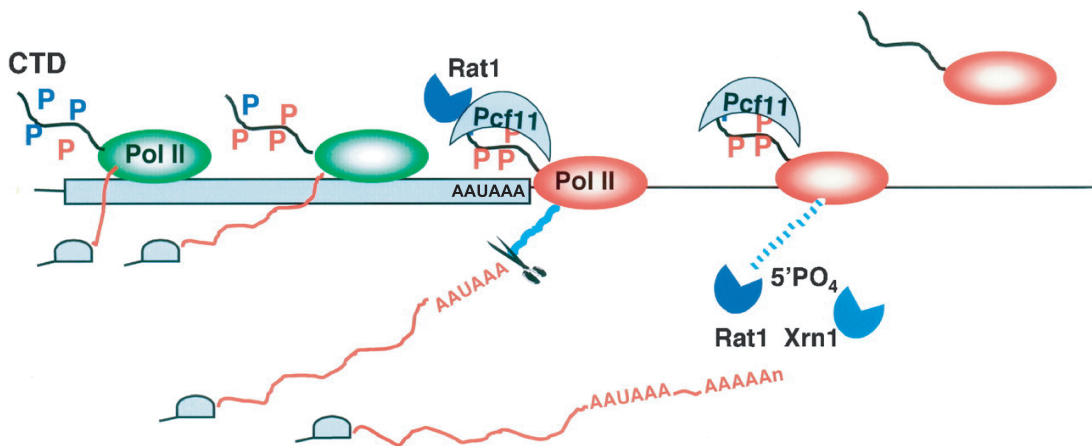


Figura 6. Modelo híbrido de terminación de la transcripción acoplada a al procesamiento del extremo 3'. Este modelo incorpora aspectos del modelo alostérico y de torpedo. Inmediatamente después de transcribir la secuencia consenso de poliadenilación AAUAAA, la RNA polimerasa II pausa la transcripción tras sufrir un cambio conformacional (mostrado como el cambio de color de verde a rojo) donde participan factores de procesamiento como Pcf11, favoreciendo la terminación. A continuación tiene lugar la escisión endonucleolítica del RNA naciente en el sitio de poliadenilación y la subsiguiente degradación por Rat1 (o Xrn2 en mamíferos) y Xrn1, facilitando así la terminación. Imagen modificada de (Luo et al., 2006).

En kinetoplastidos, la información acerca de los mecanismos y factores involucrados en este proceso es muy escasa. Además, el acoplamiento funcional de los procesos de terminación y procesamiento en 3' no parece algo plausible en organismos cuya transcripción es policistronica. En estos parásitos, el fin de la transcripción tiene lugar en las zonas donde convergen dos unidades transcripcionales (cSSR), en genes tRNA o transcritos por la RNA polimerasa I y aguas arriba del tándem *SLRNA*. Al contrario de lo que encontramos en los sitios de inicio de la transcripción, los sitios de terminación se enriquecen en tres marcas no esenciales: la base J (β -D- glucosyl-hydroxymethyluracil, una modificación de DNA específica de tripanosomátidos) y las variantes de histonas H3v y H4v (Figura 7) (Cliffe et al., 2010; Schulz et al., 2016; Siegel et al., 2009).

En *Leishmania*, la reducción de la base J conduce a una acumulación de transcritos antisentido en las cSSRs (Reynolds et al., 2014; van Luenen et al., 2012), mientras que en *T. brucei* la terminación de la transcripción se ve afectada en sitios específicos a lo largo de la PTU, lo que resulta en una mayor expresión de los genes situados aguas abajo (Reynolds et al., 2014). La base J está ausente en las formas procíclicas de *T. brucei* (Gommers-Ampt et al., 1991) y en *T. cruzi* también puede influir en la iniciación (Ekanayake et al., 2011), por tanto su papel parece ser específico de especie.

La H3v y la base J, pero no H4v, también se enriquecen en los telómeros (Cliffe et al., 2010; Siegel et al., 2009). Aunque H3v contribuye al proceso de terminación de la transcripción, un estudio reciente ha descrito que su función principal radica en la transcripción monoalélica de la VSG, ya que en los mutantes carentes de H3v tiene lugar una clara sobreexpresión de los sitios de expresión de sanguíneos (BES) y sitios de expresión de metacíclicos (MES), además de cambios en la cromatina (Kim, 2021; Müller et al., 2018). Aparentemente, H4v parece ser la principal marca de cromatina en los TTS a pesar de que los otros factores juegan un papel de apoyo en esos sitios (Figura 7) (Kim, 2021). Los tripanosomas carentes

INTRODUCCIÓN

de H4v son viables y tienen mayores niveles de transcritos antisentido que provienen de TTS (Kim, 2021). En este contexto, se ha propuesto que la maquinaria de la RNA polimerasa II no encuentra una “señal” para detenerse y continúa transcribiendo e invadiendo la unidad transcripcional adyacente tal y como se ilustra en la Figura 7.

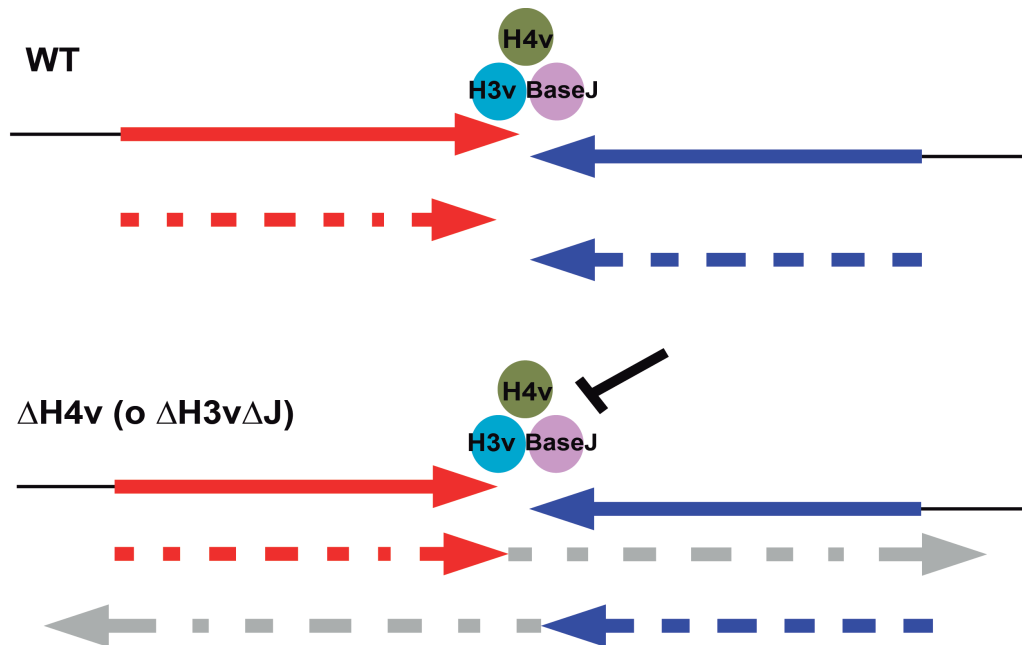


Figura 7. Marcas epigenéticas específicas de los sitios de terminación (TTS). En los mutantes carentes de H4v ($\Delta H4v$) o H3v y base J ($\Delta H3v\Delta J$) la RNA polimerasa II no se detiene y continúa generando transcritos antisentido (en color gris) hasta alcanzar el siguiente sitio de inicio de la transcripción (TSS). Basado en (Kim, 2021).

2.3. Procesamiento, transporte y degradación del RNA mensajero

2.3.1. Procesamiento: *trans-splicing* y poliadenilación

En tripanosomas, los mensajeros maduros se generan a partir de RNAs policistrónicos mediante reacciones acopladas de *trans-splicing* y poliadenilación. Estos procesos son co-transcripcionales pero hasta el momento no existen evidencias de que estén físicamente acoplados a la polimerasa como sucede en el resto de eucariotas, por lo que cualquier polimerasa (como la RNA polimerasa I o incluso polimerasas heterólogas como la polimerasa del fago T7) es capaz de generar transcritos traducibles. De hecho, en el momento en que ocurre el procesamiento, seguramente la polimerasa se encuentre aproximadamente 1 kb aguas abajo (Fadda et al., 2014).

El *trans-splicing* fue descubierto por primera vez en tripanosomas (de Lange et al., 1984; Kooter & Borst, 1984) aunque posteriormente se encontró en otros organismos (Krchňáková et al., 2017). Durante este proceso, se escinde el mensajero individual a partir del policistrón y tiene lugar la incorporación en el extremo 5' del mini-exón de 39 nt denominado *spliced leader* (SL), además de la caperuza hipermetilada. Como consecuencia, todos los RNA mensajeros de los tripanosomas comienzan por los mismos 39 nucleótidos correspondientes al mini-exón SL.

Los sitios aceptores de *splicing* (*Splice-Acceptor Sites*, SAS) suelen estar precedidos por una región de polipirimidinas, lo que se conoce en inglés como *polypyrimidine tract* (Kolev et al., 2010; Siegel et al., 2010) y cuanto mayor es su longitud, más eficiente es el procesamiento (Huang & van der Ploeg, 1991; Siegel et al., 2005). Existe una preferencia por U sobre C en *T. brucei* pero no en *Leishmania major* (Dillon et al., 2015; Smith et al., 2008), siendo el primer AG en 3' el sitio aceptor del *splicing* (Hummel et al., 2000; Kolev et al., 2010). Normalmente es posible encontrar *polypyrimidine tracts* en las regiones intergénicas y en las regiones no

INTRODUCCIÓN

codificantes en 3' (*3' Untranslated Región*, 3'-UTRs), lo que le confiere al parásito la capacidad de utilizar sitios alternativos de *splicing*. Esto da lugar a diferentes transcritos y a la incorporación o exclusión de señales de localización subcelular afectando, en última instancia, a la estabilidad del mensajero (Benabdellah et al., 2007; Rettig et al., 2012).

En cuanto al mecanismo de poliadenilación, consiste en la adición de una cola de poli(A) en el extremo 3' del transcrito. *T. brucei* posee dos poli(A) polimerasas, aunque una de ellas se localiza en el citosol y no parece ser funcional (Koch et al., 2016). El complejo de poliadenilación es convencional salvo por la presencia de dos subunidades que no tienen homólogos obvios en eucariotas no relacionados (Koch et al., 2016). A diferencia de lo que ocurre en otros eucariotas, no existe ninguna señal específica de poliadenilación. Estos sitios se eligen según la distancia a las señales de *trans-splicing* (tracto de polipirimidinas) presentes en el primer cistrón aguas abajo (Campos et al., 2008; Clayton & Michaeli, 2011; Kolev et al., 2010), por lo que se asume que ambas reacciones deben de estar funcionalmente acopladas. De hecho, cuando se inhibe el *trans-splicing* tampoco tiene lugar la poliadenilación (Hug et al., 1994). Además, al mutar el *polypyrimidine tract*, el parásito utiliza un sitio de poliadenilación aberrante (Ullu et al., 1993).

2.3.2. Transporte del RNA del núcleo al citosol

El transporte del RNA desde el núcleo es esencial para todas las células eucariotas y es un paso determinante en el control de la expresión génica. En kinetoplastidos no se dispone de mucha información acerca de este proceso. Concretamente en *T. brucei*, se sabe que cuando se silencia la expresión de Mex67 (ortólogo de un factor de exportación nuclear de levaduras), se produce una acumulación de mRNAs poliadenilados en el núcleo (Schwede et al., 2009). Mex67 y Mtr2 interactúan y funcionan como receptores de exportación de mRNA y tRNA en el núcleo (Dostalova et al., 2013; Hegedúsová et al., 2019; Schwede et al., 2009).

Este fenotipo también es detectado en ausencia del ortólogo de NMD3, pero aparentemente solo tiene lugar una acumulación de transcritos de genes asociados a la prociclina (PAGs) (Bühlmann et al., 2015). Además, parece que la célula posee un mecanismo de control que impide que los transcritos mal procesados alcancen el citoplasma y se traduzcan, reteniéndolos en la periferia nuclear en gránulos NPG (*Nuclear Periphery Granules*) (Goos et al., 2019).

2.3.3. Degradación del RNA

Por lo general, en levaduras y mamíferos la degradación de los RNAs comienza retirando la cola de poli(A) mediante una exonucleasa que puede ser PRNA, o los complejos Pan2/Pan3 o Ccr4/Caf1/Not. A continuación el mRNA puede degradarse en sentido 3'-5' por el complejo exosoma (Buttner et al., 2006), el cual se describirá más adelante. Alternativamente tiene lugar un segundo paso, que consiste en la hidrólisis de la caperuza en 5' por una serie de enzimas que pertenecen a la familia de hidrolasas similares a MutT (Dcp1, Dcp2 y la helicasa Dhh1) en un proceso que se conoce como "decapping". Finalmente, el mRNA se degrada en el citoplasma en dirección 5'-3' gracias a la exoribonucleasa Xrn1 (Keene, 2010; Parker & Song, 2004). En levaduras, Xrn1 también se encarga de degradar los transcritos inestables sensibles a Xrn1 (*Xrn1-sensitive Unstable Transcripts*, XUTs) que, en su mayoría, son antisentido (van Dijk et al., 2011). También actúa como un factor de transcripción a nivel de inicio y de elongación y está involucrada en la exportación de mRNA desde el núcleo (Medina et al., 2014).

En *Trypanosoma* encontramos tres sistemas de desadenilación que están presentes también en mamíferos: el complejo CAF1/NOT, PAN2/PAN3 y homólogos de PRNA, siendo el primero el principal responsable de la desadenilación en los parásitos (Schwede et al., 2008); la mayoría de los mensajeros sufren una desadenilación antes de su degradación (Clayton & Shapira, 2007). A pesar de poseer cinco potenciales hidrolasas similares a MutT, ninguna está implicada en el *decapping*. En su lugar, ALPH1

INTRODUCCIÓN

(fosfatasa tipo ApaH) se encarga de este proceso (Kramer, 2017), aunque se desconoce cómo identifica los mRNAs que deben ser degradados. *Trypanosoma* posee 4 exoribonucleasas similares a Xrn1: XRNA, XRNB, XRNC y XRND (Li et al., 2006) sin embargo, tan solo XRNA, que está presente tanto en el citoplasma como en el núcleo, está implicada en la degradación (Manful et al., 2011). La ausencia de XRNA es letal en ambas formas del parásito y, como consecuencia, tiene lugar la estabilización de mRNAs regulados en el desarrollo que, inicialmente, eran muy inestables (Li et al., 2006). Estudios sobre los transcritos que codifican la prociclina EP en formas sanguíneas demostraron que, excepcionalmente, estos mensajeros se degradan desde ambos extremos, con la implicación de XRNA, CAF1, PAN2 y el exosoma (Haile et al., 2003; Irmer, 2001; Li et al., 2006; Schwede et al., 2009), lo que sugiere que los mRNAs inestables son degradados sin desadenilación previa.

En levaduras y mamíferos el exosoma es un complejo multiprotéico compuesto por al menos 11 proteínas, siendo Rrp6 y Rrp44 las subunidades catalíticas con actividad exonucleasa y exonucleasa/endonucleasa respectivamente. Este complejo está involucrado en la degradación y el procesamiento de varios tipos de RNAs y se encuentra tanto en el núcleo como en el citoplasma. En el citoplasma degrada los mensajeros que contienen errores, codones de STOP prematuros (*Nonsense-Mediated Decay*, NMD) (Lejeune et al., 2003) o aquellos que no poseen codón de STOP (*Non-Stop Decay*, NSD) (Frischmeyer et al., 2002; van Hoof et al., 2002). En metazoos la helicasa dependiente de ATP Upf1 es esencial para la degradación por NMD (Delhi et al., 2011), mientras que la presencia de la proteína de unión a poli(A) (*Poly(A)-Binding Protein*, PABP) tras un codón de STOP prematuro inhibe la degradación por esta vía. *Trypanosoma* tiene dos homólogos Upf1 y Upf2, siendo Upf1 esencial en las formas procíclicas. Sin embargo, no existen evidencias claras de una ruta NMD clásica en tripanosomátidos (Delhi et al., 2011).

En el núcleo, la función principal del exosoma consiste en el procesamiento del RNA pequeño nuclear (Allmang, 1999), mientras que

en el nucleolo procesa el rRNA 5.8S y los RNAs pequeños nucleolares (Allmang, 1999; Mitchell et al., 1997; Schilders, 2005). Aparte de su función clave en la vigilancia y el recambio del RNA, en levaduras también elimina los denominados transcritos crípticos inestables (*Cryptic Unstable Transcripts*, CUTs) (Neil et al., 2009; Wyers et al., 2005) que son producto de una transcripción generalizada por parte de la RNA polimerasa II, aunque su función se desconoce (ver más abajo en el apartado 3).

La estructura del exosoma en *T. brucei* es similar a la de otros eucariotas. Está formado por 6 subunidades y posee homólogos tanto de Rrp6 (RRP6) como de Rrp44 (RRP44) y su localización es principalmente nuclear (Haile & Papadopoulou, 2007; Kramer et al., 2016). La única diferencia con respecto al exosoma de otros eucariotas es que la subunidad RRP44 no está asociada. Todos los componentes están involucrados en la maduración del rRNA 5.8S (Estévez et al, 2001). El silenciamiento de la expresión de las subunidades del exosoma es letal para el parásito (Estévez et al, 2001; Estévez et al., 2003) y causa una acumulación de policistrones de mRNA, lo que sugiere una función mayoritaria de “control de calidad” a nivel nuclear (Kramer et al., 2016). En el citosol parece que su papel no es tan importante, y su función probablemente sea la degradación 3'-5' de transcritos muy inestables.

3. Transcripción generalizada

Gracias a los avances en las tecnologías de secuenciación masiva, se han identificado transcritos no canónicos, no codificantes y principalmente antisentido que se generan a consecuencia de una “*pervasive transcription*” (transcripción generalizada o ubicua) de casi cualquier región genómica en organismos tan diversos como bacterias, levaduras y mamíferos (Jensen et al., 2013; Wade & Grainger, 2014) (Figura 8A-B). Esto quiere decir que se producen transcritos distintos a los que codifican proteínas y de los que tienen funciones conocidas como tRNA y rRNA, que hasta ahora no habíamos sido capaces de detectar.

INTRODUCCIÓN

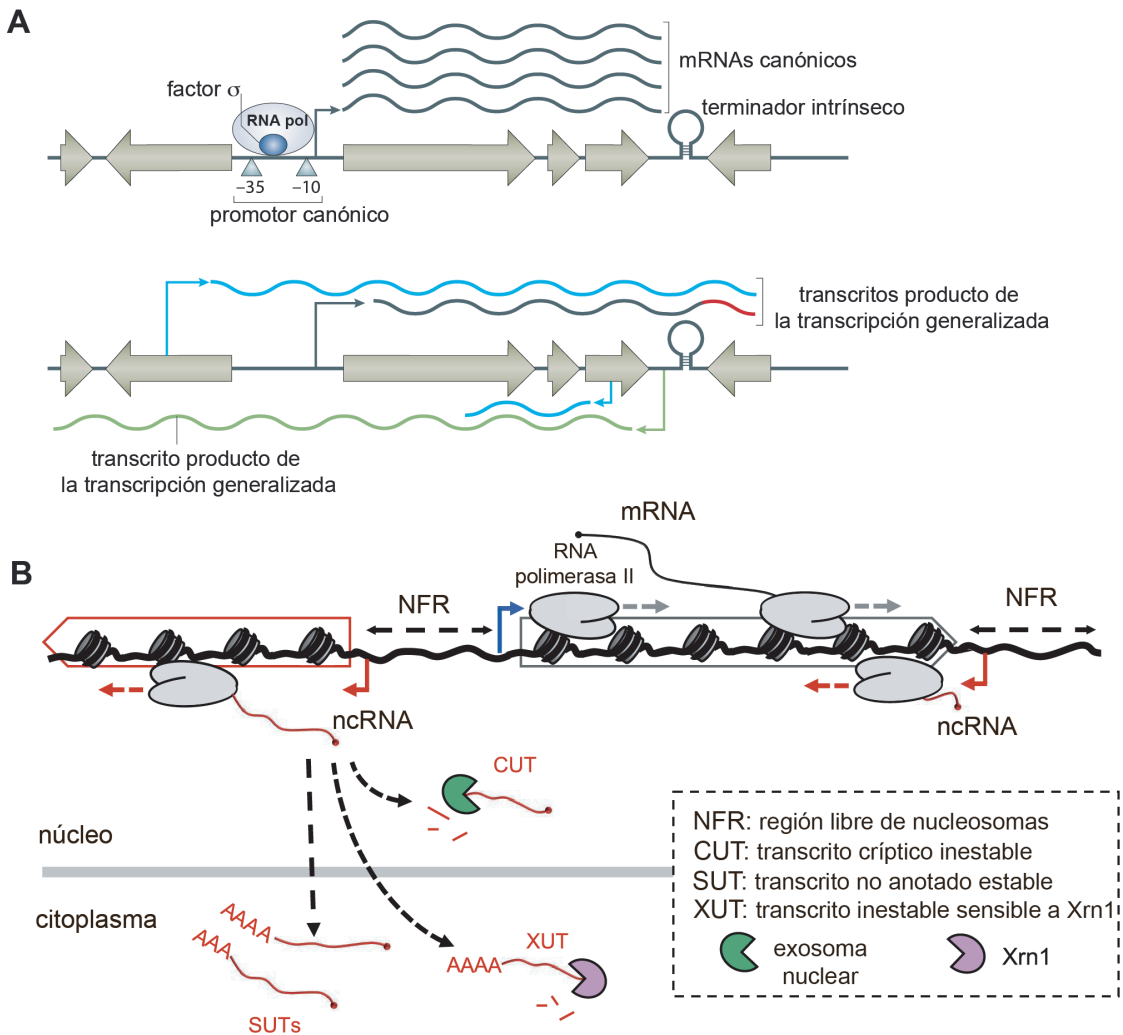


Figura 8. Transcripción generalizada. (A) Esquema en bacterias. La transcripción se inicia cuando la RNA polimerasa reconoce un promotor canónico correctamente orientado en una región intergénica y gracias a la formación del complejo de elongación. La transcripción genera mRNAs canónicas (en color negro) y finaliza cuando se encuentra con un terminador intrínseco. La transcripción ubicua tiene lugar cuando el complejo de elongación no consigue detenerse, dando como resultado la transcripción del gen situado aguas abajo (color rojo). El transcrito puede ser sentido o antisentido, según la orientación del gen aguas abajo. En el ejemplo que se muestra el gen está orientado a la inversa, generándose un RNA antisentido (en color rojo). Los promotores intragénicos pueden generar transcritos sentido o antisentido (transcrito en color azul), y lo mismo sucede con los promotores intergénicos (transcrito en color verde). Imagen adaptada de (Wade & Grainger, 2014). (B) Esquema en eucariotas. Los transcritos son generados por la RNA

polimerasa II (en color gris) desde regiones abiertas de cromatina (regiones libres de nucleosomas o NFRs) donde se encuentran los promotores (flecha azul). Esto conduce a la creación de numerosos RNA no codificantes (ncRNAs) mostrados en color rojo, muchos de los cuales tienen una orientación antisentido respecto a los genes que codifican proteínas. Los promotores crípticos se indican con flechas rojas. Los diferentes ncRNAs producidos (CUT, XUT y SUT) serán degradados según su sensibilidad vía Xrn1 o exosoma nuclear. Imagen modificada de (Villa & Porrua, 2022).

En las células animales, los RNA no codificantes (*non-coding RNAs*, ncRNA) generalmente se clasifican como pequeños o largos. La longitud de los ncRNA largos (*long non-coding RNAs*, lncRNAs) es > 200 nt y, en su mayoría, suelen originarse en regiones promotoras, por lo que se asocian a la transcripción de mRNA. Los ncRNA pequeños incluyen los RNAs que regulan el *splicing*, los RNAs pequeños nucleares (*small nuclear RNAs*, snRNAs), los tRNAs, los RNAs pequeños de interferencia, los micro-RNAs que regulan la estabilidad y la traducción del mRNA y los RNA pequeños nucleolares (*small nucleolar RNAs*, snoRNAs) que regulan el procesamiento del RNA ribosomal entre otros (Morris & Mattick, 2014). *T. brucei* tiene todos estos RNA pequeños, excepto los micro-RNAs (Rajan et al., 2020). Aunque algunos autores no consideran este tipo de RNAs producto de la transcripción generalizada (Jensen et al., 2013), la información disponible acerca de sus funciones nos puede ayudar a inferir la utilidad de los transcritos aberrantes.

Actualmente se asume que la transcripción generalizada es una característica universal de los genomas eucariotas y procariotas, que se controla principalmente a nivel de terminación de la transcripción y que está acoplada a la degradación de RNA por el exosoma en el núcleo. Si no se regula, la transcripción ubicua puede resultar muy dañina para la integridad del genoma, por lo que las células emplean diferentes estrategias de vigilancia para evitar la acumulación de estos transcritos y desarrollan mecanismos de destrucción específicos que marcan los transcritos para su degradación, o atacan de manera inespecífica a cualquier molécula de RNA no percibida como funcional (Villa & Porrua, 2022).

INTRODUCCIÓN

A continuación se describen algunos de los factores identificados hasta la fecha que podrían estar relacionados con este fenómeno.

3.1. Factores implicados en el control de la transcripción generalizada

La transcripción generalizada se identificó por primera vez en la levadura *S. cerevisiae* cuando al mutar el factor de cromatina Spt6 y posteriormente Rrp6p (una de las subunidades catalíticas del exosoma), tuvo lugar la acumulación de los denominados transcritos crípticos inestables (*Cryptic Unstable Transcripts*, CUTs) (Figura 8B), que no se detectaban en condiciones fisiológicas en células de tipo silvestre (Davis & Ares, 2006; Houalla et al., 2006; Kaplan et al., 2003). En los mutantes Spt6, dicha acumulación ocurría dentro de las regiones codificantes, mientras que en los mutantes Rrp6p los transcritos provenían de regiones intergénicas. Posteriormente se han ido identificando varios tipos de transcritos aberrantes en su mayoría antisentido: transcritos estables no anotados (*Stable Unannotated Transcripts*, SUTs) (Xu et al., 2009), que son menos sensibles a la actividad del exosoma, y transcritos inestables sensibles a Xrn1 (*Xrn1-sensitive unstable transcripts*, XUTs), que son degradados por la 5'-3' exoribonucleasa Xrn1 en el citoplasma (van Dijk et al., 2011) (Figura A8B). En células humanas se han identificado transcritos cortos y altamente inestables denominados PROMPTs (*Promoter Upstream Transcript*) (Preker et al., 2008).

Los nucleosomas protegen a las células de la transcripción ubicua, actuando como una barrera física para la RNA polimerasa II facilitando la terminación de la transcripción (Hildreth et al., 2020); sin embargo, cada vez se están descubriendo más factores implicados a diferentes niveles capaces de impedir la acumulación de estos RNAs. En *S. cerevisiae* el complejo Nrd1p-Nab3p-Sen1p (NNS) es el mejor caracterizado. Este complejo se une a secuencias específicas de los

CUTs, promoviendo la terminación de la transcripción y el reclutamiento del exosoma nuclear para una rápida degradación de estos transcritos (Tudek et al., 2014). La presencia de secuencias ricas en AU muy próximas a Nrd1 y Nab3 es importante, posiblemente porque disminuyen la probabilidad de que los RNA diana adopten estructuras secundarias que impedirían su unión a Nrd1 y Nab3 (Villa & Porrua, 2022).

Por otro lado, el complejo conservado TRAMP (Trf4/5-Air2/1-Mtr4-*polyadenylation*) es el principal cofactor del exosoma nuclear en células eucariotas y desempeña un papel esencial en la degradación de los RNAs que son objetivo del exosoma. Recientemente, en levaduras se han identificado distintas formas del complejo TRAMP, cada una de las cuales tiene afinidad por un tipo de transcrito diferente (Delan-Forino et al., 2020). Otro factor importante es la proteína de unión al RNA Npl3, que se une a diversos sitios en un gran número de transcritos, y cuya ausencia da como resultado una transcripción extendida con efectos globales en la expresión génica (Holmes et al., 2015).

En ocasiones los XUTs y algunos SUTs pueden escapar de la degradación por la vía NNS y ser exportados al citoplasma (Malabat et al., 2015; Schulz et al., 2013; van Dijk et al., 2011). Sin embargo, debido a que carecen de información codificante, una vez en el citoplasma son reconocidos por la vía de degradación NMD que promueve su degradación por la exonucleasa 5'-3' Xrn1 (Malabat et al., 2015).

En mamíferos y *Drosophila*, recientemente se ha demostrado que el complejo de proteína *Integrator* es esencial para atenuar la transcripción generalizada. Este complejo escinde co-transcripcionalmente los transcritos de una manera independiente de secuencia para su posterior degradación en el núcleo por el exosoma (Lykke-Andersen et al., 2021; Tatomer et al., 2019).

En tripanosomas, el hecho de que se detecten transcritos antisentido en ausencia de marcas de cromatina específicas de TTS, sugiere que la

INTRODUCCIÓN

transcripción ubicua también podría existir, pero hasta la fecha no se han identificado los factores involucrados en este proceso.

En definitiva, la transcripción generalizada sucede en diversos organismos, procariotas y eucariotas, independientemente de cuál sea su organización genética o su regulación transcripcional. Se trata de un fenómeno extendido que puede otorgar una nueva modalidad de regulación de la expresión génica. No podemos olvidar que cualquiera que sea la estrategia diseñada por las células para controlar los niveles de estos transcritos debe de ser altamente eficiente, ya que la ausencia de regulación puede afectar a la integridad del transcriptoma así como a la estabilidad genómica.

3.2. Posibles consecuencias de la transcripción generalizada

Tanto en levaduras como en metazoos la RNA polimerasa II parece transcribir en dirección antisentido generando, como consecuencia, transcritos no codificantes. Una gran fracción de estos transcritos ubicuos ha permanecido “oculta” durante mucho tiempo debido a que la célula posee sistemas eficientes que restringen la acumulación de estas especies aberrantes. La importancia biológica de la transcripción generalizada es actualmente tema de debate. Existen varias hipótesis que defienden que podría tener la finalidad de originar nuevas unidades transcripcionales, entre las cuales algunas podrían conferir ventajas selectivas a través de la evolución (Baalsrud et al., 2018; Carvunis et al., 2012). Además, la transcripción antisentido es capaz de alterar los niveles de expresión de genes vecinos (Huber et al., 2016; Xu et al., 2011). Sin embargo, la transcripción generalizada es potencialmente dañina para la expresión y la integridad del genoma y, por lo tanto, todos los organismos han desarrollado diferentes mecanismos para mantener este proceso bajo control basados en la degradación de este tipo de transcritos.

Una de las principales consecuencias que surge a raíz de la estabilización de estos transcritos es la interferencia transcripcional, que normalmente reprime la expresión del gen anotado asociado en *cis* (Castelnuovo et al., 2013; Martens et al., 2004; Nadal-Ribelles et al., 2014; Pinskaya et al., 2009; Shearwin et al., 2005). En levaduras el silenciamiento de Nrd1, que forma parte del complejo NNS, desregula la síntesis de RNA por represión antisentido e interferencia transcripcional (Schulz et al., 2013). A veces, el mero acto de transcribir el RNA no codificante puede provocar la deposición de determinadas marcas represivas de cromatina en las regiones promotoras de los genes que codifican proteínas. Por ejemplo, en *S. cerevisiae* el RNA no codificante SRG1 interfiere con el promotor aguas abajo del gen de respuesta al estrés *SER3*, al bloquear la unión de factores de transcripción específicos, siendo el primer ejemplo descrito de interferencia transcripcional no codificante (Martens et al., 2004). El hecho de que la RNA polimerasa II invada el promotor de un gen vecino también puede contribuir a la represión, provocando directamente el desalojo físico del complejo de pre-iniciación (*Preinitiation Complex*, PIC).

En raras ocasiones la transcripción antisentido puede estimular la expresión génica. Por ejemplo, un CUT antisentido que se origina en el extremo 3' del gen *PHO5* favorece la transcripción del gen al incrementar la accesibilidad de la RNA polimerasa II al promotor, posiblemente contribuyendo al desalojo de histonas (Uhler et al., 2007). No obstante, los pares de transcritos sentido-antisentido generalmente están negativamente correlacionados (Neil et al., 2009; Xu et al., 2009), debido a la interferencia transcripcional o a la generación de modificaciones de histonas en la región codificante que inhiben la expresión génica.

En cualquier caso, asignar una función a estos transcritos resulta complicado debido a que no están conservados y es difícil discriminar experimentalmente entre una función directa del transcrito maduro en cuestión, su precursor o el proceso de transcripción que lo produce. De hecho, la mayoría tienen una vida media corta y son muy inestables, lo

INTRODUCCIÓN

que lleva a pensar que quizás no tengan un papel demasiado relevante e incluso que puedan llegar considerarse “ruido” de la transcripción. Sin embargo, algunos estudios en levaduras demuestran que estos transcritos se regulan en respuesta a diferentes condiciones de crecimiento o durante el desarrollo (Dutrow et al., 2008; Stolc et al., 2004), lo que sugiere que deben tener algún tipo de papel regulador. Por ejemplo, en levaduras la mayoría de los XUTs se acumulan en medios que contienen litio, lo que indica un posible papel en el establecimiento de una respuesta temprana al cambio en las condiciones del medio (van Dijk et al., 2011).

El hecho de que algunos de estos transcritos no canónicos se distribuyan de forma no aleatoria por el genoma, por ejemplo asociándose a promotores, como es el caso de los CUTs o de los PROMPTs, sugiere una posible función reguladora sobre los mensajeros situados aguas abajo (Davis & Ares, 2006). A varios CUTs se les han atribuido funciones relacionadas con la regulación de la modificación de histonas con efectos posteriores en la expresión génica, como por ejemplo en el locus *GAL10-GAL1* en *S. cerevisiae* (Houseley et al., 2008; Pinskaya et al., 2009): los transcritos no codificantes que se originan en la región codificante de *GAL10* en condiciones represivas para *GAL10-GAL1* son responsables de la trimetilación de la histona 3, lo que conduce a la inhibición del PIC, atenuando la inducción de *GAL1*. Otro ejemplo en este mismo organismo sería el transcrito antisentido *TY1* RNA, cuya estabilización desencadena el silenciamiento transcripcional en *trans* del retrotransposon Ty1 e impide su movilidad (Berretta et al., 2008).

Por otro lado, se ha demostrado que la transcripción más allá de los sitios de terminación es una fuente de diversidad transcriptómica en células cancerosas (Grosso et al., 2015). Además, la transcripción antisentido también se asocia con desórdenes neuronales, diabetes y enfermedades cardiovasculares (Barman et al., 2019), lo que resalta la importancia de este mecanismo de control. Independientemente del posible papel de algunos eventos de transcripción no codificante en la regulación génica, una

alteración de alguno de los componentes conduce a la desregulación de cientos de genes y, por lo tanto, al deterioro de múltiples funciones celulares que puede terminar siendo letal (Gill et al., 2020; Schulz et al., 2016).

En resumen, la transcripción generalizada es un hecho universal pero su función biológica, si es que existe, es controvertida. Varios indicios hacen sugerir una funcionalidad general de este tipo de transcripción no codificante, dotando de plasticidad al genoma y desafiando el concepto clásico de gen y de información genética (Villa & Porrua, 2022). Independientemente de cómo las células controlen los niveles de estos transcritos aberrantes, resulta evidente que la pérdida de un control estricto sobre la transcripción por parte de la RNA polimerasa II o la persistencia de estas especies de RNA en la célula, pueden ejercer un impacto negativo en varios aspectos de la regulación génica.

4. La proteína de unión a RNA RBP33

El control de la expresión génica a nivel postranscripcional es esencial en todos los organismos. Numerosos estudios han demostrado el importante papel de las proteínas de unión a RNA (*RNA Binding Proteins*, RBPs) como factores que actúan en *trans* en la regulación del procesamiento, el transporte, la estabilidad y traducción de los RNAs. La unión de estas proteínas a los elementos en *cis* presentes en las regiones codificantes y no codificantes de los mRNA, puede afectar a la estabilidad de la transcripción, la eficiencia de la traducción o ambas (Müller-McNicoll & Neugebauer, 2013).

En tripanosomátidos, las RBPs son factores reguladores clave ya que el control de la expresión génica se lleva a cabo esencialmente a nivel postranscripcional. Mediante estudios bioinformáticos y genéticos, se ha determinado la existencia de más de 150 RBPs en estos parásitos en base a los dominios de unión a RNA que contienen. En concreto, el motivo RRM

INTRODUCCIÓN

es uno de los dominios más comunes encontrados en este tipo de proteínas en eucariotas (Lorkovič, 2002) y normalmente consta de dos hélices alfa y cuatro hebras beta antiparalelas (Hoffman et al., 1991; Nagai et al., 1990).

Las proteínas que poseen estos motivos suelen estar implicadas en la biogénesis de los RNA ribosómicos, en *splicing*, en la incorporación de la caperuza a los mensajeros, en la exportación de los RNAs desde el núcleo hacia el citosol, en el procesamiento de los RNA de transferencia o en la traducción de mensajeros. Además, una gran cantidad de proteínas con dominios RRM desempeñan papeles fundamentales en la regulación de la degradación del RNA (de Gaudenzi et al., 2005; D'Orso & Frasch, 2001).

Tb927.8.990, conocida como RBP33, es una proteína de unión a RNA de 318 aminoácidos y un tamaño teórico de 34 KDa, que ha sido previamente caracterizada en nuestro laboratorio (Fernández-Moya et al., 2014). Posee un motivo RRM cerca de su extremo N-terminal, una secuencia de localización nuclear (*Nuclear Localization Sequence*, NLS) y los dos últimos tercios de la proteína no muestran similitud aparente con ninguna proteína conocida que no sea de tripanosomátidos (Figura 9A). El silenciamiento de la expresión de RBP33 conduce a una muerte celular rápida tanto en la forma sanguínea como en la procíclica (Fernández-Moya et al., 2014) (Figura 9B), mientras que su sobreexpresión provoca una rápida detención del crecimiento en sanguíneos, así como una disminución de la cantidad total de mRNA (Cirovic et al., 2017). Esta proteína se expresa al mismo nivel tanto en sanguíneos como en procíclicos (Fernández-Moya et al., 2014) y se localiza exclusivamente en unos 60 focos en el núcleo, fuera del nucleolo (Cirovic et al., 2017).

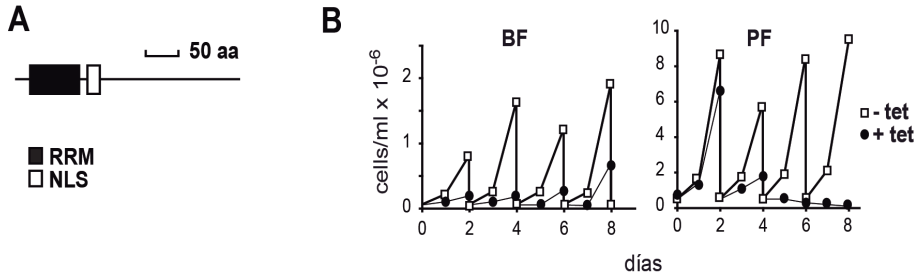
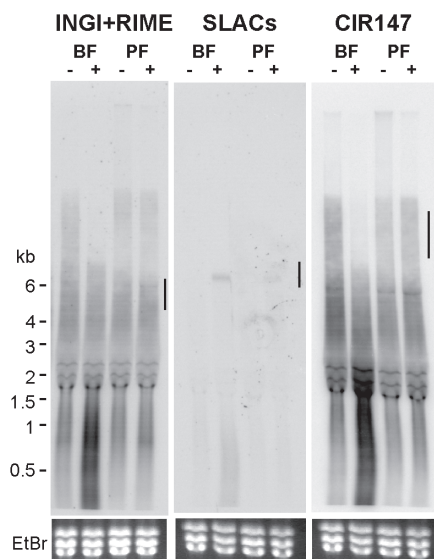


Figura 9. RBP33 es una proteína esencial. (A) Diagrama esquemático de la proteína RBP33 de *T. brucei*. (B) Curvas de crecimiento de las líneas celulares donde se silenció la expresión de RBP33 mediante RNAi inducible por tetraciclina. Las líneas celulares de RNAi se cultivaron en ausencia (cuadrados blancos) o presencia (círculos negros) de 1 mg/ml de tetraciclina para inducir el RNAi. Imagen adaptada de (Fernández-Moya et al., 2014).

RBP33 se une preferentemente a RNAs que provienen de zonas cSSRs y otras regiones genómicas silenciadas (Fernández-Moya et al., 2014). Además, el silenciamiento de RBP33 provoca un aumento de transcritos derivados de los retrotransposones INGI/RIME y SLACs, así como de repeticiones en tándem CIR147. Además, los transcritos de INGI/RIME y CIR147 de migración más lenta disminuyen en abundancia en ausencia de RBP33 en tripanosomas sanguíneos, lo que sugiere una posible implicación de esta proteína en la maduración del RNA (Figura 10) (Fernández-Moya et al., 2014).



INTRODUCCIÓN

Figura 10. Efecto del silenciamiento de RBP33 en los niveles de transcritos derivados de retrotransposones y repeticiones CIR147. Los transcritos derivados de los retrotransposones INGI/RIME y SLAC o de CIR147 se detectaron mediante northern con sondas específicas. Como control de carga se utilizó la tinción con bromuro de etidio. Imagen tomada de (Fernández-Moya et al., 2014).

Estos resultados sugieren que RBP33 podría jugar algún papel en el control de la abundancia de este tipo de transcritos. Esta posibilidad ha sido explorada durante la presente tesis doctoral.

OBJETIVOS

OBJETIVOS

El objetivo general de la presente tesis doctoral fue dilucidar el papel de la proteína de unión a RNA RBP33 en la regulación de la abundancia de transcritos generados a partir de zonas silenciadas en el genoma de *T. brucei*.

Para abordar esta idea, se establecieron los siguientes objetivos específicos:

1. Estudiar los cambios globales que ocurren en el transcriptoma y en la accesibilidad de la cromatina tras silenciar la expresión de RBP33.
2. Identificar proteínas asociadas a RBP33.
3. Analizar la estabilidad de los transcritos regulados por RBP33.

MATERIALES Y MÉTODOS

MATERIALES Y MÉTODOS

1. Materiales

1.1. Líneas celulares de *Trypanosoma brucei*

Las líneas parentales de *T. brucei* utilizadas fueron S16 ‘*Single marker*’ para sanguíneos (Wirtz et al., 1999) y 449 para procíclicos (Biebinger et al., 1997). Las líneas derivadas se indican primero con el nombre de la línea parental seguida del número correspondiente al plásmido transfectedo. La línea S16 tiene integrados en su genoma genes que codifican la RNA polimerasa del fago T7 así como represor inducible por tetraciclina (*TetR*) en su genoma, mientras que la línea 449 solo posee el gen *TetR*. Esto permite la expresión inducible de genes a voluntad mediante la adición de tetraciclina al medio de cultivo.

1.2. Oligonucleótidos

Todos los oligonucleótidos utilizados en esta tesis fueron sintetizados por Metabion international AG (Condalab).

Nombre	Secuencia 5’-3’	Propósito
AE1061	ATCCTATTTGTCGTAACAATCTCTCAAGC	Clonaje de la secuencia Tb427_110143800 para la expresión de eYFP
AE1039	GCACCGGTCTGAACCAACCGATTATTGTTTC	Clonaje de la secuencia Tb427_110143800 para la expresión de eYFP. AgeI
AE1040	AACCTGTGTCTTTTCCTTTAAC	Genotipado de la inserción de pGR431 en el locus Tb427_110143800 Hibrida en el 5’-UTR de Tb427_110143800

MATERIALES Y MÉTODOS

Nombre	Secuencia 5'-3'	Propósito
AE1019	CAACCACTACCTGAGCTACCAG	Genotipado de la inserción de pGR431 y pGR433. Hibrida en el eYFP ORF
AE331	CGAGTCAGTGAGCGAGGAAGC	Genotipado de la inserción de pGR431 y pGR433. Hibrida en la región <i>E. coli</i> ori del plásmido
AE1041	CGGAGAACTCTTTATCCCAAAG	Genotipado de la inserción de pGR431 en el locus Tb427_110143800 Hibrida en el 3'-UTR de Tb427_110143800
AE1042	GCCGAATTTATCAACTATGCATATTAGGG-CACAG	Clonaje de la secuencia Tb427_100047800 para la expresión de eYFP
AE1043	GCACCGGTTTTTTAGCGCAATAAGCCGG-GATACC	Clonaje de la secuencia Tb427_100047800 para la expresión de eYFP. Agel
AE1044	CGCTTCTGCTTTATGTCTC	Genotipado de la inserción de pGR433 en el locus Tb427_100047800 Hibrida en el 5'-UTR de Tb427_100047800
AE1081	GTTTGAAAAGCGGCAAACAC	Genotipado de la inserción de pGR433 en el locus Tb427_100047800 Hibrida en el 3'-UTR de Tb427_100047800
AE962	TGAAAAATGAAAGACCAAGAGGCGACGC	qPCR de cSSR10:6
AE961	CGCGGAGAAATTGAACTTAACCCTAAGAG	
AE1069	GTTGCTTAGGCATGGTGTTG	qPCR de cSSR11:3
AE1070	CAGGTGTTGAATGGTCGTTTG	

Nombre	Secuencia 5'-3'	Propósito
AE1071	AAACTAAGCGAGGGTGCAATTTG	qPCR de cSSR3:3
AE1072	CTGCTTCCAATTTCTGTATTTTC	
AE1054	ACTTGTCATGTGCAACATCG	qPCR de Tb427_110130400
AE1055	CTGAATTCTCCGATTGCAACC	
AE1052	GTTTTGTTTCCGCCAGCCTC	qPCR de Tb427_060007400
AE1053	GTGCCCAAGGGGAACTTTT	
AE1046	CATAGGCTCACATTCTCTGAG	qPCR de Tb427VSG-531 (Tb427_00304300)
AE1058	GTCCGTCAAGTTCTGATGTGAGC	
AE1048	TTACGCTAGCGTTCATACTAGC	qPCR de Tb427VSG-653 (Tb427_000173600)
AE1049	GAAATAAGCCGCAACATCGCATATC	
AE1175	GAGCTGGACGGCGACGTAAAC	qPCR de <i>eYFP</i>
AE1162	GTAGCGGGCGAAGCACTGC	

Tabla 1. Listado de oligonucleótidos utilizados en este trabajo.

1.3. Plásmidos

Plásmido	Inserto	Propósito	Plásmido parental	Antibiótico de selección
pGR70 (Fernández-Moya et al., 2014)	Mitad carboxilo de RBP33	RNAi	pGR19 (Clayton et al., 2005)	HYG
pGR417	ORF <i>eYFP</i>	Expresión de <i>eYFP</i> en orientación inversa	pGR406*	BLA

MATERIALES Y MÉTODOS

Plásmido	Inserto	Propósito	Plásmido parental	Antibiótico de selección
pGR431	Parte del ORF de Tb427_110143800	Expresión de eYFP en orientación inversa en el locus Tb427_110143800	pGR417 (esta tesis)	BLA
pGR433	Parte del ORF de Tb427_100047800	Expresión de eYFP en orientación inversa en el locus Tb427_100047800	pGR417 (esta tesis)	BLA
pGR149 (Gómez-Liñán et al., 2022)	ORF RBP33	N-TAP	pHD1360 (Häusler & Clayton, 1996)	HYG
pGR12 (Rico-Jiménez et al., 2021)	-	TAP control	pHD1360 (Häusler & Clayton, 1996)	BLA
pHD1157	RRP44	RNAi	(Estévez et al., 2001)	HYG
pHD1165	RRP6	RNAi	(Estévez et al., 2001)	HYG

Tabla 2. Listado de plásmidos utilizados en este trabajo

*pGR406 es pGR12 donde la luciferasa ha sido reemplazada por eYFP y el espaciador ribosomal por Tb407.04.3280

1.4. Anticuerpos

Anticuerpos primarios			
Nombre	Especie	Origen	Dilución utilizada en inmunoblot
Anti-RBP33	Conejo	(Fernández-Moya et al., 2014)	1:10.000
Anti-RRP6	Conejo	(Haile et al., 2003)	1:600
Anti- RRP44	Conejo	(Estévez et al., 2001)	1:100
Anti-Tubulina	Ratón	Sigma (clon B-5-1-2)	1:20.000
Anticuerpos secundarios comerciales			
Anti-IgG de conejo HRP	Cabra	Promega	1:5.000
Anti-IgG de ratón HRP	Cabra	Promega	1:5.000

Tabla 3. Listado de anticuerpos utilizados en esta tesis para ensayos de inmunoblot.

1.5. Medios y tampones utilizados

Todos los medios y los tampones fueron preparados con agua Milli-Q

Agar LB	1% triptona, 0.5% extracto de levadura, 1%NaCl y 1.5% agar
Citomix	2mM EGTA, 120mM KCl, 0.15mM CaCl ₂ ·2H ₂ O, 10mM fosfato potásico pH7.6, 25 mM Hepes, 5mM MgCl ₂ ·6 H ₂ O, 0.5% glucosa, 100 µg/ml BSA y 1mM hipoxantina

MATERIALES Y MÉTODOS

HMI-9	IMDM (sigma) 1.77% (p/v), 36mM NaHCO ₃ , 1mM hipoxantina, 161mM timidina, 1.5mM cisteína, 78mM batocuproina, 1mM piruvato sódico y 2µM β-mercaptoetanol. Ajustar pH a 7.5, filtrar y añadir 10% SBF.
SDM-79	MEM (GIBCO) 0.7% (p/v); medio 199 (GIBCO) 0.2 (p/v), aminoácidos esenciales MEM 50X (Sigma) 0.8% (p/v); aminoácidos no esenciales MEM 100X (Sigma) 0.6% (p/v); 5.6 mM glucosa, 30 mM Hepes, 24 mM MOPS, 24 mM NaHCO ₃ , 1mM piruvato sódico, 2.24 mM D-L-alanina, 574 µM L-arginina, 2 mM L-glutamina, 469 µM D-L-metionina, 484 µM L-fenilalanina, 5.2 mM L-prolina, 571 µM D-L-serina, 1.28 mM taurina, 2.94 mM D-L-treonina, 552 µM tirosina, 35 µM guanosina, 9 µM ácido fólico, 232 µM D(+) glucosamina, 14,6 µM ácido p-aminobenzoico, 800 µM biotina. Filtrar y añadir BSF al 10% y 11.5 µM hemina.
Citomix	2mM EGTA, 120mM KCl, 0.15mM CaCl ₂ ·2H ₂ O, 10mM fosfato potásico pH7.6, 25 mM Hepes, 5mM MgCl ₂ ·6 H ₂ O, 0.5% glucosa, 100 µg/ml BSA y 1mM hipoxantina
Laemli 5X	0.625 M Tris-HCl pH 6.8, 10% (p/v) SDS, 40% (p/v) glicerol, 7.5% (v/v) β-mercaptoetanol, 1.5% azul de bromofenol
PBS	10mM Na ₂ HPO ₄ y 1.8 mM de KH ₂ PO ₄ pH7.4; 140 mM NaCl, 2.7 mM KCl
PBST	PBS + 0.1% Tween 20
Tampón de electroforesis SDS-PAGE 10X	0.25 mM Tris, 1.92 glicina y 1% SDS
Tampón de transferencia 10X	0.25 mM Tris, 1.92 glicina y 20% metanol
Tampón de lisis TAP	10 mM Tris-HCl pH 7.4, 1mM DTT, 0.1 % Igepal, inhibidor de proteasas 1X (Roche)
IPP-150	10mM Tris-HCl pH7.4, 150mM NaCl y 0.1% Igepal

TEV buffer	10 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% Igepal 0.5 mM EDTA
TCA-DOC	Disolver 40 mg de sodio desoxicolato en 10 ml de 100% Tricloroacético (TCA)

1.6. Antibióticos

Antibiótico	Abreviatura	Stock	BF	PF	<i>E. coli</i>
Higromicina	Hyg	50 mg/mL	5 µg/mL	50 µg/mL	-
Blasticidina	Bla	10 mg/mL	5 µg/mL	10 µg/mL	-
Geneticina	G418	2.5 mg/mL	2.5 µg/mL	-	-
Tetraciclina	Tet	1 mg/mL	1 µg/mL	1 µg/mL	-
Fleomicina	Fleo	20 mg/mL	-	1 µg/mL	-
Ampicilina	Amp	100 mg/mL	-	-	100 µg/ mL

Tabla 4. Concentraciones de antibióticos utilizadas en esta tesis. BF, tripanosomas sanguíneos; PF, tripanosomas procíclicos.

1.7. Programas informáticos y bases de datos

El análisis de datos de secuenciación masiva, RNA-seq y ATAC-seq, se describen en los apartados 2.14. y 2.16. de la sección de materiales y métodos, respectivamente.

1.7.1. Programas informáticos

Adobe Photoshop CS5.1 versión 12.1

Adobe Illustrator CS5.1 versión 15.1.0

FlowJo versión 10.0

MATERIALES Y MÉTODOS

Gene construction kit version 4.0

Microsoft Office 2011 (Word, Excel, Power Point)

Google Chrome

BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

Oligocalc (<http://biotools.nubic.northwestern.edu/OligoCalc.html>)

DISOPRED3 (<http://bioinf.cs.ucl.ac.uk/psipred>)

1.7.2. Bases de datos

NCBI (<https://www.ncbi.nlm.nih.gov/>)

TriTrypDB (<https://tritrypdb.org/tritrypdb/app>)

Eukaryotic Linear Motif (ELM) (<http://elm.eu.org/index.html>)

2. Métodos

2.1. Cultivo de *Trypanosoma brucei*

La forma sanguínea de *T. brucei* fue cultivada en medio HMI-9 (Hirumi & Hirumi, 1989) previamente suplementado con 10% de suero bovino fetal inactivado, en un incubador a 37°C en presencia de 5% de CO₂. La forma procíclica se cultivó en medio SDM-79 (Brun et al., 1979) previamente suplementado con 10% de suero bovino fetal inactivado y hemina (2.5 mg/mL), en un incubador a 28°C. Para su mantenimiento en fase exponencial, los tripanosomas se diluyeron en medio fresco, previamente calentado a la temperatura adecuada, cada 48 horas. La densidad celular se determinó contando los tripanosomas en un Coulter Counter (Beckman).

2.2. Transfección estable de tripanosomas mediante nucleofección

Para la transfección de tripanosomas se utilizó el mismo protocolo con algunas modificaciones según se tratara de tripanosomas sanguíneos o procíclicos.

2.2.1. Tripanosomas sanguíneos

Se recogieron un total de 40×10^6 tripanosomas sanguíneos (por cada transfección) fase exponencial a 1000g durante 10 min a temperatura ambiente y se resuspendieron en 100 μ L de citomix a 37°C. Los 100 μ L se transfirieron a una cubeta estéril junto con el DNA plasmídico linearizado (1 μ g/ μ L). La transfección se llevó a cabo en un Nucleofector Amaxa utilizando el programa X-001 (Burkard et al., 2007). Inmediatamente después, las células se transfirieron a 10 mL de HMI-9 y se diluyeron 1:10. La dilución se repartió en placas de 24 pocillos y, tras incubar durante 24 horas a 37°C, se añadió el antibiótico de selección correspondiente a la concentración apropiada. Los transfectantes fueron visibles a partir del día 5-7.

2.2.2. Tripanosomas procíclicos

Se recogieron un total de 40×10^6 (por cada transfección) de tripanosomas procíclicos en fase exponencial a 1000g durante 10 min a temperatura ambiente y se resuspendieron en 100 μ L citomix a 4°C. Los 100 μ L se transfirieron a una cubeta esteril junto con el DNA plasmídico linearizado (1 μ g/ μ L). La transfección se llevó a cabo en un Nucleofector Amaxa utilizando el programa X-014 (Burkard et al., 2007). Inmediatamente después, las células se transfirieron a 10 mL de SDM-79 suplementado al 20% de SBF. Tras su incubación durante 24 horas a 28°C sin selección, se diluyeron 1:10 y 1:100. Las diluciones se repartieron en placas de 96 pocillos y se añadió el antibiótico de selección correspondiente a la concentración apropiada. Los transfectantes fueron visibles a partir del décimo día.

2.3. Congelación y descongelación de tripanosomas

Las células en fase exponencial se centrifugaron a 1000g durante 10 min a temperatura ambiente para su congelación. Las células se resuspendieron en medio de cultivo con 10% (v/v) de glicerol en criotubos y se congelaron lentamente hasta los -80°C. Para su conservación durante largos períodos de tiempo, se transfirieron a nitrógeno líquido. Para su descongelación, las células se dejaron a temperatura ambiente hasta su total descongelación y se transfirieron rápidamente a medio aatemperado (37°C para sanguíneos y 28 °C para procíclicos).

2.4. Ribointerferencia (RNAi)

Los cultivos transfectados con los plásmidos RNAi fueron seleccionados con los marcadores de resistencia correspondientes. El cultivo se dividió en dos subcultivos y uno de ellos fue inducido con 1 µg/mL de tetraciclina. Se comparó el crecimiento a las 48 horas y se seleccionaron aquellos clones donde el crecimiento se vio afectado para su confirmación por inmunoblot.

2.5. Clonaje de vectores

Los protocolos de corte de DNA con enzimas de restricción, ligación y desfosforilación de fragmentos de DNA utilizados en esta tesis están descritos en (Sambrook et al., 1989).

2.6. Aislamiento de DNA genómico en *T. brucei*

Para el aislamiento del DNA genómico de cultivos de *T. brucei*, se recogieron 10-20x 10⁶ células mediante centrifugación a 1400g durante 10 min a temperatura ambiente y el pellet se resuspendió en 200 µl de DNAzol. A continuación se añadieron 100 µl de etanol absoluto, se dejó 5 minutos

a -20°C y se centrifugó 20 min a 4°C a máxima velocidad. Posteriormente se lavó dos veces con 70% etanol, se dejó secar y el pellet se resuspendió en 100 µl de agua Milli-Q. Se dejó disolver durante toda la noche a 4°C.

2.7. Preparación de células competentes de *E. coli*

Se utilizó la cepa de *E. coli* DH5α preparadas siguiendo el protocolo descrito por (Inoue et al., 1990).

2.8. Aislamiento de DNA plasmídico de *E. coli*

Se utilizó el kit de purificación de plásmidos de Macherey-Nagel (Alemania) siguiendo las instrucciones recomendadas por el fabricante.

2.9. Transformación de *E. coli* por choque térmico

Para su transformación con plásmidos de DNA, las células de *E. coli* se pusieron en hielo y se añadieron 3 µL de la mezcla de la ligación (de un total de 10 µL). Se incubaron en hielo durante 25-30 min y se llevó a cabo el choque térmico introduciéndolas primero en un baño a 42°C durante 30 segundos y, posteriormente, devolviéndolas al hielo durante 2 min. Las células se plaquearon directamente en placas de agar LB con ampicilina (atemperadas a 37 °C) y se incubaron durante toda la noche a 37 °C.

2.10. Amplificación de DNA por PCR

El DNA se amplificó por la reacción en cadena de la polimerasa, (*polymerase chain reaction*, PCR) utilizando las polimerasas *Phusion™ High-Fidelity* siguiendo el protocolo recomendado por el fabricante. Los productos de PCR se purificaron con el kit de purificación de PCRs de Macherey-Nagel (Alemania). Los oligonucleótidos se diseñaron utilizando el programa Gene

MATERIALES Y MÉTODOS

Construction kit 4.0 y OligoCalc.

2.11. Precipitación de DNA

Se añadieron 0.1 volúmenes de 3M acetato sódico pH 7.0, 2.5 volúmenes de etanol y 1µL de glucógeno a la mezcla de DNA digerido para su precipitación. Incubamos a -20°C durante 30 min y centrifugamos a 16.000g durante 30 min a 4°C. El pellet de DNA se lavó con etanol al 70%, se dejó secar al aire y se resuspendió en agua Milli-Q.

2.12. Extracción de RNA de cultivos de *T. brucei*

Se centrifugaron las células en crecimiento exponencial a 1400g durante 10 min a temperatura ambiente. El pellet se resuspendió en 350 µL del buffer RA1 del kit de extracción Nucleospin RNA de de Macherey-Nagel (Alemania) siguiendo el protocolo descrito por el fabricante. Las muestras se almacenaron a -80°C.

2.13. RT-PCR cuantitativa

El RNA extraído se cuantificó mediante NanoDrop® system. El DNA complementario (cDNA) fue sintetizado a partir de 250 ng de RNA utilizando 0.5 µg de hexámeros (Invitrogen) y 0.25 mM de cada dNTP, en un volumen final de 14.5 µL. Las muestras se incubaron 5 min a 65°C y se les añadió 5x Buffer RT (Thermo Scientific), 20 unidades de Ribolock y 200 unidades de Maxima reverse transcriptasa (Thermo Scientific) siendo el volumen final 20 µL. Se incubó 5 min a temperatura ambiente, 30 min a 50 °C y finalmente, 5 min a 85°C.

El ensayo de PCR cuantitativa se realizó en un termociclador *CFX96 (Bio-Rad)* en placas de 96 pocillos (Thermo Scientific). La reacción contenía 0.5 µL de cDNA, 0.5 µM de cada oligo específico, 1X de Maxima

SYBR Green qPCR Master Mix (Thermo Scientific) y agua Milli-Q hasta un volumen final de 10 μ L. Las condiciones del termociclador fueron: 50°C durante 2 min y 95°C 10 min, seguido de 40 ciclos [95°C, 15s; 55°C, 1 min]. Se realizó un análisis de curvas de *melting* para confirmar un solo amplicón para cada mRNA probado.

Los cambios en la expresión se calcularon usando el método de $2^{-\Delta\Delta CT}$ (Livak & Schmittgen, 2001) utilizando los mRNAs de *Actina* o *ZFP2* como referencia para normalizar. Se realizaron al menos tres réplicas biológicas independientes.

2.14. Secuenciación masiva de RNA (RNA-seq)

El RNA de las muestras control (S16 o 449 parentales) y de las muestras inducidas por RNAi (transfectadas con el plásmido pGR70) se extrajo utilizando el kit de Rneasy Mini (Qiagen). Las librerías de RNA se prepararon a partir de transcritos poliadenilados previamente seleccionados utilizando el protocolo estándar de preparación de muestras de TrueSeq Stranded (Illumina). Los triplicados biológicos se secuenciaron en la Unidad de Genómica del IPBLN-CSIC (Granada, España) a través de la plataforma NextSeq 500 (Illumina).

La calidad de las secuencias generadas de 76 nt se verificó mediante FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Las librerías fueron submuestreadas utilizando seqtk (version 1.2, <https://github.com/lh3/seqtk>), para que las diferencias entre los tamaños de muestra fueran menores de un 10%, y luego se alinearon con el genoma de *T. brucei* Lister 427 2018 (v4.9) utilizando el programa '*align*' del paquete Subread [Rsubread versión 1.34.7(Liao et al., 2013)]. Solo se consideraron las lecturas mapeadas de forma única. La asignación de lecturas a genes se realizó utilizando el programa '*featureCounts*' del paquete subread [version 1.5.0-p, (Liao et al., 2014)] con las opciones `-p -B -C`. El archivo GFF con las anotaciones de los genes fue TriTrypDB-49_TbruceiLister427_2018.

MATERIALES Y MÉTODOS

gff (descargado de <https://tritrypdb.org/tritrypdb/app/downloads/>) que se complementó con 126 regiones de cambio de hebra (*strand-switch regions*) anotadas manualmente asignadas de forma arbitraria a la cadena Watson (o cadena +). Los perfiles de recuento de lecturas antisentido y sentido se generaron de forma independiente (opción `-s` establecida en 1 o 2 en `featureCounts`, respectivamente) y luego se combinaron y analizaron para determinar la expresión diferencial usando `edgeR` [versión 3.36.0, (Robinson et al., 2010)]. Solo se consideraron los genes que contenían más de un recuento por millón de lecturas mapeadas (CPM) en al menos tres muestras.

Para la correlación de Pearson y el análisis de componentes principales (PCA), los CPM se normalizaron mediante el método TMM en `edgeR` (Robinson et al., 2010) y se transformaron a \log_2 en `edgeR`. Los gráficos PCA se generaron utilizando el módulo `sklearn.decomposition` (versión 0.24.2, (Pedregosa et al., 2012)). La amplitud de cobertura se calculó a partir de los valores de profundidad de secuenciación obtenidos utilizando el módulo `'depth'` de `samtools` (versión 1.9, (Danecek et al., 2021) con opciones `-d 0 -aa`. El enriquecimiento posicional de los transcritos regulados se analizó mediante pruebas exactas de Fisher de cola derecha. Los valores de *p* se corrigieron por *false discovery rate* de acuerdo con el método de Benjamini-Hochberg ((Benjamini et al., 2001). Para los gráficos de cobertura, las regiones de interés se agruparon y contaron (ventana deslizante, 100 pb; tamaño de paso, 10 pb) utilizando el módulo `'countReadsPerBin'` de `deepTools` (versión 3.5.0, (Ramírez et al., 2016)) y se corrigieron por el tamaño de la librería. Para mapear los sitios aceptores de *splicing* (SASs), se seleccionaron lecturas que contenían al menos los últimos 14 nucleótidos de la secuencia del *spliced-leader* (SL), que son únicos y no aparecen en otro sitio del genoma (Siegel et al., 2010). Se eliminó la secuencia de SL, las lecturas se alinearon con el genoma de *T. brucei* 427 como se describió anteriormente, y se asignaron los sitios aceptores a la primera la primera posición de todas las lecturas que contenían SL mapeadas de forma única. Solo aquellos SASs

respaldados por un mínimo de cinco lecturas en al menos dos réplicas fueron considerados para su posterior análisis.

Los datos externos correspondientes a una línea celular “triple knock-out” que carece de las marcas de cromatina H3v, H4v y base J (Kim, 2021), así como los datos del control de la línea celular parental, se descargaron del *European Nucleotide Archive* (proyecto PRJNA727846; muestras SRR9088967, SRR9088968, SRR9088969, SRR9088997, SRR9088998 y SRR9088999) y se analizaron de la forma descrita anteriormente.

2.15. Citometría de flujo

La citometría de flujo se llevó a cabo con 1×10^6 células de cultivos de tripanosomas sanguíneos control, no inducidos o inducidos con tetraciclina. Las células se centrifugaron a 1000g, 10 min a temperatura ambiente y se resuspendieron en PBS atemperado y suplementado con 20mM de glucosa atemperado. La fluorescencia de eYFP se midió en un citómetro FaCSCalibur (Becton Dickinson) equipado con un láser de 488 nm, y se registraron 20.000 eventos por muestra. Los datos se adquirieron con el software CellQuest Pro y se analizaron con el software FlowJo (versión 10).

2.16. Ensayo de cromatina accesible por transposasa con secuenciación masiva (ATAC-seq)

Para los ensayos de ATAC-seq, llevamos a cabo el método descrito por Müller et al., (2018), adaptado de la siguiente manera: 1×10^7 tripanosomas sanguíneos parentales (S16) o inducidos por RNAi (transfectados con el plásmido pGR70) se lavaron con PBS frío y se resuspendieron en 100 μ l de tampón de permeabilización (10 mM Tris, pH 8,0, 100 mM KCl, 25 mM EDTA e inhibidores de proteasas libres de EDTA). Se añadió digitonina (Millipore

MATERIALES Y MÉTODOS

300410) a una concentración final de 40 μ M (de un stock 4 mM recién preparado en tampón de permeabilización) y la suspensión se incubó durante 5 min a temperatura ambiente. A continuación, las células se centrifugaron a 4°C durante 5 min a 1400g, se lavaron en 600 μ l de tampón isotónico frío (10 mM Tris pH 8,0, 100 mM KCl, glicerol al 5%, 1 mM DTT, 1 mM PMSF), se resuspendieron en 50 μ l de reacción de transposición (25 μ l de tampón 2X TD, 22,5 μ l de agua libre de nucleasas, 2,5 μ l de transposasa Tn5 del kit Illumina Nextera FC-121-1030) y se incubaron durante 30 min a 37 °C en agitación. Como control, una muestra que contenía 200 ng de DNA genómico se incubó con la reacción de transposición durante 5 min a 55 °C en las mismas condiciones, excepto que se utilizaron 1,25 μ l de transposasa. Las reacciones se purificaron utilizando el kit Qiagen MiniElute y se eluyeron en 11 μ l de Tris 10 mM, pH 8,0. La amplificación de las librerías se realizó con 2X KAPA HiFi mix y 1.25 μ M de oligonucleótidos de Nextera durante 7 ciclos; el número de ciclos se determinó mediante PCR cuantitativa utilizando las condiciones descritas en (Buenrostro et al., 2015). La selección de fragmentos se realizó con el kit *NucleoMag for clean up and size selection of NGS library prep reactions* (Macherey-Nagel 744970) siguiendo un esquema de selección de tamaño doble (primera ronda, volumen de bolitas = 0,6x volumen de muestra; segunda ronda, volumen de bolitas = 0,7x volumen de muestra; volumen de elución final = 17 μ l). La selección del tamaño de la muestra genómica se realizó en un solo paso utilizando un volumen de bolitas igual al de la muestra. Los triplicados biológicos se secuenciaron como se describió en el apartado 2.14. de materiales y métodos.

Las lecturas se alinearon con el genoma de *T. brucei* Lister 427 2018 (v4.9) usando bowtie 2.0 (Langmead & Salzberg, 2012) con las opciones --no-unal --no-mixed -5 5 -3 5. Las lecturas correspondientes a regiones libres de nucleosomas (*nucleosome-free regions, NFR*) fueron seleccionadas usando el módulo 'alignmentSieve' de deeptools (Ramírez et al., 2016) con las opciones --minMappingQuality 30 --maxFragmentLength 100 -- ATACshift --ignoreDuplicates; también se incluyó la opción -bl para eliminar las lecturas correspondientes al DNA

mitochondrial (unitig_851_maxicircle_Tb427v10). Para la correlación de Pearson y PCA, se obtuvieron coberturas de lectura de todo el genoma (tamaño = 100 pb) o de genes utilizando el módulo 'multiBamSummary' del paquete de programas 'deeptools' (Ramírez et al., 2016), normalizadas para generar señal de ATAC, y transformadas a log2. Para el PCA, la señal de ATAC se normalizó además respecto al control de DNA genómico.

Los gráficos con resolución de una base se generaron utilizando el módulo 'countReadsPerBin' del paquete de programas 'deeptools' (Ramírez et al., 2016) y se normalizaron según el tamaño de la librería. Para determinar los picos de lecturas libres de nucleosomas se utilizó MACS2 [versión 2.2.7.1 (Zhang2008)] a través del módulo 'callpeak' con los siguientes parámetros: -g 3.5e7 --nomodel -f BAMPE -B --keep-dup all. Como control se utilizó el DNA genómico. Los datos externos correspondientes a tripanosomas sanguíneos que carecen de las marcas de cromatina H3v y H4v (Müller et al., 2018), así como los datos de la línea celular parental y las muestras de DNA genómico, se descargaron del *European Nucleotide Archive* (proyecto PRJNA487259; muestras SRR7739654, SRR7739658, SRR7739651, SRR7739655, SRR7739659 y SRR7739660) y se analizaron de la misma forma.

2.17. Ensayos de vida media

Para determinar la vida media de los RNA mensajeros, se incubaron tripanosomas procíclicos control o inducidos por RNAi (transfectados con el plásmido pGR70) con 10 µg/ml de actinomicina D (a partir de una solución 5 mg/ml disuelta en DMSO) y 2 µg/ml de sinefungina (a partir de una solución 2 mg/ml disuelta en agua) para inhibir la transcripción y el *trans-splicing* respectivamente (C.-H. Li et al., 2006). Se añadió sinefungina 5 minutos antes que la actinomicina D. El RNA total se obtuvo a partir de muestras tomadas en diferentes momentos hasta el trascurso total de 1 hora utilizando el protocolo descrito en el apartado de materiales y métodos 2.12. La abundancia de transcritos se midió mediante RT-PCR cuantitativa

MATERIALES Y MÉTODOS

y se normalizó con *ZFP2* (Tb427_110167900/Tb927.11.14950), un transcrito que es estable a lo largo del tiempo establecido en los ensayos (Fadda et al., 2014). La constante de velocidad de degradación k , se obtuvo después de ajustar los datos a una cinética de primer orden mediante la ecuación $[RNA]_t = [RNA]_0 \cdot e^{-kt}$. Los valores de vida media se calcularon como $\ln(2)/k$ (Chen et al., 2008).

2.18. Preparación de lisados de proteínas

Para la obtención de extractos totales de tripanosomas procíclicos y sanguíneos, las células se centrifugaron en fase de crecimiento exponencial a 1400g, durante 10 minutos a temperatura ambiente; el pellet se lavó con PBS, se resuspendió en buffer Laemmli 1X y se calentó a 95°C durante 10 minutos. Las muestras se cargaron en geles de poliacrilamida. Como marcador se utilizó Benchmark™ (Invitrogen) para determinar el tamaño.

2.19. Inmunoblot

Se utilizaron geles al 8-10% SDS-PAGE (Sambrook & Russell, 2006) de 0.75 mm de grosor, en un sistema de electroforesis Protean (Bio-Rad) a 200V durante 50 min. Para la transferencia de las muestras, se utilizaron membranas de nitrocelulosa (Millipore) con un sistema de transferencia Mini-Protean (Bio-Rad) a 100V, durante 45 min en frío.

Una vez finalizada la transferencia, la membrana se tiñó con Ponceau durante 5 minutos para comprobar la calidad de la electroforesis y de la transferencia. La membrana se bloqueó durante 30 min con PBS suplementado con 0.05% de Tween 20 (p/v) y 5% de leche en polvo (p/v, Sveltesse). Se incubó 1h 30 min con los anticuerpos primarios diluidos en PBST-leche y se lavó 3 veces con PBST. Se incubó con el anticuerpo secundario diluido en PBST-leche durante 1h y 30 min, se lavó de nuevo 3 veces y se reveló con agentes de quimioluminiscencia en una solución preparada en el laboratorio

compuesta por 10 ml de solución A (250 mg/L luminol disueltos en 100mM de Tris-Hcl pH 8.5), 100 µl de solución B (180 mg 4-yodofenol disueltos en 20 ml de DMSO) y 4µl de 30% peróxido de hidrógeno (MERCK).

2.20. Purificación por afinidad en tándem (TAP)

Se utilizó una línea celular procíclica que expresaba el TAP-RBP33 en el locus endógeno generada anteriormente en el laboratorio por Sandra Fernández Moya, en su tesis doctoral titulada 'Caracterización funcional de las proteínas de unión a RNA RBP33 y DRBD3 como reguladoras de la expresión génica de *Trypanosoma brucei*, 2013 (resultado publicado en Gómez-Liñán et al., 2022).

Los complejos proteicos se purificaron utilizando el método TAP (Puig et al, 2001) con las modificaciones descritas en (Fernández-Moya et al, 2012) y que se detallan a continuación. Se recogieron 1×10^{10} células en fase logarítmica de crecimiento y se lavaron con PBS frío. El pellet se resuspendió en 10 ml de tampón de lisis (10mM Tris-Hcl pH7.6, 2mM DTT, 0.1% (p/v) Igepal CA-630 e inhibidores de proteasas de Roche) y se pasó 3 veces por una aguja 27G, en hielo, para obtener el extracto proteico. El lisado se centrifugó a 16.000g durante 10 min a 4°C. Se añadió NaCl a una concentración final de 150 mM. Se llevó a cabo el procedimiento de purificación tal y como se describe en (Puig et al., 2001) con las siguientes modificaciones: se digirió con la cisteín-proteasa TEV (*Tobacco Etch Virus*) durante toda la noche a 4°C y la unión a la resina calmodulina se llevó a cabo durante 4 horas a 4°C. Para la elución de los complejos de proteínas se utilizó 1 ml de tampón de elución (10 mM Tris-HCl pH7.4, 150 mM CaCl₂ y 20 mM EGTA) y se precipitaron durante 30 min en hielo con 20% de ácido tricloroacético y 0.08% de desoxicolato de sodio en tubos siliconizados. El pellet se lavó con acetona y se dejó secar.

El material purificado por TAP se sometió a cromatografía líquida acoplada a espectrometría de masas (LC-MS/MS) en un equipo EASY-nLC

MATERIALES Y MÉTODOS

(Proxeon) acoplado a una trampa de iones amaZon speed ETD (Bruker) en el servicio de proteómica del IPBLN-CSIC. Las proteínas se identificaron utilizando los softwares ProteinScape (Bruker) y MASCOT (Matrix Ciencia). Se consideraron solo aquellas proteínas con una puntuación mayor de 60 y con al menos dos péptidos detectados.

RESULTADOS

RESULTADOS

1. La proteína de *Trypanosoma brucei* RBP33 es específica de tripanosomátidos

La proteína de unión a RNA RBP33 de *T. brucei* (Tb927.8.990) fue descrita previamente en nuestro laboratorio (Fernández-Moya et al., 2014); posee un motivo de reconocimiento de RNA (RRM), una señal de localización nuclear, los dos últimos tercios de la proteína están intrínsecamente desordenados y no muestra similitud aparente con ninguna proteína conocida fuera de la familia Trypanosomatidae (Fernández-Moya et al., 2014).

Para comprobar si RBP33 era una proteína exclusiva de tripanosomátidos, se realizó una búsqueda exhaustiva de posibles homólogos, comparando los dos últimos tercios de su secuencia proteica en las bases de datos de proteínas tanto generales como específicas de organismo, utilizando diversas estrategias: BLAST, BLAST recíprocos, PSI-BLAST y motivos ocultos de Markov (búsqueda HMMER). A excepción de los homólogos de RBP33 identificados en los genomas de tripanosomátidos, no se pudieron encontrar otras secuencias que mostraran una similitud significativa con esta proteína.

También se analizó la similitud a nivel de estructura secundaria. Sin embargo, a excepción del motivo RRM, la proteína RBP33 está intrínsecamente desordenada y no fue posible detectar elementos estructurales adicionales (Figura 11A). No obstante, contiene una pequeña región, de aproximadamente 25 aminoácidos, que está muy conservada entre las secuencias de tripanosomátidos (recuadro verde en la Figura 11A). Este tipo de motivos cortos “ordenados” embebidos dentro de grandes regiones “desordenadas” es algo muy común en las proteínas intrínsecamente desordenadas, y se denominan SLiMs (*Short Linear Motif*, motivos lineales cortos). Se cree que los SLiMs median interacciones proteína-proteína y se estima que el genoma humano contiene al menos

RESULTADOS

100.000 SLiMs (Bugge et al., 2020). De hecho, existe una base de datos dedicada a dichos motivos (<http://elm.eu.org/index.html>), aunque no se encontraron SLiMs similares a los de RBP33.

Dado que los dos últimos tercios de RBP33 son muy divergentes entre los tripanosomátidos (con la excepción del SLiM mencionado previamente), se utilizó la secuencia correspondiente a los últimos dos tercios de doce proteínas ortólogas diferentes (y sinténicas) de parásitos monoxenos y heteroxenos, para buscar en bases de datos generales de proteínas. También se investigó en bases de datos individuales de procariotas, arqueas, protistas y metazoos incluyendo *Homo sapiens*, *Saccharomyces cerevisiae*, *Dictyostelium discoideum*, *Plasmodium falciparum*, *Arabidopsis thaliana*, *Escherichia coli*, *Sulfolobus solfataricus*, *Thermococcus kodakarensis*, *Methanosarcina acetivorans*, *Methanococcus maripaludis*, *Halobacterium salinarum*, *Haloferax volcanii*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Thiogestelium lacteum*, *Entamoeba histolytica*, *Acanthamoeba castellanii*, *Cryptomonas paramecium*, *Giardia lamblia*, *Emiliana huxleyi*, *Sphaeroforma artica*, *Chroomonas mesostigmatica*, *Thecamonas trahens*, *Capsaspora owczarzaki*, *Fonticula alba* y *Euglena gracilis*. Aún así, no fue posible identificar ningún homólogo de RBP33 o incluso una proteína con similitud significativa.

Utilizando el SLiM como consulta, se identificó un ortólogo sinténico de RBP33 en *Bodo saltans*, una especie de vida libre que pertenece al orden Eubodonida, que están estrechamente relacionados con los tripanosomátidos parásitos (Figura 11B). Sin embargo, no se encontraron homólogos en otras especies de Kinetoplastea de los órdenes Prokinetoplastida, Parabodonida o Neobodonida. Todas estas observaciones proporcionan evidencias suficientes para considerar a RBP33 como una proteína específica de tripanosomátidos, sin homólogos en otros eucariotas. De hecho, ni siquiera está presente en otras especies de kinetoplástidos.

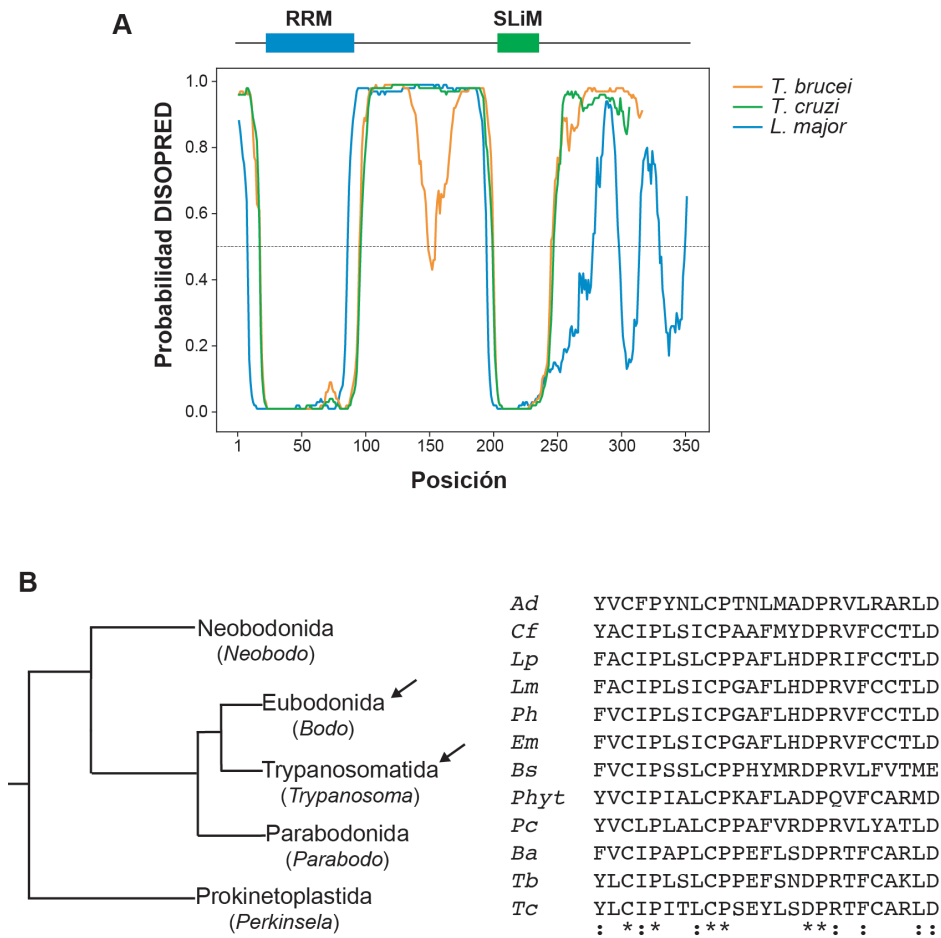


Figura 11. Ortólogos sinténicos de RBP33. (A) Las secuencias de aminoácidos correspondientes a *Trypanosoma brucei*, *Trypanosoma cruzi* y *Leishmania major* se analizaron utilizando DISOPRED3. El motivo RRM en la región N-terminal se representa como un recuadro azul. La línea discontinua horizontal indica el umbral de probabilidad (P=0.5) por encima del cual un aminoácido determinado se considera parte de una región desordenada. Los dos últimos tercios de la proteína RBP33 están, en su mayoría, intrínsecamente desordenados y no se identificaron elementos estructurales adicionales, a excepción de un motivo lineal corto (SLiM, recuadro verde). (B) Árbol filogenético de kinetoplastidos (a la izquierda). Las flechas señalan a los dos únicos órdenes donde se encontraron homólogos de RBP33 (Eubonida y Trypanosomatida). Basado en (Kostygov et al., 2021). Se muestran las secuencias de aminoácidos correspondientes al SLiM (a la derecha) para *Angomonas deanei* (*Ad*), *Crithidia fasciculata* (*Cf*), *Leptomonas pyrrochocorys* (*Lp*), *Leishmania major* (*Lm*), *Porcisia hertigi* (*Ph*), *Endotrypanum monterogeii* (*Em*), *Bodo saltans* (*Bs*), *Phytomonas* sp. (*Phyt*), *Paratrypanosoma confusum* (*Pc*), *Bleptomonas ayalai* (*Ba*), *Trypanosoma brucei* (*Tb*) y *Trypanosoma cruzi* (*Tc*). Los asteriscos (*) y los dos puntos (:) indican residuos de aminoácidos idénticos y sustituciones conservadas en todas las secuencias, respectivamente.

RESULTADOS

2. El silenciamiento de la expresión de la proteína RBP33 conduce a un marcado incremento en los niveles de transcritos antisentido.

RBP33 se une preferentemente a transcritos que provienen de regiones de cambios de hebra convergentes (*convergent strand switch regions*, cSSRs) y otros loci genómicos que normalmente se encuentran silenciados en *T. brucei* (Fernández-Moya et al., 2014). Para comprender mejor la función de RBP33, se analizó el transcriptoma de manera global tanto en la forma sanguínea del parásito como en la procíclica, mediante la técnica de RNA-seq. En todo momento se compararon las células control (S16 para sanguíneas y 449 para procíclicas) con células donde se silenció la expresión de la proteína RBP33 por interferencia de RNA (RNAi) utilizando el plásmido pGR70 (Fernández-Moya et al., 2014).

Las lecturas obtenidas se alinearon con el genoma de referencia Lister 427 (versión 2018) y la asignación a genes se hizo por separado para las lecturas sentido y antisentido. Los coeficientes de correlación de Pearson entre las distintas réplicas fueron $> 0,980$, lo que indica una buena reproducibilidad entre repeticiones (Figura 12).

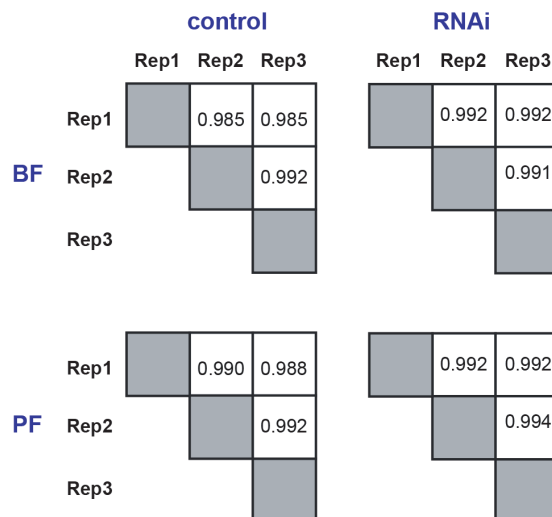


Figura 12. Coeficientes de correlación de Pearson entre réplicas de RNA-seq. BF, formas sanguíneas; PF, formas procíclicas.

Como se muestra en el gráfico de dispersión, se detectó un gran aumento en la abundancia de transcritos antisentido en ambas formas de desarrollo tras el silenciamiento de la expresión de la proteína RBP33 (Figura 13A). Tal y como se describe en la introducción, este fenómeno también sucede en tripanosomas que carecen de H3v y base J (Schulz et al., 2016), en un mutante *knockout* para la H4v (Kim, 2021) y en una línea celular triple *knockout* Δ H3v Δ H4v Δ J (Kim, 2021). Por ello, se analizaron las muestras de RNA-seq correspondiente al transcriptoma del mutante triple KO Δ H3v Δ H4v Δ J con objeto de comparar el grado de acumulación de los transcritos antisentido con respecto a nuestros datos.

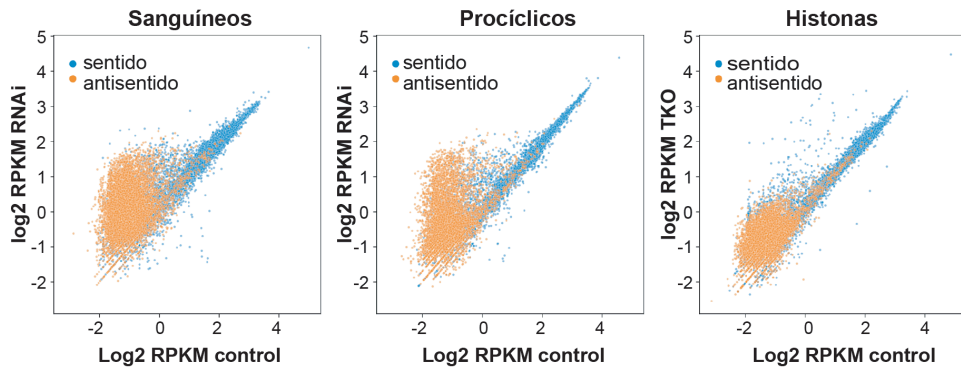
Como se puede observar en la Figura 13A, el aumento en la abundancia de transcritos antisentido fue más pronunciado en la línea celular RBP33 RNAi que en el mutante triple KO. De hecho, las lecturas antisentido fueron entre 7 a 8 veces más altas que en el control en el caso de RBP33, mientras que en la línea celular triple KO se observó un aumento de solo 1,8 veces (Figura 13B y (Kim, 2021)). Además, tras realizar un análisis de componentes principales de las muestras de RBP33 y Δ H3v Δ H4v Δ J, se detectó que la distancia entre las réplicas RNAi RBP33 a sus correspondientes controles fue mayor que la observada en las muestras triple KO (Figura 13C), lo que sugiere un mayor cambio en la expresión génica tras el silenciamiento de RBP33.

A continuación, se realizó un análisis de expresión diferencial para investigar qué genes estaban regulados por RBP33 de manera significativa. Para ello, se combinó el número de lecturas sentido y antisentido y se analizaron conjuntamente con edgeR. Se consideraron solo aquellos genes que vieron aumentada o disminuida su expresión al menos dos veces ($\log_2 |fold-change| \geq 1,0$) y cuya tasa de descubrimiento falso fuera menor del 1% (*false discovery rate*, FDR < 0,01). Observamos que el 40% (en sanguíneos) o el 20% (en procíclicos) de los 17.243 genes anotados en el genoma de *Lister 427 2018* se expresaron diferencialmente en la línea RNAi RBP33. Entre estos, la gran mayoría (> 94%) vieron aumentada su

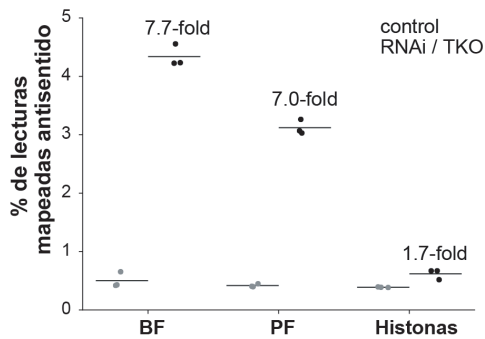
RESULTADOS

expresión principalmente de forma antisentido (Figura 13D). Por otro lado, en los tripanosomas $\Delta H3v\Delta H4v\Delta J$ solo el 6% de los genes se expresaron diferencialmente, el 96% de los cuales se sobreexpresaron de manera similar en sentido y antisentido (Figura 13D).

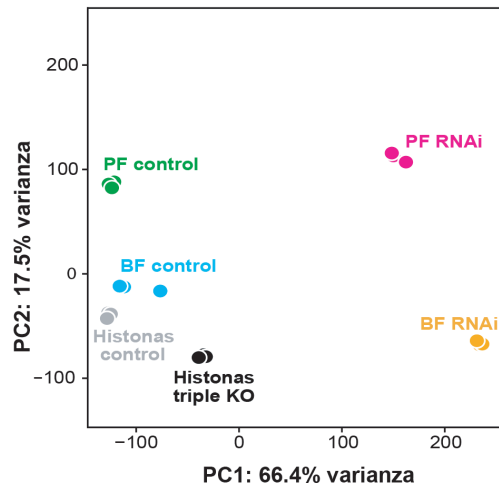
A



B



C



D

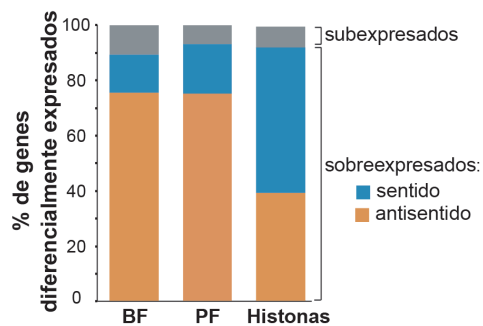


Figura 13. Efecto del silenciamiento de la expresión de RBP33 mediante RNAi en el transcriptoma de *T. brucei*. (A) Gráficos de dispersión que comparan los valores de expresión en el control frente a las células donde la expresión de RBP33 se silenció o en células de tipo silvestre frente al triple KO $\Delta H3v\Delta H4v\Delta J$ (TKO). Los recuentos de lectura promedio entre réplicas se normalizaron a lecturas por kilobase de transcrito por millón de lecturas mapeadas (RPKM) y se transformaron a \log_2 . Cada punto representa un transcrito. (B) Cuantificación de lecturas antisentido correspondientes a las muestras de RBP33 e histonas. El porcentaje de lecturas antisentido mapeadas obtenidas tras la asignación de lecturas a genes, relativas al total de lecturas mapeadas, se representa como la media (línea) de tres réplicas biológicas de RNA-seq (puntos). BF, formas sanguíneas; PF, formas procíclicas. (C) Análisis de componentes principales (PCA) de las muestras de RNA-seq de RBP33 (este estudio) y $\Delta H3v\Delta H4v\Delta J$ triple KO (Kim, 2021). Cada punto representa una réplica. PC, componente principal. (D) Porcentaje de genes diferencialmente expresados de manera sentido o antisentido en las muestras de RBP33 e histonas.

3. Los transcritos regulados por RBP33 derivan principalmente de regiones genómicas silenciadas y de la transcripción antisentido de regiones centrales

En *T. brucei* los cromosomas se dividen en regiones centrales o 'core' que se transcriben y en regiones subteloméricas que se asume que no se expresan (Müller et al., 2018). Dentro de los genes diferencialmente expresados, al silenciar la expresión de RBP33 tanto en tripanosomas sanguíneos como procíclicos, existe una gran abundancia de transcritos que no se detectan en condiciones fisiológicas, tales como regiones de cambio de cadena (SSRs) o subtelómeros, y también de lecturas antisentido derivadas de genes de mantenimiento celular (*housekeeping*). Para cuantificar estas observaciones de manera más precisa, los genes de *T. brucei* fueron clasificados en las siguientes categorías según su localización en el genoma: genes 'core' que codifican proteínas de mantenimiento celular (*housekeeping*) dentro de las regiones centrales; genes 'subteloméricos' no expresados y ubicados en regiones subteloméricas; 'tRNAs', genes tRNA dentro de las regiones centrales; 'cSSR' y 'dSSR', sitios de cambio de cadena convergentes y divergentes dentro de regiones centrales; 'primeros', los primeros tres genes dentro de una PTU (flanqueando las dSSRs); 'últimos', últimos tres genes dentro de una PTU (flanqueando las cSSRs); y genes 'en inverso' localizados en orientación inversa dentro de una PTU

RESULTADOS

(Figura 14A). Pudimos observar que la mayoría de los RNAs regulados al alza por RBP33 tanto en sanguíneos como en procíclicos, derivaban de subtelómeros (sentido) o regiones centrales (transcripción antisentido), como se muestra en la Figura 14B.

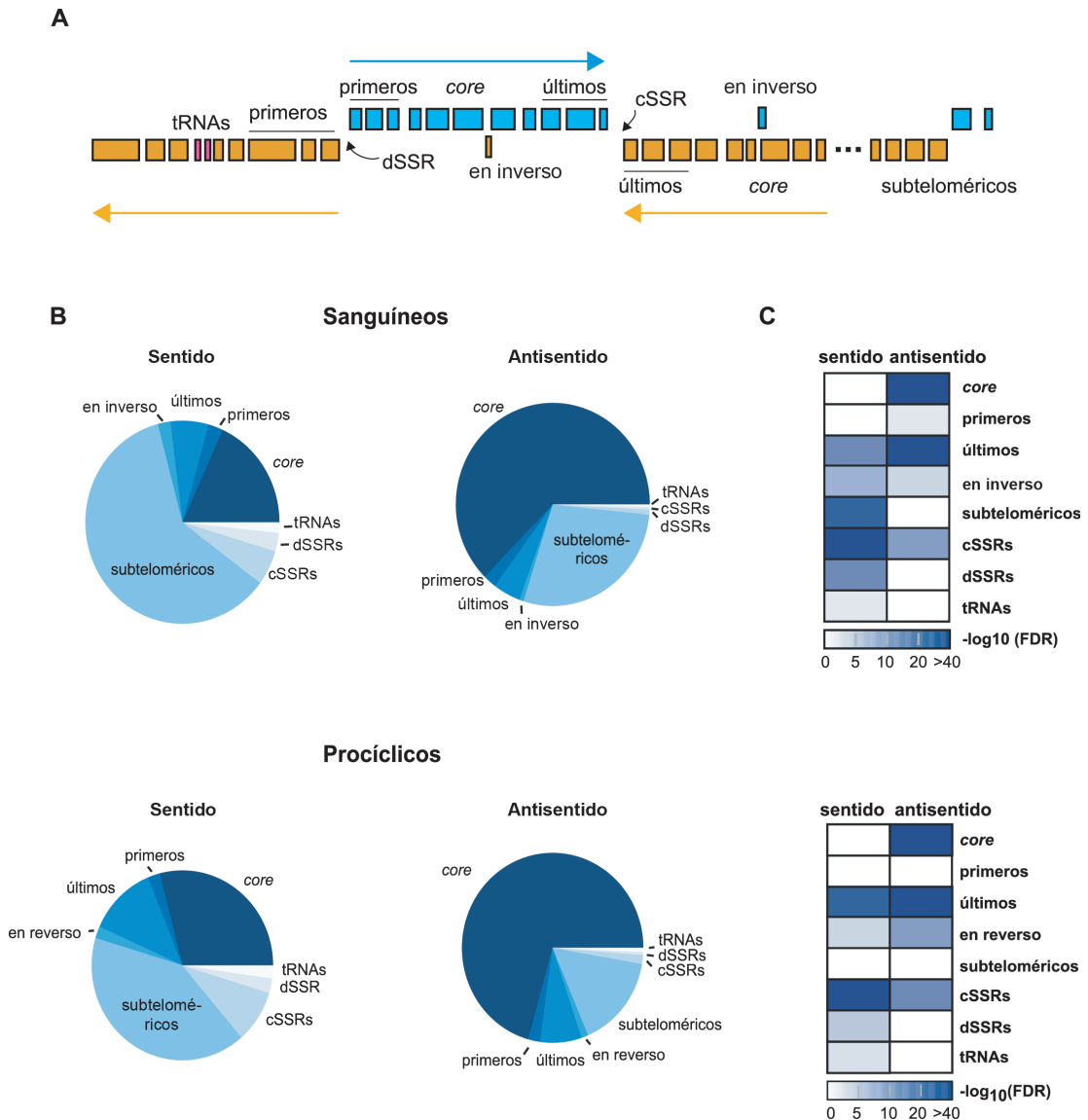


Figura 14. Clasificación de los genes regulados por RBP33 según su localización genómica. (A) Representación esquemática de un cromosoma de *T. brucei* donde se indican las diferentes localizaciones utilizadas en este estudio. (B) Proporciones relativas de genes sobreexpresados dentro de cada categoría de manera sentido (izquierda) y

antisentido (derecha) en tripanosomas sanguíneos (arriba) y procíclicos (abajo). (C) Se analizó el enriquecimiento de genes sobreexpresados en cada categoría mediante la prueba exacta de Fisher en tripanosomas sanguíneos (arriba) o procíclicos (abajo), en ausencia de RBP33.

A continuación, se determinó el enriquecimiento estadístico de los transcritos regulados por RBP33 dentro de cada una de las categorías mediante una prueba exacta de Fisher. Se observó un enriquecimiento significativo ($FDR < 0,05$) de los transcritos sentido derivados de subtelómeros, cSSR, dSSR, 'últimos', tRNAs y 'en inverso' y de transcritos antisentido originados en regiones 'core', 'últimos', 'primeros', 'en inverso' y 'cSSR' (Figura 14C para sanguíneos (arriba) y procíclicos (abajo)). Lo más llamativo fue que 58 (en sanguíneos) o 56 (en procíclicos) de las 59 cSSRs anotadas se sobreexpresaron de manera significativa en ausencia de RBP33.

La mayoría de los subtelómeros parece que se encuentran transcripcionalmente silenciados en condiciones fisiológicas. Sin embargo, nuestros datos indicaron que los genes subteloméricos se expresaban en ausencia de RBP33, especialmente en las formas sanguíneas. Para visualizar el alcance de este fenómeno, se calculó la amplitud de cobertura de cada subtelómero como el porcentaje de bases que se secuenciaron al menos una vez en al menos dos réplicas de RNA-seq (profundidad = 1X). La Figura 15 muestra los valores de amplitud para los 33 subtelómeros correspondientes a las muestras de RBP33 e histonas. Se analizaron diferentes profundidades de secuenciación y se observó en todos los casos un aumento pronunciado en la amplitud de la cobertura tras el silenciamiento de la expresión de RBP33 en sanguíneos: así solo el 6% de las bases subteloméricas se secuenciaron al menos una vez (profundidad = 1X) en las células control, mientras que este valor aumentó al 43% en ausencia de RBP33. Los valores de amplitud también aumentaron en las muestras de procíclicos e histonas, aunque de una manera menos pronunciada (profundidad = 1X; 4.5 veces y 2.5 veces, respectivamente).

RESULTADOS

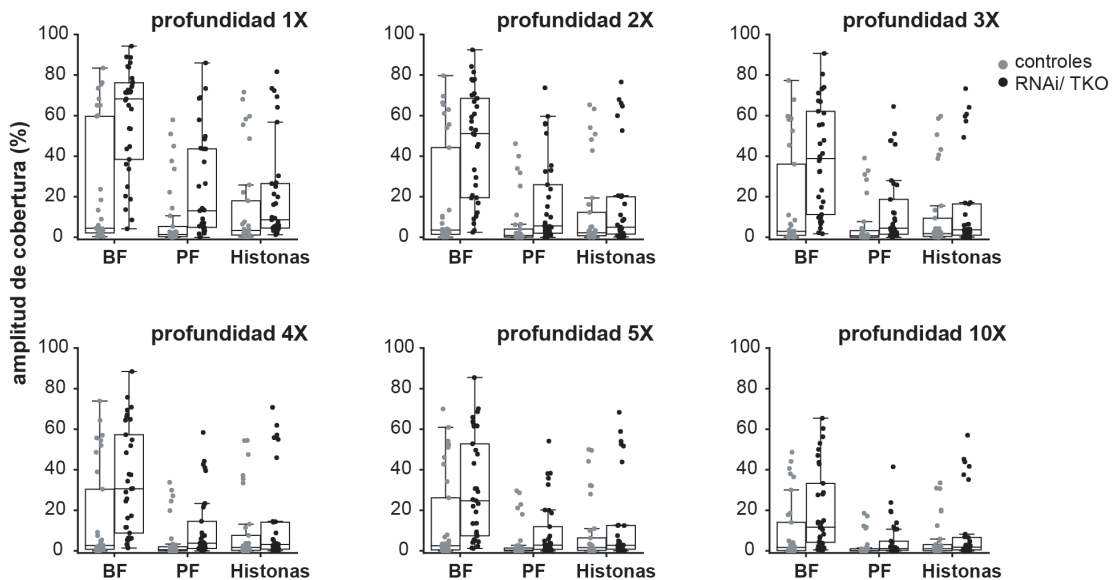
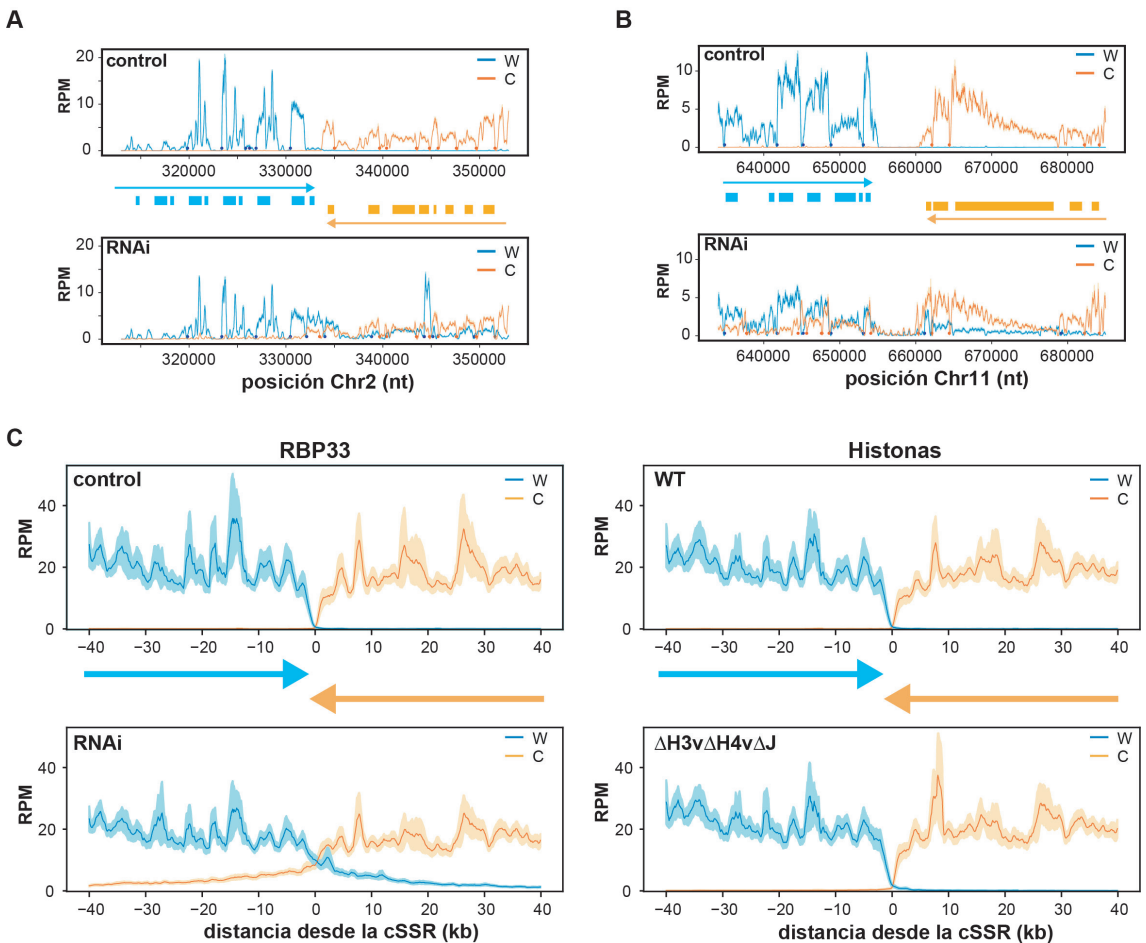


Figura 15. Amplitud de cobertura en regiones subteloméricas en las muestras de RBP33 (BF, tripanosomas sanguíneos; PF, procíclicos) y del triple KO H3vΔH4vΔJ (Histonas). Los diagramas de caja representan el porcentaje de bases en los subtelómeros secuenciadas a distintas profundidades. Los rectángulos indican el rango intercuartílico (*interquartile range*, IQR); bigotes, ± 1.5 IQR; líneas horizontales dentro de los rectángulos, medianas; puntos, subtelómeros individuales.

En vista del notable aumento observado en los niveles de transcritos procedentes de cSSRs, nos propusimos estudiar con más detalle la distribución de las lecturas mapeadas alrededor de las cSSRs. Se utilizaron ventanas deslizantes para analizar las regiones genómicas correspondientes a zonas convergentes (considerándolas como el centro) y 15-20 kb de extensión a ambos lados de las mismas. Las Figuras 16A-B muestran los perfiles de cobertura correspondientes a dos cSSRs representativas (se muestran ejemplos adicionales en la Figura suplementaria S1A-F en Anexo I). Los valores de lecturas por millón de lecturas mapeadas (*reads per million mapped reads*, RPM) se generaron por separado para cada hebra. En las muestras control apenas se detectaron transcritos provenientes de zonas convergentes ni de las regiones adyacentes. Sin embargo, la ausencia de RBP33 produjo un claro

aumento de transcritos antisentido, representados como picos naranjas en la PTU aguas arriba (izquierda) o como picos azules en la PTU aguas abajo (derecha).

Este fenómeno se visualizó mejor al trazar el promedio de RPM en 30 cSSRs diferentes flanqueadas por PTUs de al menos 50 kb de largo (Figura 16C izquierda). Se puede observar que la aparición de estos transcritos antisentido se extiende hasta 40 kb desde la cSSR. Este fenotipo también se observó en el triple KO $\Delta H3v\Delta H4v\Delta J$ (Kim, 2021; Schulz et al., 2016), aunque el efecto fue mucho menos pronunciado (Figura 16C derecha).



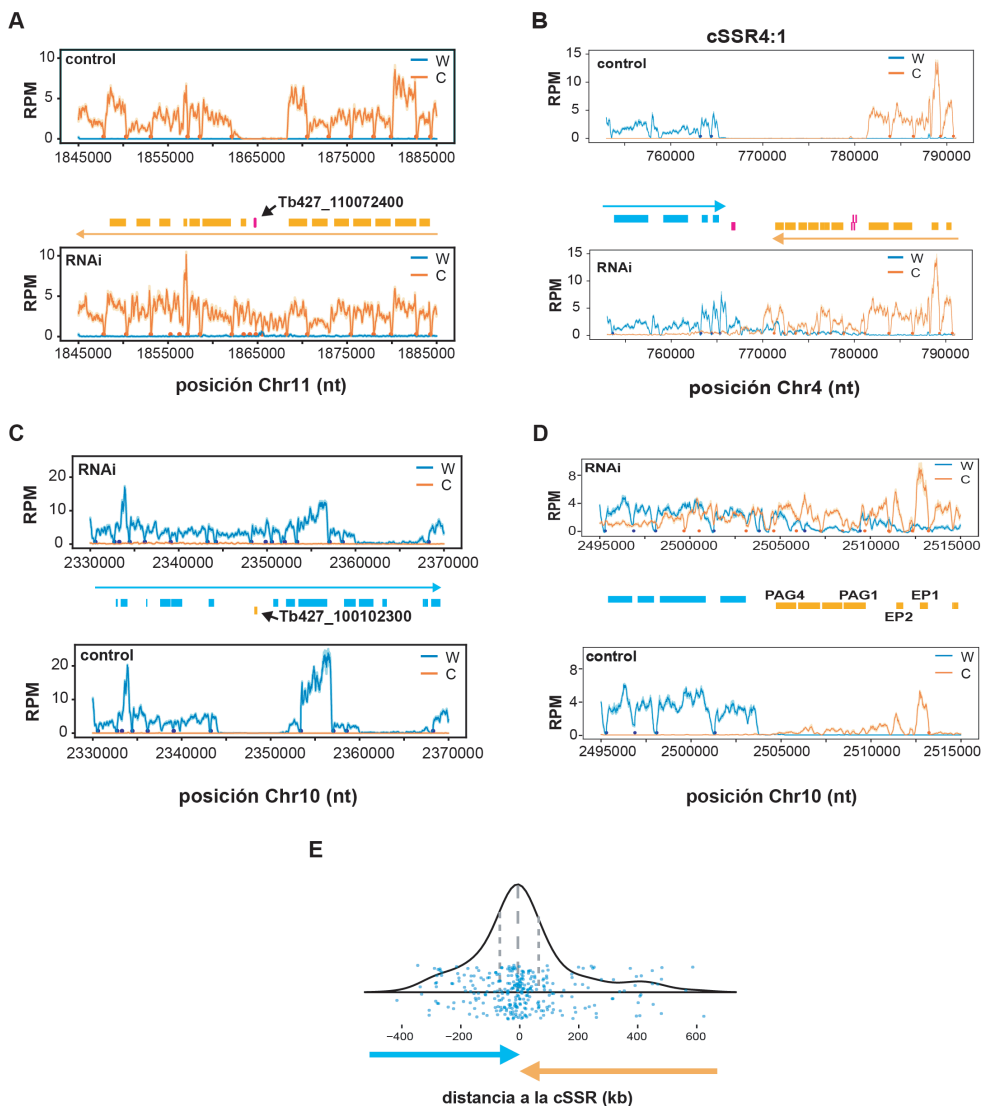
RESULTADOS

Figura 16. Gráficos de cobertura de regiones genómicas representativas (I). Los recuentos de lecturas se obtuvieron por separado para la hebras de Watson (W, líneas azules) o Crick (C, líneas naranjas) en tripanosomas sanguíneos utilizando ventanas deslizantes (con 'bins' de 100 pb, de 10 pb en 10 pb para todos los paneles excepto en el panel **C**, que se analizó con 'bins' de 1000 pb y de 100 pb en 100 pb y normalizando por el tamaño de la librería (lecturas por millón, RPM). Los valores correspondientes a las muestras control (gráficos superiores panel **A** y **B**) se compararon con los de las muestras en ausencia de RBP33 (RNAi, gráficos inferiores panel **A** y **B**) y se representaron como la media (líneas continuas) \pm SEM (área sombreada) de tres réplicas biológicas de RNA-seq, excepto para el panel **C**, donde la media \pm SEM hace referencia a la cobertura de 30 cSSRs. Los genes se representan como rectángulos (azul, genes que codifican proteínas en la cadena W; naranja, genes que codifican proteínas en la cadena C; rosa, genes que codifican tRNAs) y las flechas de colores indican la dirección de la transcripción en cada unidad policistrónica (PTU). Los círculos azules y naranjas indican los sitios aceptores de splicing asignados a las hebras W y C, respectivamente. (**A** y **B**) Ejemplos de regiones de cambio de cadena convergentes, cSSR2:2 y cSSR11:2, respectivamente. (**C**) Cobertura media en 30 cSSRs en el control frente a células donde se silenció la expresión de RBP33 mediante RNAi (izquierda) o en el control frente al triple KO H3v Δ H4v Δ J (derecha).

Anteriormente se ha mencionado que la abundancia de tRNAs y de transcritos correspondientes a genes que poseen una orientación inversa dentro de una PTU, aumentaba de forma significativa tras el silenciamiento de la expresión de RBP33 por ribointerferencia. La mayoría de los genes de tRNAs están ubicados en las cSSRs, lo que explicaría su sobreexpresión, aunque otros también residen dentro de las PTUs; el fenotipo fue similar en ambos casos (Figura 17A-B y Figura suplementaria S1C en Anexo I). En cuanto a los genes en orientación inversa, observamos un fenómeno semejante (Figura 17C y Figura suplementarias S1D-E en Anexo I), que también fue detectado en los límites entre unidades transcripcionales de RNA polimerasa I y RNA polimerasa II (locus *EP1*, Figura 17D).

Solo una pequeña fracción del conjunto de transcritos diferencialmente regulados (6% en sanguíneos) vio disminuida su abundancia tras el silenciamiento de RBP33. Casi todos estos RNAs (98%) estaban regulados a la baja de manera sentido. Esto podría ser el resultado de la interferencia de RNA desencadenada por la formación de

RNA bicatenario debido al aumento de los niveles de transcritos antisentido. Si ese fuera el caso, cabría esperar que los genes regulados a la baja se ubicaran preferentemente cerca de las cSSRs, que son las regiones donde se generan principalmente los transcritos antisentido. Para comprobar esta hipótesis, se analizó la localización genómica de los genes subexpresados con respecto a la cSSR más cercana, y se observó que la mayoría se encontraban próximos a estas regiones (Figura 17E). Cabe mencionar que al menos 141 de los genes regulados a la baja en sanguíneos y 11 en formas procíclicas son esenciales para el crecimiento del tripanosoma de acuerdo con cribados RNAi de todo el genoma (Alford et al., 2011).



RESULTADOS

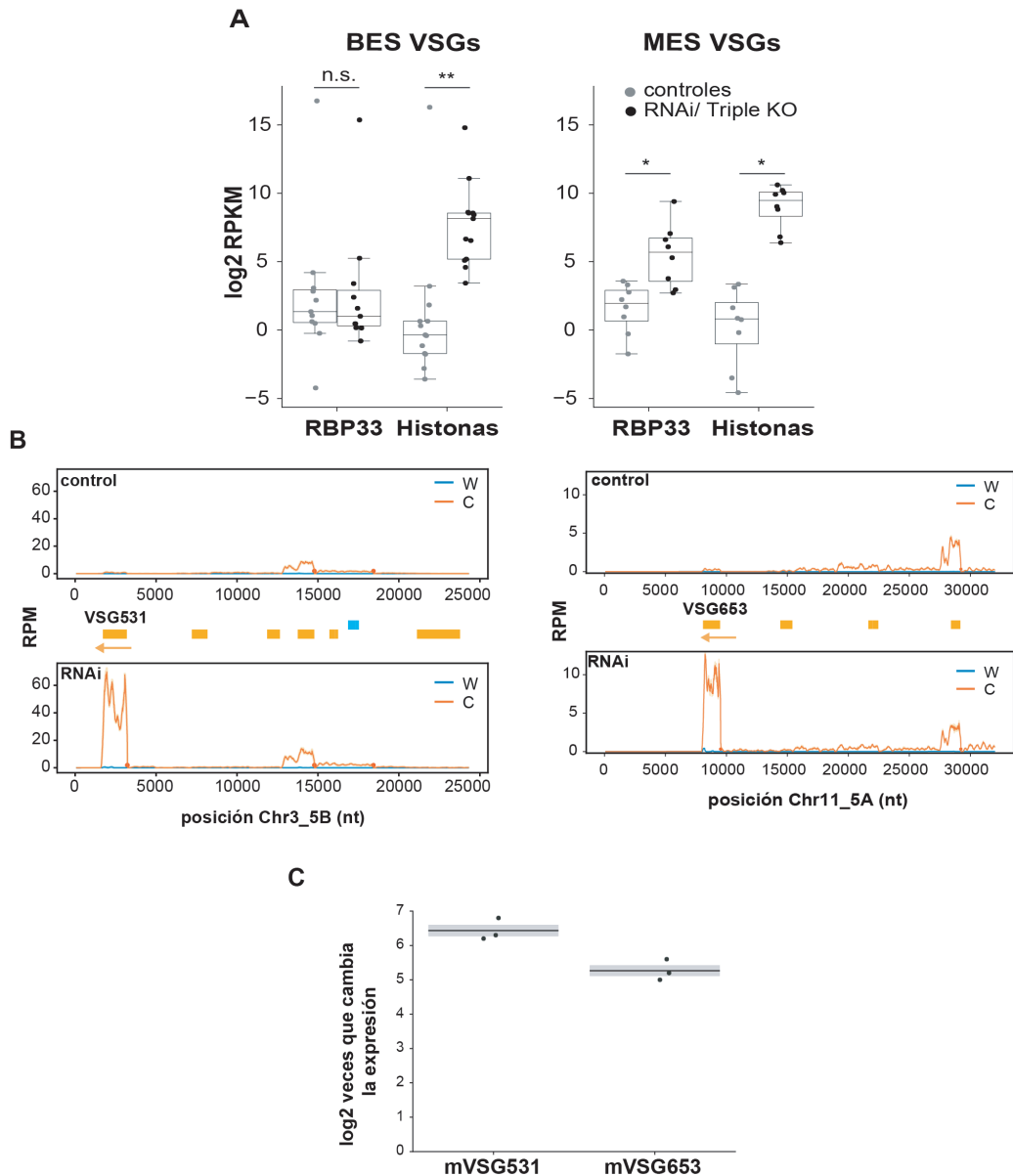
Figura 17. Gráficos de cobertura de regiones genómicas representativas (II). (A) Gen de tRNA Tb427_110072400 (Tb927.11.tRNA_Gly_1). (B) cSSR4:1 incluidos los genes agrupados de tRNAs (recuadros rosas) Tb427_040034500, Tb427_040034600 y Tb427_040034700 (en la cSSR) y Tb427_040035400, Tb427_040035500, Tb427_040035600 y Tb427_040035700 (dentro de la PTU en la hebra de Crick); (C) Ejemplo de un gen que codifica una proteína que se encuentra en orientación inversa dentro de una PTU, Tb427_100102300 (Tb927.10.9530). Se pueden encontrar más ejemplos en Figura suplementaria S1. (D) Locus del gen *EP1*. (E) Gráfico de violín partido que muestra la distribución de genes regulados negativamente por RBP33 dentro de PTUs convergentes en las formas sanguíneas. La distancia fue medida desde el comienzo de cada gen hasta la mitad de la cSSR correspondiente. Cada punto representa un gen. La línea discontinua en la mitad dentro del gráfico de violín representa la mediana y las líneas discontinuas externas indican el rango intercuartílico (*interquartile range*, IQR).

4. Los transcritos que codifican las VSGs metacíclicas se sobreexpresan en ausencia de RBP33

En las regiones teloméricas de *T. brucei* la transcripción se encuentra fuertemente reprimida (Cestari & Stuart, 2018; Müller et al., 2018). Estos loci suelen albergar genes que codifican las VSGs, los cuales se organizan en sitios de expresión de la forma sanguínea (BES) o metacíclica (MES) (ver apartado 1.4. de introducción). Tanto H3v como la base J se encuentran enriquecidas en regiones teloméricas en *T. brucei* (Cliffe et al., 2009; Siegel et al., 2010) y, en la línea $\Delta H3v\Delta H4v\Delta J$, las VSGs de BES y MES se sobreexpresan (Kim, 2021; Müller et al., 2018).

Dado que nuestros resultados hasta ahora sugerían que RBP33 parecía estar involucrada en el silenciamiento de zonas que no se expresan en condiciones fisiológicas, decidimos estudiar si el silenciamiento de RBP33 tenía algún efecto en la expresión de las VSGs. Para ello, se compararon los valores normalizados de recuento de lecturas correspondientes a 17 genes BES VSG y 8 MES VSG, entre las muestras control y las muestras RNAi. Como referencia, los datos $\Delta H3v\Delta H4v\Delta J$ se analizaron de la misma manera. Todos los transcritos de MES VSGs vieron aumentada su expresión en tripanosomas sanguíneos en ausencia de RBP33, mientras que los niveles de mRNA que codifican BES VSGs apenas

se vieron afectados (Figura 18A, Figura suplementaria S2A en Anexo I). Para ilustrar este fenómeno, en la Figura 18B y en la Figura suplementaria S2A y S2B en Anexo I se muestran gráficos de cobertura de algunas regiones teloméricas. Por el contrario, el mutante triple KO $\Delta H3v\Delta H4v\Delta J$ sobreexpresa ambos tipos de transcritos VSGs ((Kim, 2021); Figura 18A, Figura suplementaria S2 en Anexo I). Los cambios en la expresión de dos MES VSGs se confirmaron mediante RT-PCR cuantitativa (Figura 18C).



RESULTADOS

Figura 18. Los transcritos que codifican VSGs metacíclicas están sobreexpresados en ausencia de RBP33 en las formas sanguíneas del parásito. (A) Efecto del silenciamiento de RBP33 y del triple KO $\Delta H3v\Delta H4v\Delta J$ en la expresión de genes VSG metacíclicos (MES) y sanguíneos (BES). Los recuentos de lecturas se normalizaron a CPM con edgeR, se convirtieron a RPKM y se transformaron a \log_2 . Se calcularon los valores promedio de RPKM de las réplicas de RNA-seq para 17 transcritos BES VSG y 8 MES VSG, y se representaron en diagramas de caja (las cajas representan el rango intercuartílico (*interquartile range*, IQR); bigotes, ± 1.5 IQR; líneas horizontales dentro de los rectángulos, medianas; puntos, transcritos de VSG individuales). Solo se consideraron los transcritos de VSG que mostraban valores de RPKM ≥ 1 en al menos tres réplicas biológicas. Se usaron pruebas de la U de Mann-Whitney de dos colas para evaluar si había una expresión diferencial significativa entre el RNAi de RBP33 o el triple KO de histonas y sus respectivas muestras control; n.s., no significativo ($p > 0,5$); *, $p < 0,005$; **, $p < 0,0005$ ($n = 11/13$ BES VSG; $n = 8$ MES VSG). (B) Gráficos de cobertura correspondientes a los loci que albergan las VSG metacíclicas VSG531 y VSG653. (C) RT-PCR cuantitativa para confirmar la regulación positiva de dos transcritos MES VSG. Los cambios en la expresión se representan como la media (líneas horizontales) \pm SEM (área sombreada) de tres inducciones independientes de RNAi en tripanosomas sanguíneos.

5. Los transcritos regulados por RBP33 están correctamente procesados

Para evaluar si los transcritos regulados por RBP33 estaban correctamente procesados, se realizó una búsqueda global de sitios aceptores de *splicing* (SAS) en nuestros datos de RNA-seq. Se identificaron 8553 SASs en el control y 10426 SAS en la línea RNAi RBP33. En ambas condiciones se encontró un dinucleótido canónico aceptor AG en el 98% de los SASs, lo que está de acuerdo con otros estudios (Kolev et al., 2010; Rastrojo et al., 2013). Los SASs mapeados se indican como círculos azules (hebra W) o naranjas (hebra C) en la Figura 16, en la Figura 17 y en la Figura suplementaria S1 en Anexo I y dichos SASs coinciden con picos en los gráficos de cobertura. Se detectaron 42 SASs en el control y 707 en ausencia de RBP33, para los cuales el gen aguas abajo más cercano estaba ubicado en la cadena opuesta, lo que es indicativo de posibles transcritos antisentido. No pudimos abordar si estos transcritos también estaban poliadenilados debido al método de preparación de la librería. Sin embargo, dado que las muestras de RNA se procesaron mediante cromatografía con bolitas magnéticas con oligo(dT) antes de la construcción de la librería, es razonable suponer que los transcritos regulados por RBP33 contienen colas de poli(A). De hecho, en nuestro

laboratorio ya demostramos mediante ensayos de digestión con RNasa H que los transcritos que derivan de cSSRs, en ausencia de RBP33, están poliadenilados (Fernández-Moya et al., 2014). Estos resultados sugieren que al menos algunos de los transcritos que surgen tras el silenciamiento de RBP33 se procesan correctamente y, por tanto, son potencialmente traducibles.

6. Expresión dependiente de RBP33 de un gen reportero insertado en orientación antisentido dentro de una región genómica silenciada

El silenciamiento de RBP33 da como resultado la aparición de nuevos transcritos que parecen estar correctamente procesados y que podrían traducirse. Esta cuestión fue abordada directamente mediante la inserción de un gen que codifica la proteína fluorescente amarilla mejorada (eYFP), en una orientación antisentido en dos loci cSSR diferentes en tripanosomas sanguíneos (flechas verdes en la Figura 19A). Para ello, el gen de *eYFP* presente en p2675 (Kelly et al., 2007) se amplificó por PCR y se clonó en orientación inversa en el plásmido pGR12 (Rico-Jiménez et al., 2021) para generar pGR417, donde el ORF de *eYFP* está flanqueado por la 5'-UTR de *EP1* y la 3'-UTR del gen que codifica la actina. La secuencia del espaciador de rDNA en pGR417 se reemplazó por fragmentos de DNA correspondientes a los genes Tb427_110143800 o Tb427_100047800 para generar pGR431 y pGR433, respectivamente. Los plásmidos se linealizaron con EcoRI (pGR431) o MfeI (pGR433) y se transfetaron en células control S16. Las líneas celulares estables así obtenidas fueron seleccionadas en presencia de 5 µg/ml de blasticidina, y se comprobó la correcta integración de los plásmidos mediante PCR de DNA genómico. El proceso de generación de las construcciones, así como la comprobación de la integración en el genoma, se representa en la Figura 19B. Las líneas celulares que tenían insertada la eYFP, se transfetaron adicionalmente con pGR70 (Fernández-Moya et al., 2014) para el silenciamiento por ribointerferencia de la proteína RBP33.

RESULTADOS

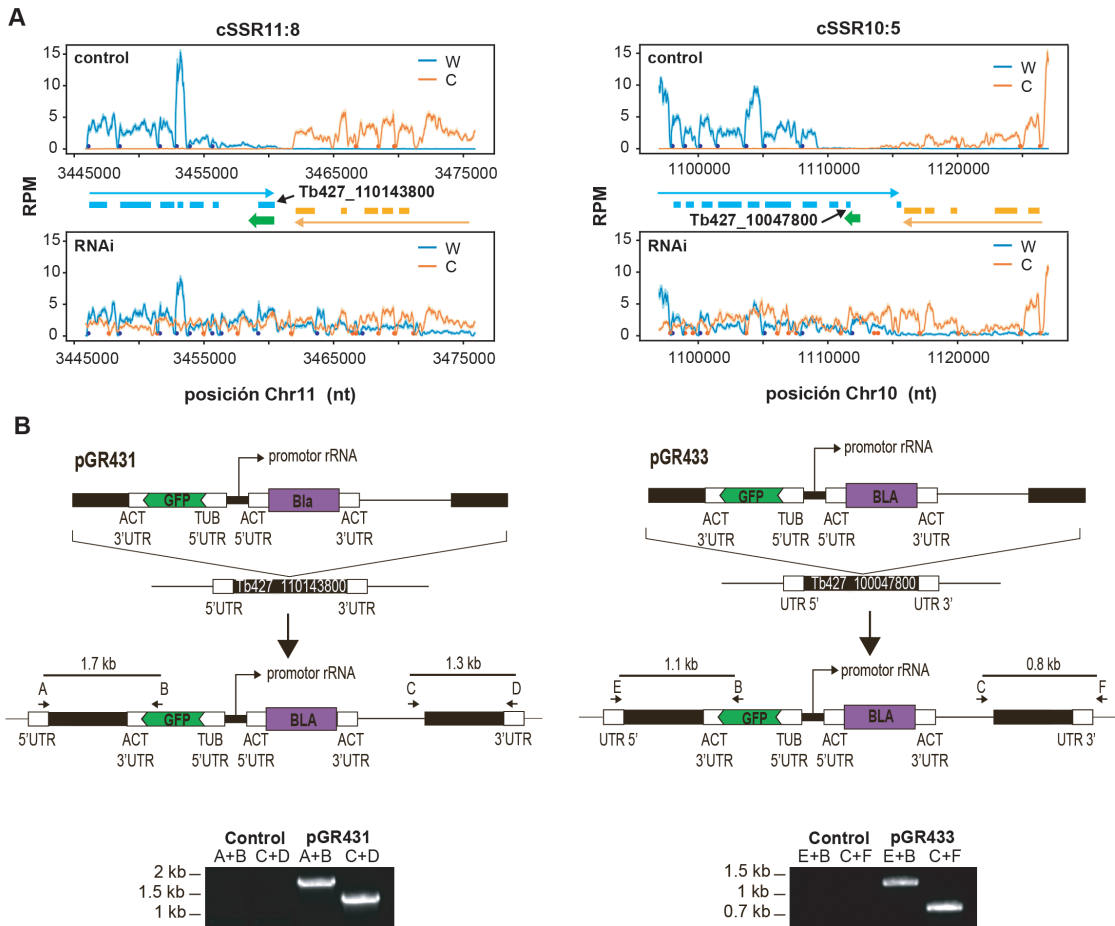


Figura 19. Generación de líneas celulares portadoras del gen *eYFP* insertado en orientación inversa en dos loci cSSR distintos. (A) Gráficos de cobertura correspondientes a las regiones cSSR11:8 (izquierda) y cSSR10:5 (derecha) donde se insertó el gen *eYFP*. Los genes utilizados como diana para la inserción se indican con flechas negras. (B) Representación esquemática de las construcciones de expresión pGR431 (dirigida al gen Tb427_110143800) y pGR433 (dirigida al gen Tb427_100047800). Como marcador de selección, se utilizó un gen de resistencia a blasticidina transcrito a partir de un promotor constitutivo en dirección opuesta con respecto al gen *eYFP* (rectángulos morados). Los esquemas correspondientes a los loci genómicos tras la integración de los plásmidos de expresión se representan debajo de la flecha vertical. Se comprobó la correcta integración del plásmido mediante genotipado utilizando una combinación de cebadores específicos del vector (cebador B y cebador C) y de cebadores que hibridan fuera de las regiones de recombinación (cebador A y cebador D para pGR431; cebador E y cebador F para pGR433). Los tamaños del producto de PCR esperados se indican debajo de cada construcción. Como control negativo se utilizó el DNA genómico obtenido de la línea celular parental S16.

A continuación, la expresión de eYFP se midió mediante citometría de flujo. Como era de esperar, se detectó muy poca fluorescencia en los cultivos no inducidos. Sin embargo, la expresión de eYFP aumentó entre 3 y 10 veces tras el silenciamiento de RBP33 por RNAi (Figura 20). Estos resultados sugieren que los transcritos regulados por RBP33 podrían llegar a traducirse a proteínas.

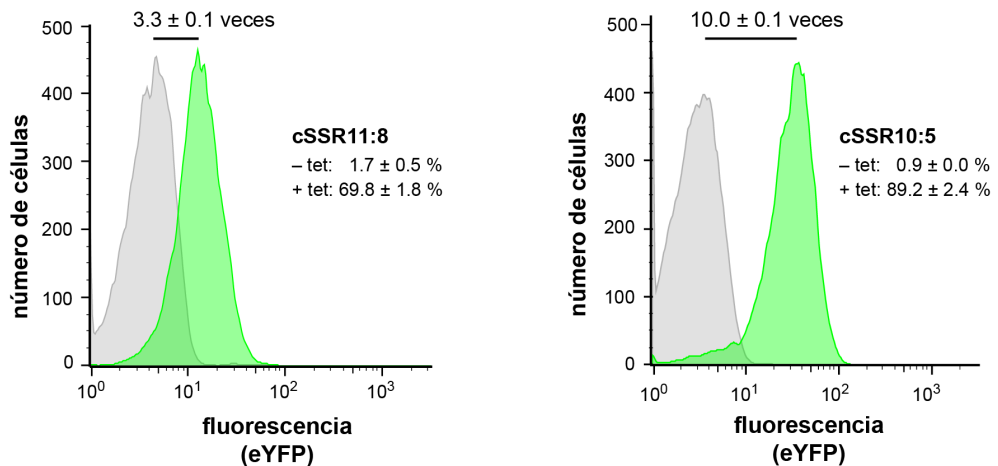


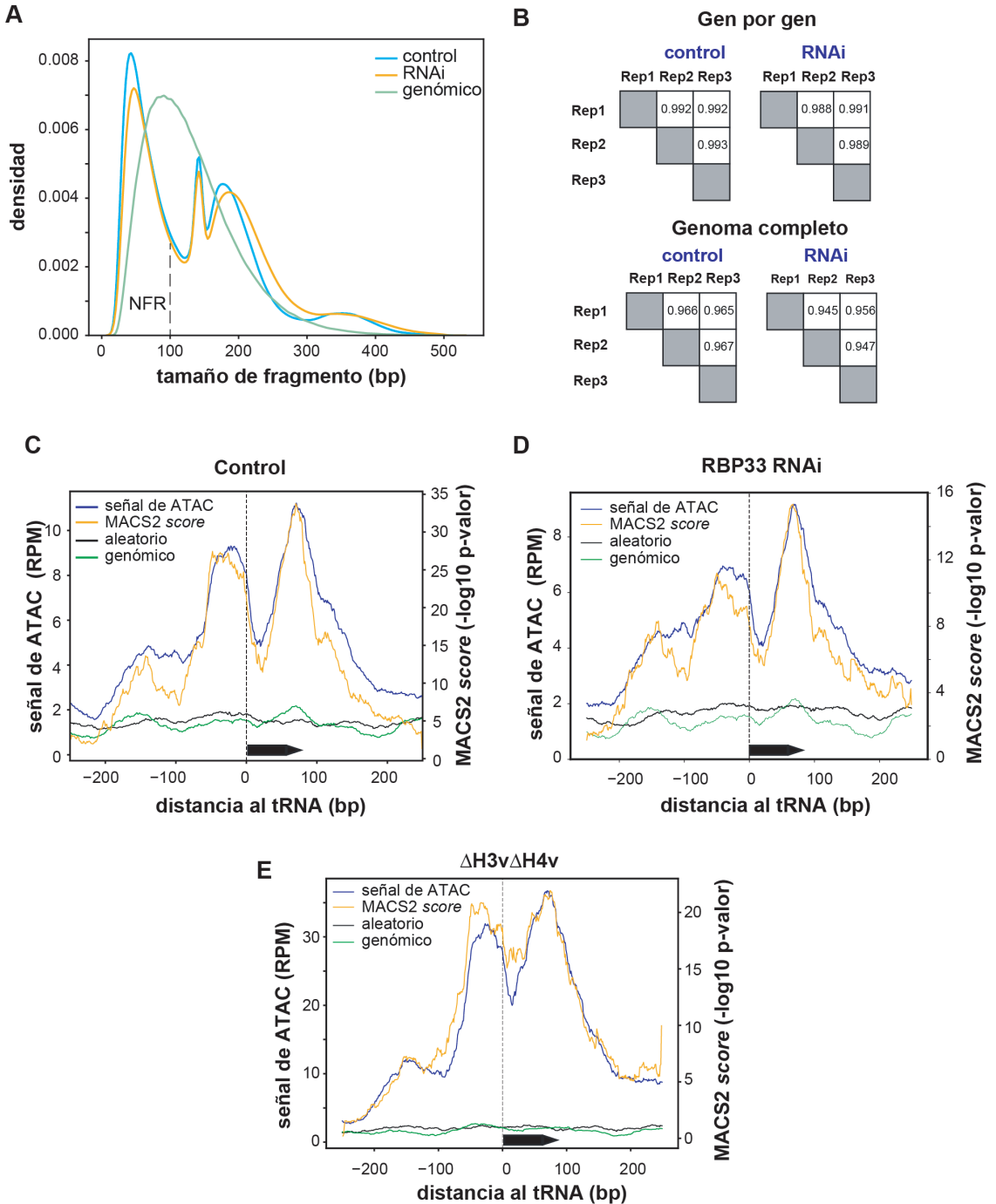
Figura 20. Cuantificación de la expresión eYFP en tripanosomas sanguíneos mediante citometría de flujo. Las líneas celulares con el gen *eYFP* insertado en dos loci cSSR diferentes (cSSR11:8 y cSSR10:5), se transfectaron con el plásmido pGR70 (Fernández-Moya et al., 2014) para el silenciamiento de la expresión de RBP33 por ribointerferencia. Se muestran los histogramas de citometría de flujo correspondientes a tripanosomas no inducidos (curvas grises) e inducidos con tetraciclina (curvas verdes). Se utilizó el software FlowJo v10 para calcular la media de la fluorescencia por célula en cada muestra. Se indican los valores de cambios en la expresión de la fluorescencia (veces) en tripanosomas donde se silenció la expresión de RBP33 mediante RNAi en relación a los controles sin inducir, así como el porcentaje de células eYFP positivas en cada caso. Los valores se representan como la media \pm SEM de tres inducciones de RNAi independientes.

7. Análisis de la accesibilidad de la cromatina en ausencia de RBP33

La regulación positiva de los transcritos que se originan en las cSSRs y MES VSGs podría deberse a un aumento en la accesibilidad de la cromatina causada por el silenciamiento de la expresión de RBP33. De hecho, la ausencia de H3v, H4v y base J, que están enriquecidas en estos loci genómicos, conduce a un fenotipo similar al observado para RBP33 (Kim, 2021). Para comprobar esta hipótesis, se determinó la accesibilidad de la cromatina mediante el ensayo de cromatina accesible por transposasa usando secuenciación masiva (ATAC-seq), en líneas control y RNAi. Se compararon nuestros resultados con los obtenidos en un análisis ATAC-seq previamente publicado, realizado en un doble mutante KO Δ H3v Δ H4v (Müller et al., 2018). Los fragmentos libres de nucleosomas (*nucleosome-free regions*, NFR) fueron seleccionados tras alinear las lecturas de secuenciación con el genoma de referencia (umbral de tamaño = 100 pb; Figura 21A) y se usaron para análisis adicionales. Los coeficientes de correlación de Pearson fueron $> 0,940$, lo que indica una buena reproducibilidad entre réplicas (Figura 21B).

Un control de calidad estándar en los experimentos ATAC-seq incluye el análisis de cobertura alrededor de regiones NFR conocidas, como por ejemplo los sitios de inicio de la transcripción de los genes que codifican proteínas (Yan et al., 2020). Sin embargo, no hay promotores bien definidos asociados a genes individuales en tripanosomátidos, ya que los genes están organizados en PTUs. En lugar de eso, lo que hicimos fue medir la cobertura en los loci de tRNA, que generalmente están libres de nucleosomas en *T. brucei* y en otros eucariotas (Maree et al., 2017; Nagarajavel et al., 2013). Se observó una NFR clara al trazar el promedio de RPM de diez genes de tRNAs aislados (es decir, no agrupados) diferentes (Figura 21C, línea azul). La identificación de picos de cromatina abierta utilizando MACS2 generó prácticamente el mismo resultado (Figura 21C, línea naranja), mientras que ni la muestra control de DNA genómico (línea verde) ni el promedio de diez loci genómicos de la misma longitud seleccionados al azar (línea negra) mostró un aumento notable en la señal

ATAC-seq. Se obtuvieron perfiles de cromatina abierta similares en los loci de tRNA cuando se analizaron las muestras de RBP33 RNAi y Δ H3v Δ H4v (Figuras 21D-E).



RESULTADOS

Figura 21. Análisis de ATAC-seq. (A) Distribución del tamaño de los fragmentos de ATAC-seq. Los tamaños de los fragmentos se calcularon utilizando el módulo 'bamPEFfragmentSize' de deepTools. Los valores promedio de las tres réplicas biológicas se representaron utilizando gráficos de densidad de kernel. Las lecturas correspondientes a las regiones libres de nucleosomas (NFR <100 pb) se seleccionaron para los análisis posteriores. (B) Coeficientes de correlación de Pearson en las muestras control y en las réplicas de RNA-seq en ausencia de RBP33 (RNAi). Los valores se obtuvieron tras analizar la cobertura de lecturas teniendo en cuenta todos los genes anotados en el genoma de *T. brucei* (gen por gen) o en todo el genoma, dividiendo el genoma en 'bins' de 100 pb (genoma completo). (C, D y E) Gráfico de cobertura de base única para la señal de ATAC correspondiente a genes individuales de tRNAs en tripanosomas sanguíneos control (C), en ausencia de RBP33 (D) y en las muestras del doble KO $\Delta H3v\Delta H4v$ (E). Las flechas negras gruesas representan un gen de tRNA. Se obtuvieron recuentos de lecturas para regiones genómicas centradas en genes de tRNA individuales y para las 250 pb extendidas a cada lado, normalizados por tamaño de librería (RPM) y representados como el promedio de valores de RPM (línea azul) o de MACS2 (línea naranja) en diez loci de tRNA diferentes. También se muestran las señales de ATAC correspondientes al promedio de diez regiones elegidas al azar del mismo tamaño (línea negra) y a un control de DNA genómico (línea verde).

También se observaron NFRs en loci de tRNA agrupados (Figura suplementaria S3 en Anexo I). Se observó un patrón de doble pico en la señal de ATAC en la mayoría de los genes de tRNA con un valle inmediatamente aguas abajo del sitio de inicio de la transcripción (Figura 21A). Esto podría interpretarse como una huella (*footprint*) causada por un factor de transcripción RNA polimerasa III unido al elemento box A del promotor del gen tRNA.

A continuación, se compararon las muestras de RBP33 e histonas mediante un análisis de componentes principales. Los ratios de cobertura relativos al DNA genómico (tras la reacción de transposición) se obtuvieron bien para todos los genes anotados o bien para el genoma completo tras dividir el genoma en *bins* de 100 pb no solapantes. En las muestras correspondientes a RBP33, las réplicas de las muestras control y RNAi agruparon juntas, mientras que se observó una marcada separación entre los controles de histonas (de tipo silvestre) y el doble KO $\Delta H3v\Delta H4v$ (Figura 22A-B). Además, cuando analizamos los cambios en la señal de ATAC respecto a los controles, vimos un efecto mucho mayor en la accesibilidad de la cromatina en las muestras de histonas que en las de RBP33 (Figura 22C-D).

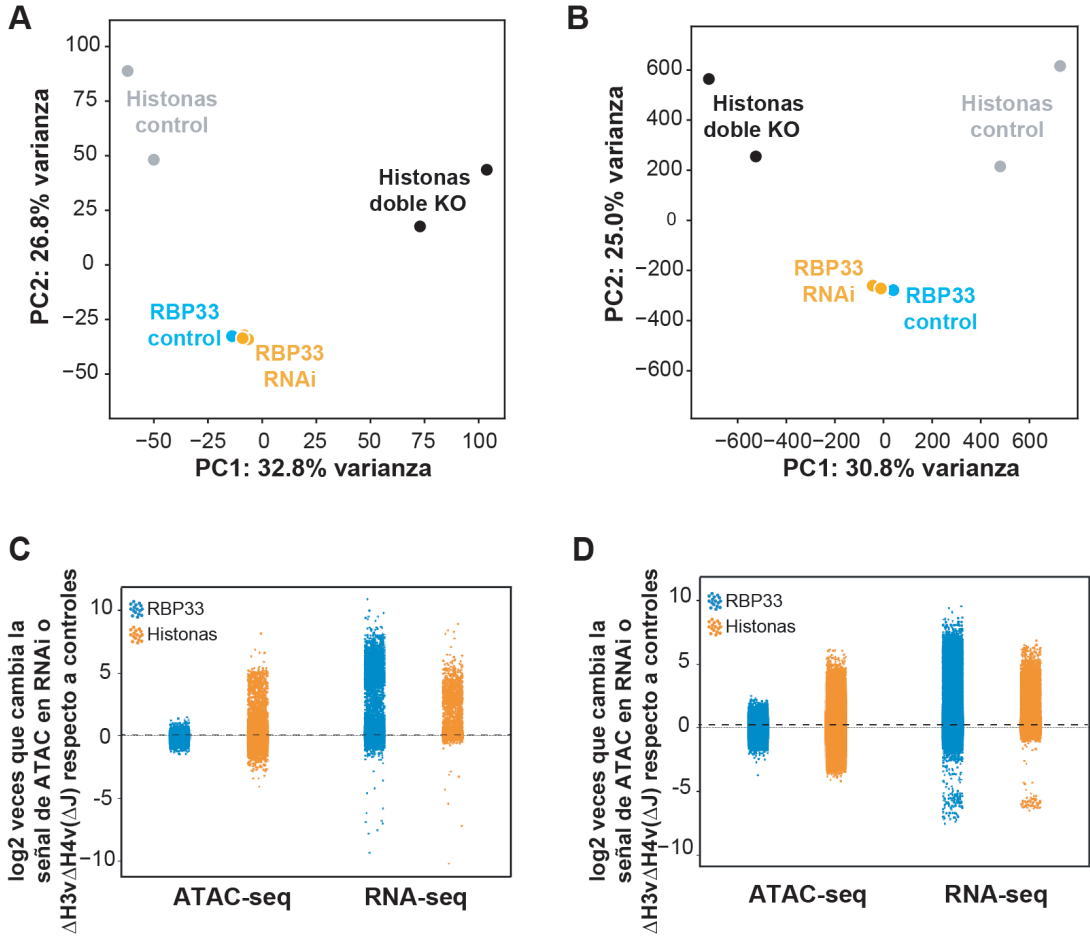


Figura 22. Comparación entre las muestras de ATAC-seq de RBP33 y el doble KO $\Delta H3v\Delta H4v$. (A y B) Análisis de componentes principales (PCA) de las muestras de RBP33 y el doble KO $\Delta H3v\Delta H4v$ a nivel de genes (A) o de todo el genoma (B). Cada punto representa una réplica. PC, componente principal. (C y D) Gráfico de bandas donde se representa el promedio de los ratios log₂ de la señal de ATAC del RNAi de RBP33 relativo a su control (azul), o del doble KO $\Delta H3v\Delta H4v$ relativo a sus controles (naranja), a nivel de genes (C) o de todo el genoma (D). Para comparar, se muestra al lado el mismo análisis realizado con las muestras de RNA-seq de RBP33 y del triple KO $H3v\Delta H4v\Delta J$. Cada punto representa un gen (C) o una 'bin' de 100 pb (D). La línea discontinua indica log₂=0, es decir, que no hay cambios en la accesibilidad de la cromatina (ATAC-seq) o en la expresión génica (RNA-seq).

Estos hallazgos indicaron que no hay grandes cambios en la accesibilidad de la cromatina en ausencia de RBP33 en comparación con los observados en el doble KO $\Delta H3v\Delta H4v$. De hecho, cuando se analizaron más específicamente los loci correspondientes a MES VSGs, cSSRs y a

RESULTADOS

genes que flanquean a las cSSRs (genes ‘últimos’), donde se enriquecen H3v y H4v, se observó que la cromatina estaba significativamente más abierta en los tripanosomas $\Delta H3v\Delta H4v$ tal y como se esperaba, mientras que el silenciamiento de RBP33 no produjo un aumento significativo en la accesibilidad de la cromatina en estas zonas (Figura 23A). Los gráficos de cobertura que representan el promedio de la señal de ATAC en 47 cSSRs diferentes de más de 1,2 kb de extensión confirmaron estas observaciones (Figura 23B).

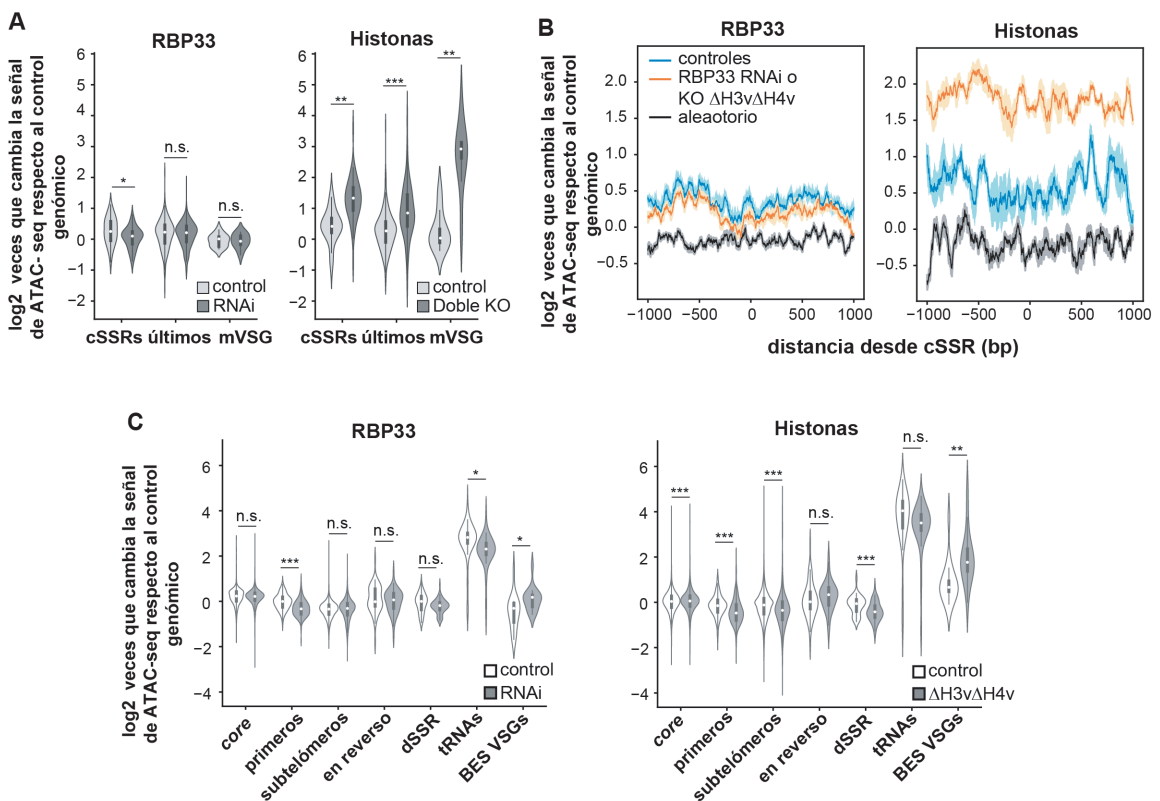


Figura 23. Análisis de la señal de ATAC en zonas representativas del genoma en las muestras de ATAC-seq de RBP33 y el doble KO $\Delta H3v\Delta H4v$. (A) Gráficos de violín que muestran la accesibilidad de la cromatina en los controles (violines gris claro), en tripanosomas donde se silenció RBP33 mediante RNAi y en el doble KO $\Delta H3v\Delta H4v$ (violines gris oscuro). Se obtuvieron los valores medios de la señal de ATAC en las distintas réplicas para cSSRs, genes VSG “últimos” y genes de VSG metacíclicas, y se expresaron como el ratio en relación al control de DNA genómico (transformado a log₂); solo se consideraron los loci con valores de CPM ≥ 1 en al menos dos réplicas biológicas. Dentro de cada violín se muestran diagramas de cajas en miniatura (las cajas representan los rangos intercuartílicos (*interquartile range*,

IQR); bigotes, $\pm 1,5$ RIC; círculos blancos, medianas). Se usaron pruebas t de Student de dos colas para evaluar si la accesibilidad de la cromatina se alteró significativamente en las muestras donde se silenció RBP33 mediante RNAi y en el doble KO $\Delta H3v\Delta H4v$, en relación a las muestras control. n.s., no significativo ($p > 0,5$); *, $p < 5 \times 10^{-3}$; **, $p < 5 \times 10^{-6}$; ***, $p < 5 \times 10^{-20}$ ($n=58$, $n=279$ y $n=8$ para cSSRs, genes 'últimos' y VSG metacíclicas, respectivamente). (B) Gráficos de cobertura de base única de ATAC-seq para los loci cSSRs. Se obtuvieron señales de ATAC en regiones genómicas centradas en la cSSR y extendidas 1000 pb a cada lado. La señal de ATAC se obtuvo en regiones genómicas centradas en cSSRs y extendidas 1000 pb a cada lado, se normalizaron por tamaño de biblioteca (RPM), y se representan como el ratio relativo al control de DNA genómico (transformado a \log_2). Los valores se representan como la media (líneas continuas) \pm SEM (áreas sombreadas) de las señales obtenidas en 47 cSSRs diferentes de más de 1,2 kb. También se muestran las señales de ATAC correspondientes al promedio de 47 regiones elegidas al azar del mismo tamaño (línea negra). (C) Gráficos de violín que muestran la accesibilidad de la cromatina en los controles (violines blancos), en ausencia de RBP33 (RNAi) o en el doble KO $\Delta H3v\Delta H4v$ (violines grises). Se utilizaron pruebas t de Student de dos colas para evaluar si la accesibilidad de la cromatina se alteró significativamente en muestras RBP33 RNAi o en el doble KO $\Delta H3v\Delta H4v$ en relación con sus respectivas muestras control. n.s., no significativo ($p > 0,15$); *, $p < 0,05$; **, $p < 0,005$; ***, $p < 5 \times 10^{-6}$.

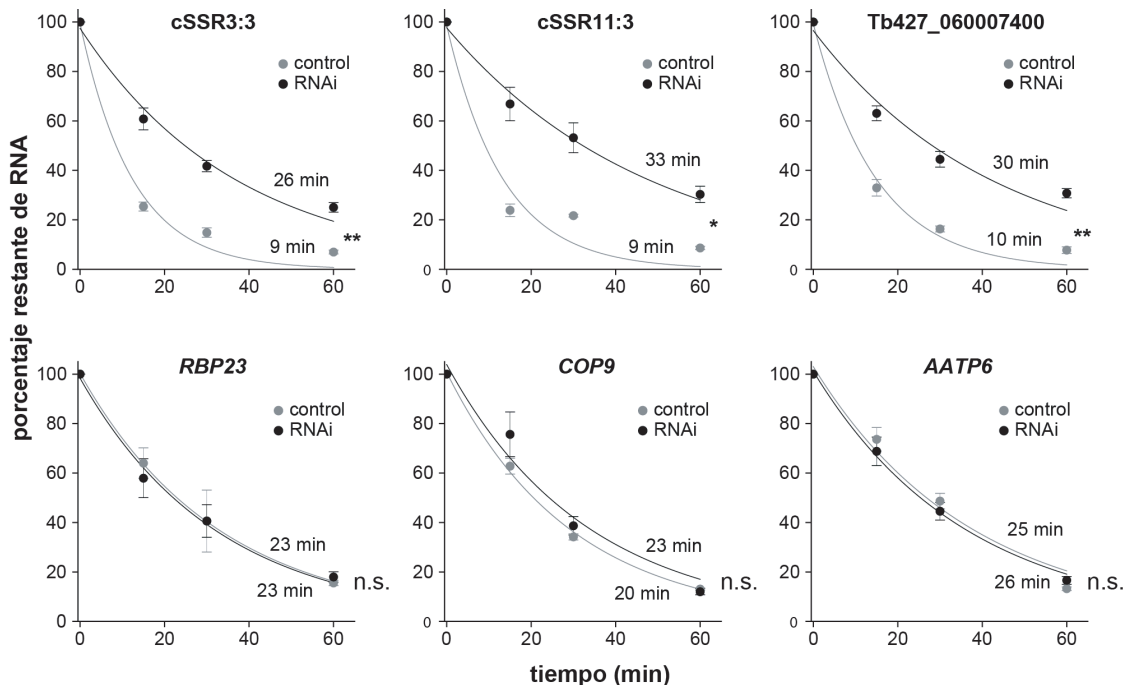
Estos resultados indican que la regulación positiva observada de los transcritos derivados de estas regiones no parece deberse a una mayor accesibilidad de la cromatina en las células RNAi RBP33. De hecho, la cromatina parece estar significativamente menos abierta en cSSRs tras el silenciamiento de RBP33 (Figura 23A) así como en los loci de tRNA, dSSR y genes 'primeros' (Figura 23C). Por otro lado, los loci BES VSGs mostraron un aumento pequeño pero significativo ($p = 0,03$) en la accesibilidad (Figura 23C); sin embargo, estos cambios puntuales en la estructura de la cromatina no dieron como resultado una regulación positiva de los transcritos de BES VSG.

8. El silenciamiento de la expresión de RBP33 por RNAi da como resultado la estabilización de los transcritos que se regulan

Para evaluar si los cambios observados en la abundancia de RNA se debieron a alteraciones en las tasas de degradación del RNA se analizó, en tripanosomas procíclicos, la degradación de los transcritos derivados de dos loci cSSR diferentes y de un gen 'último', Tb427_060007400, que flanquea una tercera cSSR. La localización genómica de estos transcritos se muestra en la Figura 24.

RESULTADOS

Como controles, medimos la degradación de tres transcritos que codifican proteínas *housekeeping*: *AATP6* (Tb427_080052600/Tb927.8.4700), *COP9* (Tb427_100084600/Tb927.10.7800) y *RBP23* (Tb427_100119800/Tb927.10.11270). *AATP6* y *COP9* son genes 'core' ubicados en medio de dos PTUs, mientras que *RBP23* es el primer gen que se transcribe dentro de una tercera PTU. Las células se incubaron con actinomicina D y sinefungina para inhibir la síntesis y la maduración del mRNA simultáneamente (Li et al., 2006) y se recogieron muestras a diferentes tiempos durante un período total de una hora. Los seis mensajeros siguieron un patrón de degradación exponencial, como se muestra en la Figura 25 ($R^2 > 0,95$ para todos los ajustes de curva). Se observó que las vidas medias de los transcritos derivados de las cSSRs y Tb427_060007400 aumentaron aproximadamente tres veces en ausencia de RBP33, mientras que la degradación de los mensajeros control no se alteró significativamente.



RESULTADOS

Figura 25. Efecto del silenciamiento de RBP33 en la estabilidad de los transcritos.

Se añadió actinomicina D y sinefungina a los cultivos de tripanosomas procíclicos control (gris) o RBP33 RNAi (negro). Las muestras de RNA se recogieron a varios tiempos y se analizaron mediante RT-PCR cuantitativa. Cada punto representa el valor de la media \pm SEM de tres inducciones de RNAi independientes. Se indican los valores de vida media estimados (superior, RBP33 RNAi; inferior, control). Se usaron pruebas t de Student de dos colas para evaluar si las vidas medias se alteraron significativamente tras el silenciamiento de RBP33. n.s., no significativo ($p > 0,4$); *, $p < 0,05$; **, $p < 0,005$.

9. Interacción de RBP33 con otras proteínas

Se usó el método de purificación por afinidad en tándem (TAP) acoplado a espectrometría de masas para analizar si RBP33 se une a otras proteínas en la célula. Para ello, se utilizó una línea celular previamente generada en nuestro laboratorio por Sandra Fernández Moya, en su tesis doctoral titulada 'Caracterización funcional de las proteínas de unión a RNA RBP33 y DRBD3 como reguladoras de la expresión génica de *Trypanosoma brucei*, 2013 (resultado publicado en (Gómez-Liñán et al., 2022)). Esta línea celular procíclica expresa, desde el locus endógeno, una versión funcional de RBP33 fusionada a una etiqueta TAP en el extremo N-terminal.

Los complejos proteicos fueron purificados usando el método TAP (Puig et al., 2001) con las modificaciones descritas en (Fernández-Moya et al., 2012), en dos purificaciones independientes. El material purificado se sometió a espectrometría de masas (LC-MS/MS) y las proteínas se identificaron como se describe en el apartado 2.20. de materiales y métodos. Como controles, se realizaron purificaciones en paralelo utilizando extractos de proteínas obtenidos de líneas celulares que expresaban la etiqueta TAP sola o el complejo de ribonucleoproteínas PuREBP1/2 (Rico-Jiménez et al., 2021) utilizando PuREBP1-TAP o TAP-PuREBP2 como cebo. También se compararon las proteínas asociadas a RBP33 identificadas con una lista de supuestos contaminantes comúnmente observados en las purificaciones TAP (Ouna et al., 2012). Se consideraron proteínas asociadas a RBP33 sólo las que cumplieron los siguientes criterios: i) se detectaban en ambas purificaciones TAP-RBP33; ii) no estaban presentes más de una vez en la

lista de posibles contaminantes, TAP sola y purificaciones PuREBP1/2; y iii) se consideran proteínas nucleares según las anotaciones de la base de datos o estudios proteómicos (Goos et al., 2017). Tras este filtro se obtuvo una lista de seis proteínas todas ellas posibles proteínas de unión a RNA (Tabla 5). Cabe destacar que tres de las proteínas asociadas a RBP33 sean conocidos factores de *splicing*: RBSR1 (ortólogo al factor de *splicing* humano SRSF7 (Wippel et al., 2019)), el factor de *splicing* basal SF1 (Gupta et al., 2013) y U2AF65/RBSR4 (Vazquez et al., 2009).

GeneID	Descripción
Tb927.10.13720	Proteína de unión a RNA RBP29
Tb927.6.1470	Proteína hipotética, conservada
Tb927.9.6870	Proteína de unión a RNA RBSR1
Tb927.10.9400	Factor de <i>splicing</i> SF1
Tb927.2.4710	Proteína de unión a RNA RRM1
Tb927.10.3500	Factor de <i>splicing</i> U2AF2 (RBSR4)

Tabla 5. Proteínas asociadas a RBP33 en células procíclicas de *T. brucei*.

10. El exosoma de *T. brucei* está involucrado en la degradación de los transcritos derivados de cSSRs

Como se ha mostrado previamente, los transcritos originados en los loci de cSSR se estabilizan tras el silenciamiento de la expresión de RBP33 mediante RNAi, lo que indica que normalmente estos mensajeros son degradados en condiciones fisiológicas. En otros eucariotas, el complejo exosoma nuclear juega un papel central en la vigilancia del RNA, al degradar los transcritos producidos como consecuencia de la transcripción ‘no productiva’, garantizando así la calidad de las moléculas del mRNA (ver apartado 2.3.2. de introducción).

RESULTADOS

Los tripanosomas también poseen un complejo exosoma (Estévez et al., 2001; Haile et al., 2003) que se localiza principalmente en el núcleo (Haile & Papadopoulou, 2007; Kramer et al., 2016). Para evaluar si el exosoma es responsable de la degradación de los transcritos que derivan de cSSRs, se generaron dos líneas celulares RNAi (tanto en sanguíneos como en procíclicos), que expresaban de manera inducible RNA de doble cadena dirigido a los mRNAs que codifican las subunidades catalíticas del exosoma RRP44 y RRP6 (como se describe en (Estévez et al., 2001)). El silenciamiento de la expresión de las dos subunidades fue confirmado mediante inmunoblot (Figura 26A).

A continuación, se seleccionaron transcritos procedentes de cuatro loci cSSRs diferentes y se analizaron los cambios en su expresión mediante RT-PCR cuantitativa en ausencia del exosoma o de RBP33. Se observó una marcada acumulación de transcritos derivados de cSSRs en los cuatro loci en las líneas celulares sanguíneas (Figura 26B) y procíclicas (Figura 26C) donde se había silenciado la expresión del exosoma. En los tripanosomas procíclicos, el silenciamiento del exosoma condujo a un fenotipo incluso más acusado que el observado tras la ausencia de RBP33, mientras que el efecto observado en las formas sanguíneas fue similar tanto en ausencia del exosoma como de RBP33. Esto podría deberse a una menor eficiencia de RNAi en sanguíneos, ya que se pueden detectar niveles de proteína residual tras la inducción del RNAi en las muestras sanguíneas, especialmente en la línea celular RNAi de RRP6 (Figura 26A).

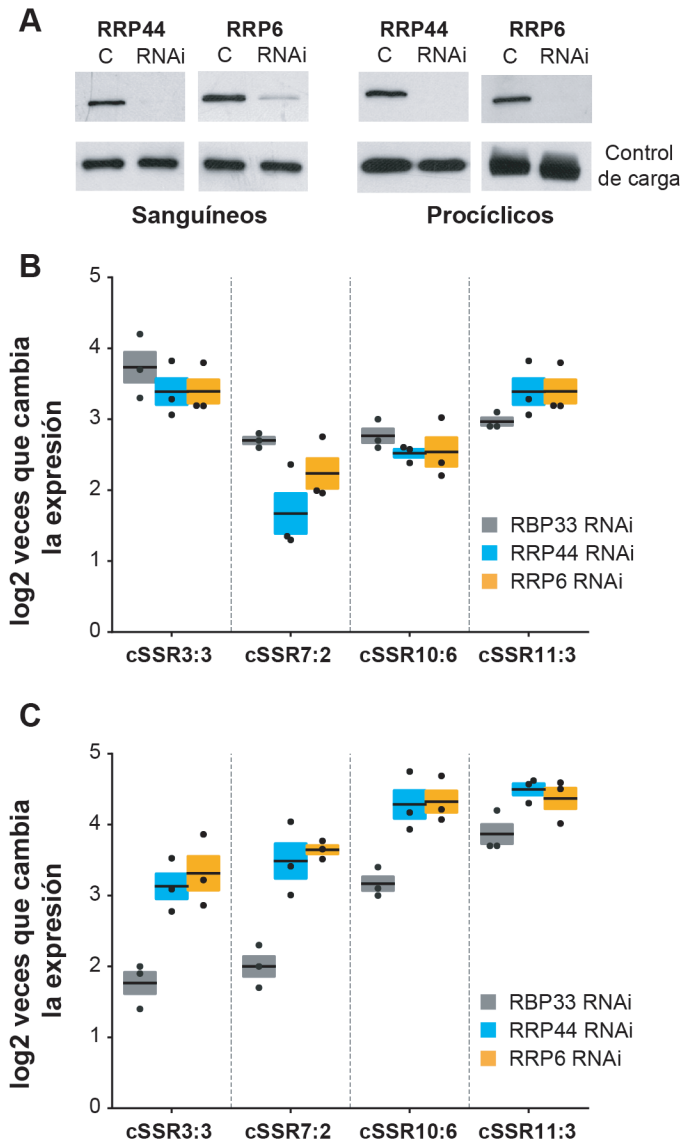


Figura 26. Efecto del silenciamiento del exosoma en la abundancia de los transcritos derivados de cSSRs. (A) Se generaron líneas celulares procíclicas y sanguíneas que expresaban RNA bicatenario correspondiente a las subunidades RRP44 o RRP6 del exosoma, en un sistema inducible por tetraciclina. El silenciamiento de cada proteína fue comprobado mediante inmunoblot tras 48 h de inducción por tetraciclina (RNAi). C, línea control (parental). RBP33 se utilizó como control de carga en todas las muestras excepto en el inmunoblot de procíclicos, donde se usó α -tubulina en su lugar. (B y C) análisis por RT-PCR cuantitativa para medir los niveles de transcritos derivados de cuatro SSR convergentes. Los cambios en la expresión se representan como la media (líneas horizontales) \pm SEM (área sombreada) de tres inducciones de RNAi independientes en tripanosomas sanguíneos (B) o procíclicos (C).

DISCUSIÓN

DISCUSIÓN

Los tripanosomátidos son protozoos parásitos que incluyen especies que causan enfermedades importantes en el ser humano. En concreto, *T. brucei* es el agente etiológico de la tripanosomiasis africana humana (HAT) o enfermedad del sueño, cuyo vector de transmisión es la mosca tsé-tsé y sigue siendo un problema de salud pública en muchos países del África subsahariana. Los tripanosomátidos se separaron de otros linajes eucariotas hace al menos 500 millones de años, lo que les otorga una gran importancia como modelo de estudio de determinados procesos que han evolucionado de manera distinta (Dacks et al., 2008; Lukeš et al., 2014). Además, estos organismos se pueden cultivar y manipular genéticamente de manera sencilla en el laboratorio, por lo que son excelentes modelos para el estudio de la regulación de la expresión génica.

Una de las características más llamativas de estos protozoos es su organización genética: los genes se agrupan en unidades de transcripción policistrónica (PTUs), de manera que la transcripción se inicia y termina en unos pocos sitios a lo largo del cromosoma, que están enriquecidos en distintas variantes de histonas (Martínez-Calvillo et al., 2003; Siegel et al., 2009). Más de cien genes pueden co-transcribirse de manera constitutiva a partir de una misma PTU. Aunque el orden de los genes está parcialmente conservado (Jackson et al., 2008), a diferencia de los operones bacterianos, no existe una relación funcional entre las proteínas que codifican. De hecho, los transcritos derivados de una misma PTU pueden mostrar patrones de expresión diferentes a lo largo del ciclo celular, en los distintos estadios del ciclo de vida del parásito o bajo condiciones de estrés. Los mensajeros monocistrónicos se generan mediante procesos de *trans-splicing* y poliadenilación del RNA policistrónico, y el control de los niveles de mRNA se realiza principalmente a nivel postranscripcional, en lugar de a través de promotores o factores de transcripción específicos como ocurre en el resto de eucariotas.

DISCUSIÓN

Para lograr una infección exitosa y persistente, las formas sanguíneas del parásito expresan una densa capa superficial que consiste en una glicoproteína de superficie variante (VSG), que van intercambiando para evadir la respuesta inmune del huésped mamífero; este mecanismo se conoce como variación antigénica (Horn, 2014). Los genes VSG que se expresan, se encuentran cerca de los telómeros y un mecanismo estrictamente regulado de expresión monoalélica asegura que la RNA polimerasa I solo transcriba un gen VSG en un determinado momento desde uno de los sitios de expresión (BES) (Figueiredo et al., 2009; Günzl et al., 2003). Las formas metacíclicas también están cubiertas por una capa de VSG que se transcribe desde uno de los sitios de expresión metacíclica (MES), que les sirve de preadaptación para la infección del huésped mamífero (Graham & Barry, 1995). Los tripanosomas metacíclicos muestran diferentes variantes de VSG; sin embargo, cada célula individual expresa solo un MES VSG específico (Tetley et al., 1987; Vigneron et al., 2020). A excepción del gen VSG activo, la mayoría de los subtelómeros están silenciados en *T. brucei* (Müller et al., 2018).

Los avances en las tecnologías de secuenciación de RNA han revelado que la transcripción por parte de la RNA polimerasa II es ubicua y los transcritos generados cubren casi todo el genoma eucariótico (Jensen et al., 2013). La transcripción generalizada da lugar a una gran cantidad de RNAs no codificantes y aparentemente no productivos, que se generan principalmente en los sitios de terminación y se degradan mediante vías de vigilancia que involucran al complejo exosoma (Lykke-Andersen et al., 2021; Villa & Porrua, 2022). La terminación de la transcripción en animales y hongos está relacionada con la poliadenilación del RNA, pero este no puede ser el caso en organismos como los tripanosomátidos, donde la transcripción es policistronica. De hecho, estos organismos carecen de señales de poliadenilación específicas y los sitios de poliadenilación se eligen según la distancia a las señales de *trans-splicing* presentes en el cistron inmediatamente aguas abajo. Además, la transcripción y el procesamiento parecen estar desacoplados en estos protozoos.

Los sitios de iniciación de la transcripción (TSS) generalmente se ubican en dSSRs que separan PTU divergentes, mientras que los sitios de terminación (TTS) generalmente están restringidos a cSSRs y regiones próximas a genes transcritos por otras polimerasas (Clayton, 2019; Siegel et al., 2009). La información sobre la terminación de la transcripción por parte de la RNA polimerasa II es muy escasa en los tripanosomas. Los TTS se enriquecen en las marcas de cromatina H3v, H4v y una modificación de DNA específica de tripanosomátidos, la base J. La pérdida de algunos de estos factores conduce a la aparición de transcritos antisentido, así como a una desregulación de la expresión de la VSG (Kim, 2021; Müller et al., 2018; Reynolds et al., 2014, 2016; Schulz et al., 2016; van Luenen et al., 2012). En general, se asume que la presencia de marcas epigenéticas específicas (H3v, H4v y base J) es suficiente para una eficiente terminación de la transcripción en los tripanosomátidos, y que tanto los subtelómeros como las cSSRs no se expresan simplemente porque no se transcriben (Kim, 2021; Müller et al., 2018).

Un estudio previo en nuestro laboratorio mostró que la proteína de unión a RNA RBP33 era capaz de unirse específicamente a RNAs derivados de cSSRs y de otras regiones silenciadas en el genoma de *T. brucei*, por lo que esta proteína podría estar implicada en la regulación de la abundancia de este tipo de transcritos (Clayton, 2019; Fernández-Moya et al., 2014). En este trabajo, hemos investigado esta hipótesis y hemos demostrado que el silenciamiento de la expresión de RBP33 mediante RNAi conduce a un aumento de 7 a 8 veces en los niveles de transcritos antisentido que surgen principalmente de TTS tanto en tripanosomas sanguíneos como en procíclicos. También observamos una sobreexpresión de los genes subteloméricos en ausencia de RBP33, especialmente en las formas sanguíneas. No sabemos si esto se debe a diferentes eficiencias del RNAi o si refleja funciones distintas de RBP33 en cada forma de vida. Podemos descartar que este fenotipo sea consecuencia de la muerte celular provocada por la ausencia de RBP33, ya que no se observa acumulación de transcritos derivados de cSSRs tras la eliminación de DRBD3, una

DISCUSIÓN

proteína de unión a RNA no relacionada con RBP33 y que también es esencial (Fernández-Moya et al., 2014). Además, el transcriptoma apenas se altera en células carentes de otra proteína esencial de unión a RNA, PuREBP1 (Rico-Jiménez et al., 2021). El hecho de que RBP33 se una a los transcritos que luego vemos sobreexpresados en su ausencia, respalda claramente un papel en la regulación de la abundancia de estos RNAs. Los efectos globales observados en el transcriptoma al silenciar la expresión de RBP33 nos indican que existe una transcripción generalizada en los tripanosomas tal y como se ha observado en otros organismos durante la última década.

Al comparar nuestros datos con otros previamente publicados donde se estudiaron mutantes $\Delta H3v\Delta H4v(\Delta J)$ (Kim, 2021; Müller et al., 2018), identificamos una serie de diferencias importantes. RBP33 es esencial (Fernández-Moya et al., 2014), mientras que ninguna marca de cromatina es esencial por sí misma, aunque su ausencia dé lugar a una acumulación de transcritos provenientes de TTS (Kim, 2021; Reynolds et al., 2014). Es importante destacar que los tripanosomas procíclicos carecen de base J y pudimos observar una marcada acumulación de transcritos derivados de TTS incluso en esta forma del parásito. También vimos un efecto mucho mayor en la abundancia de transcritos antisentido y la sobreexpresión de genes subteloméricos en ausencia de RBP33 en comparación con los datos de $\Delta H3v\Delta H4v(\Delta J)$. Sin embargo, aunque las muestras de RNA-seq de $\Delta H3v\Delta H4v(\Delta J)$ se obtuvieron a partir de líneas celulares *knockout* condicionales inmediatamente después del *knockout*, estas células se basan en mutantes $\Delta H4v$ (Kim, 2021) y, por lo tanto, no podemos descartar que los efectos menos pronunciados observados en el triple KO puedan deberse a una adaptación por parte de los tripanosomas.

Tal y como hemos descrito en la introducción de esta tesis doctoral, la variación antigénica es otra de las características más importante de este parásito. Por ello, nos pareció interesante averiguar si la abundancia de los transcritos VSG (tanto BES como MES) se alteraban en ausencia de RBP33. Tras analizar las regiones subteloméricas se determinó que

los transcritos que codifican los genes *VSG* de MES se sobreexpresaban tras el silenciamiento de *RBP33*, mientras que la expresión de los genes *VSG* de BES no se alteró. Esto sugiere que los mecanismos involucrados en la represión de la *VSG* son, al menos en parte, diferentes en MESs y BESs, como ya propusieron otros autores (Christiano et al., 2017). Estudios realizados en el laboratorio del Dr. Siegel (Munich, Alemania), utilizando la captura de conformación cromosómica (Hi-C) del genoma completo, demostraron que los BES (pero aparentemente no los MES) se agrupan mediante interacciones intercromosómicas que los unen a sitios nucleares específicos (Müller et al., 2018). Según estos datos, el alto grado de compactación entre los BES los protegería de una transcripción “esporádica” y garantizaría la expresión monoalélica de un solo gen *VSG* desde un determinado BES. En el mutante $\Delta H3V$ se perturbarían dichos enlaces intercromosómicos, lo que conduciría a una pérdida de las restricciones que anclan los BES a sitios nucleares específicos, resultando en la expresión de muchos transcritos de *VSG*. Esto podría explicar por qué no vemos este fenotipo tras silenciar la expresión de *RBP33*, ya que estas restricciones físicas continuarían existiendo y protegiendo a los BES (pero no a los MES) de la transcripción.

Con respecto a los MES *VSG*, hay muy poca información disponible en la bibliografía debido a la difícil obtención de esta forma del parásito. Sin embargo, en la actualidad es posible conseguir formas metacíclicas de una forma relativamente sencilla, simplemente sobreexpresando la proteína *RBP6*, lo que facilita su estudio (Kolev et al., 2012). Tal y como se describe en la introducción, los MES presentan algunas diferencias con respecto a los BES: son unidades de transcripción monocistrónica en lugar de policistrónica como ocurre con las BES y, en estos últimos, los *ESAGs* y el gen *VSG* suelen estar separados por una región de 5-30 kb compuesta por repeticiones de 70-76 pb (Campbell et al., 1984), mientras que en los MES esta región es muy corta o no existe (Barnes et al., 1990). Quizá esta extensión de repeticiones actúe como una barrera para asilar aún más al gen *VSG* en los BES. Además, podría ser que los MES no se adhieran a

DISCUSIÓN

sitios específicos en el núcleo; de hecho, a diferencia de los tripanosomas sanguíneos, los metacíclicos pueden mostrar diferentes variantes de VSG, lo que sugiere interacciones más débiles o ausentes entre los MES. Nuestros resultados sugieren que los MES, pero no los BES, se transcriben de manera generalizada y, por tanto, solo los transcritos correspondientes a las MES VSG se regulan al alza en las células sin RBP33.

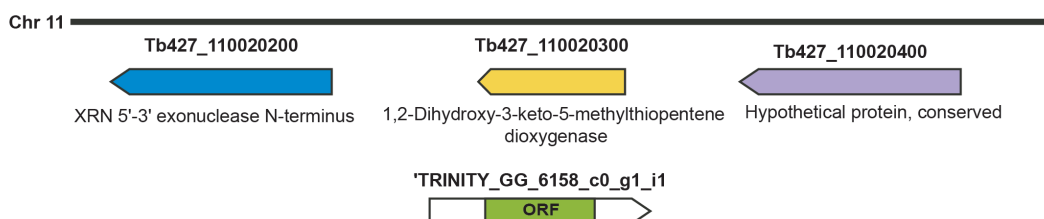
Por otro lado, otros autores determinaron que la transcripción de RBP33 se regulaba positivamente en células metacíclicas generadas *in vitro* mediante la expresión inducible de una versión mutante de la proteína de unión a RNA RBP6 (Shi et al., 2018). Sin embargo, los niveles de proteína y transcrito RBP33 no se vieron alterados en tripanosomas metacíclicos obtenidos mediante la expresión de RBP6 de tipo silvestre (Christiano et al., 2017). Se sabe muy poco acerca de cómo todos los MES se mantienen inactivos en las formas sanguíneas de *T. brucei*. Investigar si RBP33 tiene un papel en el establecimiento de formas metacíclicas podría ser una futura vía de investigación.

La ausencia de RBP33 también provocó una regulación a la baja de muchos genes esenciales. Esto de por sí puede explicar por qué RBP33 es esencial. De hecho, se ha demostrado que la transcripción antisentido altera la expresión de genes cercanos en otros organismos (Huber et al., 2016; Xu et al., 2011). Asimismo, hemos demostrado que al menos algunos de los transcritos antisentido que surgen como consecuencia del silenciamiento de RBP33 parecen procesarse correctamente y, por lo tanto, son potencialmente traducibles. De hecho, un transgén *eYFP* insertado en orientación opuesta en dos loci convergentes distintos, que en condiciones fisiológicas se encuentran silenciados, se expresa eficazmente de manera dependiente de RBP33. Por lo tanto, el silenciamiento de RBP33 podría conducir a la aparición de proteínas y péptidos potencialmente dañinos.

Por ejemplo, los lncRNAs carecen de ORFs canónicas anotadas y, por lo tanto, se asume que no tienen funciones codificantes, sino más bien mediadas por su secuencia y estructura del RNA (Statello et al., 2021).

Sin embargo, cientos de RNAs largos no codificantes (lncRNAs) están poliadenilados y se asocian a ribosomas en el citoplasma, lo que indica su idoneidad para ser traducidos y producir micropéptidos (Zaheed et al., 2021). Además, estudios recientes en *Drosophila melanogaster* demuestran que de entre todos los lncRNAs sin capacidad codificante identificados previamente, alrededor del 30% sí que son capaces de ser traducidos y originar péptidos cortos (Patraquim et al., 2022). Los autores proponen que, dado que los que se traducen están sujetos a selección natural y poseen un origen filogenético más antiguo que los no traducidos, pueden dar lugar a nuevos genes con el paso del tiempo (Patraquim et al., 2022). En *T. brucei*, algunos de los transcritos que surgen tras la pérdida de RBP33 contienen péptidos con un alto potencial codificante (Figura 27A-B), por lo que podrían dar lugar al establecimiento de nuevos genes durante el proceso evolutivo. Aunque la función de estos péptidos se desconoce, su estudio en profundidad es interesante tanto del punto de vista terapéutico como evolutivo.

A



B

ATTAGAACAGTTTCTGACTATATTGACATACTTTTTCTTTTTCTTTTCATGCCTTTGTCCATGTGGTGATGATAATTGTTATT
 AGCGTAAGATAACAAGACGTAAAACATGAAAAAAAAATAATAAAATAAAGATGAAGATGAAGATAAAAAAGGAAGCGTG
 GAAAGGCATCACCTGAATTACAACACTACAATTTCTCCACCCACTTCTCATAACCCTCACAGGGAGTGGACTCTCTGCAAATA
 TCTTTTAATCCGCCCTCGCCTGCTGGAACACCATAAATGAGGAATTCAGTCACCTTAATGAACGGATGCGTGCGGGT
 AGAGATATTTGGGTTACCCAAATATACGGAGCGTTCAACATTGCATTGCACATAAACGACGCGGGCGCTTCTCCTTCCG
 TGCCCTCACTGCGCGCTGCACCATTTGGCTCTGCAAGAACACACGGAGGACACCACGACTTGTGGGTTTTGGGATGCG
 ACACACCTTTTATAAGCACCACAAGGATATCACCGGGTTGCAGTGCAGCTTTGACCTTCTCAAGTGTAGCATCGAAGT
 CCCTTGGATTTTTACCAGAATGTTATCTGTATCATTGCTGGGCCATCGCTGTTTCGACTGGCGGATTGTTAATAAAA
 AACATGTCTCCTCATGCGCTTCCACCGCTTCCCTTGACAAGTGCTTCGGAAAAGTCCAGCGGTGTCACGAGGAATGGG
 GCGAAAGGTTTCTGCGCTTAAAAAGACGGAGGAAGTGGCGGGTCCGAGTTGGAAGAAACAACCTCGATGATAC
 ATTCCGGGAGGGAGAACAATCACGTCACCCATTTCCCTCAATGCGTACCCACTCATCTTCTCATTGCGCACATCCA
 CATAAATAATACCCTTTCCACAATTCGCATTTGTTTCATCCACGTTAAGGTGTTCCATACTCTCATCCTTCGACGATGTG
 GTAACGTAACCTTCTCCAGTGTGTTTGTAAACCGAGGTTCCAAATGAGTCGCTGGGCGTACGTACGGCCTTCATCGTCAT
 CGGGGTGGCGTTCGTGAAGTGTTCGCGACATCTATGCTGCGACTATAAACAAGAAATGTTTCCAATTGCACATCCGTC
 CGGGGACGCTCGGTGAGCAGTGACATGGTGCAGTAATATCACCCCTTGTGTGTTCAATAACCAACTGCTCGGCAATGTT
 TGATCTTCCGTTTTGTTTTCTTTTTTCTTTTTTGCACCTTAATTAACCTACCTGCACGGGTTTCGAAAAGTTGACGATGT
 CTGTAACCTCGGTGGGCCCCAATTAAGTCCACGATATCGGGAACGGGAAGGAAGAGGAGGAGAAAGAGGAGAGC

DISCUSIÓN

Figura 27. Ejemplo de un transcrito antisentido derivado de una región convergente en ausencia de RBP33. (A) Esquema correspondiente a la posición 446000-452513 del cromosoma 11 (Chr11). Se muestra el transcrito antisentido (TRINITY_GG_6158_c0_g1_i1) y la supuesta ORF identificada en color verde. Para calcular el potencial codificante se utilizaron los programas CPC2 (<http://cpc2.gao-lab.org>) y RNAsamba (<https://rnasamba.lge.ibi.unicamp.br>). Ambos programas dieron como resultado un potencial codificante >0.99. (B) Secuencia de nucleótidos del transcrito antisentido. Se indica en negrita y color azul la secuencia correspondiente al mini-exón SL y la supuesta ORF codificante en negrita.

Datos previos demuestran que en los tripanosomas que carecen de H3v, H4v y base J tiene lugar una sobreexpresión de los transcritos VSG de MES y BES. Además, estas líneas celulares exhiben una alta accesibilidad al DNA en los BES reprimidos y un aumento de transcritos derivados de TTS (Kim, 2021; Müller et al., 2018). Se sabe que la cromatina abierta favorece la iniciación de la transcripción por parte de la RNA polimerasa II en tripanosomátidos (McAndrew et al., 1998). Por lo tanto, resulta lógico pensar que el silenciamiento de RBP33 pueda ocasionar la apertura de cromatina en las regiones silenciadas, lo que a su vez permitiría la expresión de estos loci cromosómicos. Sin embargo, no se observó un aumento significativo en la señal de ATAC-seq en los loci cSSR y MES VSG en ausencia de RBP33, mientras que estas regiones sí que estaban significativamente más abiertas en los mutantes Δ H3v Δ H4v. Por lo tanto, los cambios en la accesibilidad de la cromatina no parecen ser responsables del aumento de los niveles de transcritos que se observan tras la pérdida de RBP33.

Es importante resaltar que hemos demostrado que los transcritos derivados de tres loci cSSR diferentes se estabilizan significativamente tras el silenciamiento de RBP33, mientras que las vidas medias de otros tres transcritos no regulados permanecieron inalteradas. Esto descarta una inhibición general de la maquinaria de degradación del RNA causada por la carencia de RBP33 o la propia muerte celular. Además, también mostramos que el exosoma parece ser responsable de la degradación de los transcritos derivados de cSSRs. Dado que RBP33 se une a los RNA

derivados de los sitios de terminación de la transcripción (Fernández-Moya et al., 2014), nuestros resultados sugieren que RBP33 está involucrada en guiar dichos transcritos, producto de la transcripción generalizada, hacia el exosoma para su degradación. Esta afirmación está respaldada por el hecho de que la sobreexpresión de RBP33 da como resultado una disminución general de los niveles de todos los mRNAs en *T. brucei* (Cirovic et al., 2017). Sin embargo, no podemos descartar que RBP33 ejerza efectos adicionales directos o indirectos sobre la transcripción, por ejemplo, mermando las tasas de elongación cerca de los TTS. Los llamativos efectos que se observan en el transcriptoma tras el silenciamiento de RBP33 proporcionan evidencias de la existencia de una transcripción generalizada en los tripanosomas tal y como se ha observado en otros organismos. Por lo tanto, muchos loci genómicos considerados tradicionalmente en *Trypanosoma* como “no expresados”, como las cSSRs y los subtelómeros, permanecerían silenciados no porque no se transcriban, sino porque los transcritos derivados de estas regiones están destinados a la degradación.

El hecho de que la transcripción generalizada se encuentre en los tripanosomas puede no parecer muy notable, ya que este concepto, aunque es relativamente nuevo, ya se ha descrito en varios organismos durante la última década. Sin embargo, los tripanosomas divergieron muy pronto del linaje eucariota y la regulación de la expresión de sus genes presenta aspectos aparentemente únicos. Desde una perspectiva evolutiva, el hecho de que la transcripción ubicua se observe en un eucariota en el que la transcripción por la RNA polimerasa II es constitutiva, y en el que la transcripción y el *splicing* están desacoplados, es bastante relevante. Por lo tanto, estos parásitos deben contar con mecanismos diferentes para hacer frente a la transcripción generalizada. De hecho, RBP33 parece ser un factor específico de estos organismos. Por tanto, los tripanosomátidos podrían ser organismos de gran interés para estudiar nuevos mecanismos involucrados en la regulación y el destino de los transcritos aberrantes en eucariotas.

DISCUSIÓN

La función biológica de la transcripción ubicua es actualmente un tema muy debatido. Una de las cuestiones principales es qué proporción de estos transcritos es realmente funcional. Muchos de los RNAs recientemente identificados no tienen potencial codificante y podrían ser un efecto secundario provocado por la carencia de nucleosomas que permite que la RNA polimerasa II continúe transcribiendo. Sin embargo, algunos aspectos de la transcripción generalizada contradicen la idea de que estos RNAs aberrantes carezcan de utilidad. Varios de estos transcritos son estables, por ejemplo los SUTs, que están presentes en levaduras silvestres aunque a niveles bajos. Otros tienen efectos en *cis* también en estos organismos. Por ejemplo, un CUT controla la expresión del gen metabólico *PHO84* (Camblong et al., 2007): la estabilización del CUT es responsable del silenciamiento génico a través de la desacetilación de histonas. Otro ejemplo sería en los loci *IMD2* y *URA2*, donde la transcripción de un CUT regula negativamente la expresión del gen aguas abajo. En cualquiera de los loci, tanto el CUT como el mensajero se transcriben desde el mismo promotor pero se originan en diferentes sitios de inicio de la transcripción y se ha propuesto que compiten por el complejo de preiniciación (Kuehner & Brow, 2008).

Dado que la mayoría de los transcritos generados durante la transcripción generalizada son antisentido, la información disponible en la bibliografía sobre estas especies de RNA puede arrojar luz acerca de sus posibles funciones. Existen numerosos ejemplos de cómo la transcripción antisentido es capaz de regular la expresión génica. Por un lado, puede afectar al inicio de la transcripción mediante interferencia transcripcional (Shearwin et al., 2005) mediante la oclusión de los sitios de unión debido al paso de la RNA polimerasa o por modificaciones de la cromatina. Esta interferencia también puede ser co-transcripcional, una vez iniciada la transcripción; las RNAs polimerasas colisionan o incluso una puede ser un obstáculo para otra que está elongando en ese momento (Gelfand et al., 2011; Pelechano et al., 2013). La transcripción antisentido también puede actuar a nivel postranscripcional, activando/inhibiendo la traducción de un

determinado transcrito (Carrieri et al., 2012; Kawano et al., 2007) o afectando a la estabilidad del mensajero (Faghihi et al., 2010; Lasa et al., 2011).

Por otro lado, la propia transcripción (en lugar del transcrito antisentido generado) puede inducir modificaciones en la cromatina en determinadas regiones y afectar a la expresión génica (van Werven et al., 2012). Por tanto, aunque algunos de los eventos de transcripción no codificantes puedan ser triviales, cada vez es más evidente que otros sí son funcionales, ya sea a través del acto de transcribir *per se* o, alternativamente, a través de la producción de un transcrito regulador (Berretta & Morillon, 2009).

En cualquier caso, las moléculas de RNA poseen una serie de características que las convierten en elementos con un gran potencial regulador: son móviles y se sintetizan/degradan rápidamente, lo que permite una rápida adaptación celular a las condiciones ambientales, y pueden unirse no solo a proteínas sino al DNA y al RNA a través de la complementariedad de sus secuencias. En consecuencia, los RNAs pueden actuar en *cis* o en *trans* a nivel transcripcional o postranscripcional. Se ha demostrado que algunos transcritos no codificantes que están presentes en condiciones fisiológicas en la célula, están regulados en diferentes condiciones de crecimiento (Dutrow et al., 2008; Xu et al., 2009) o durante el desarrollo (Stolc et al., 2004), por lo que quizá los transcritos originados a partir de la transcripción no productiva podrían tener un papel de este tipo.

Sin embargo, experimentalmente no es sencillo discriminar entre una función directa del RNA maduro en cuestión, su precursor naciente o el propio proceso transcripcional que lo origina. Tal vez por eso se han descrito muy pocos mecanismos funcionales para, por ejemplo, los lncRNAs en humanos (Guttman & Rinn, 2012; Hu et al., 2012; Jensen et al., 2013; Ulitsky & Bartel, 2013; Wang & Chang, 2011). Entre las funciones de los lncRNAs destacan la formación de ribonucleoproteínas, hacer de guía de complejos modificadores hacia sus dianas moleculares,

DISCUSIÓN

actuar como cebos para secuestrar otros RNA/proteínas, o los llamados “repelentes”, que impiden la unión de una proteína a su sitio de acción (Jensen et al., 2013). Es posible que los transcritos antisentido producto de la transcripción ubicua tengan funciones similares, aunque esto aún no se ha demostrado.

Quizás sea la propia transcripción generalizada la que tenga una importancia biológica, por ejemplo favoreciendo un determinado estado de la cromatina o mejorando el reclutamiento de factores que se asocian a la polimerasa (Schulz et al., 2013). En este contexto, una iniciación de la transcripción menos estricta también puede dar paso a una mayor modulación de la regulación. Por ejemplo, impidiendo que grandes regiones genómicas se silencien por completo, la célula evitaría la formación de dominios de cromatina demasiado compactos que de otro modo serían difíciles de reabrir. Podría haber miles de nuevos loci genéticos que han escapado a nuestra atención y que podrían ser clave para comprender determinados procesos más complejos en eucariotas (Mattick, 2007). La transcripción ubicua también expone al transcriptoma a una presión selectiva, lo que proporciona un abanico de posibilidades para la selección natural a nivel del RNA. Sin embargo, sigue siendo un desafío constante desentrañar cuáles de estos eventos son tolerados por la célula simplemente porque no son dañinos y cuáles se mantienen evolutivamente porque son útiles para la regulación (Jensen et al., 2013).

Otro aspecto fundamental es que la transcripción ubicua puede ser potencialmente dañina para la expresión y la integridad del genoma si no se mantiene regulada. Dada la importancia de este hecho, todos los organismos han desarrollado diferentes mecanismos para mantener este proceso bajo control, por ejemplo gracias a los nucleosomas y a las vías de terminación de la transcripción asociadas a la degradación de RNA, cuyo objetivo es atenuar la acumulación de transcritos no deseados. En levaduras una vez que la RNA polimerasa II y el transcrito naciente han sido liberados de la cromatina, el complejo NNS permanece unido

al RNA y promueve la degradación de los transcritos no codificantes por el exosoma nuclear y su cofactor, el complejo TRAMP. La ausencia de cualquiera de los componentes del complejo es letal (Han et al., 2020; Schulz et al., 2013).

Aunque dos de los componentes principales del complejo NNS (Nrd1 y Nab3) no se conservan en metazoos, es posible establecer un paralelismo entre levaduras y organismos más complejos. De hecho, los metazoos poseen un complejo de múltiples subunidades denominado *Integrator*, que juega un papel clave en la restricción de la transcripción generalizada al provocar una terminación temprana en los loci que producen RNA cortos no codificantes (Lykke-Andersen et al., 2021). La delección o mutación de las subunidades del complejo en ratón, *Drosophila*, *Caenorhabditis elegans* y el pez cebra tiene efectos letales en el desarrollo (Han et al., 2006; Hata & Nakayama, 2007; Kapp et al., 2013; Rutkowski & Warren, 2009). En este contexto, RBP33 podría ser una de las proteínas encargadas de amortiguar los productos de la transcripción generalizada en *T. brucei*.

¿Cómo actúa RBP33 en la degradación de los transcritos? Dado el amplio catálogo de transcritos regulados por RBP33, es difícil imaginar un mecanismo mediado por el reconocimiento de motivos de secuencia de RNA específicos por parte de RBP33. De hecho, no se pudieron identificar en el conjunto de transcritos unidos a RBP33 (Fernández-Moya et al., 2014), y se ha demostrado que RBP33 tiene poca especificidad para la secuencia de RNA *in vitro* (Cirovic et al., 2017). En mamíferos y *Drosophila*, la degradación del RNA nuclear vía exosoma no parece ser selectiva, sino que más bien existe una competición entre la destrucción y el procesamiento del RNA (Schmid & Jensen, 2018). Esta posibilidad también sería válida para los tripanosomas, donde se ha demostrado que la competición entre el procesamiento del mRNA y la degradación co-transcripcional del RNA juega un papel fundamental a la hora de determinar los niveles en estado estacionario de los mensajeros maduros (Fadda et al., 2014).

DISCUSIÓN

En células de mamífero, la destrucción de transcritos no productivos mediada por *Integrator* tampoco parece ser específica de secuencia, sino que depende de la velocidad de la RNA polimerasa II. Como consecuencia, la degradación del RNA es más eficiente cuando la polimerasa se detiene o se ralentiza, siendo especialmente notable en unidades transcripcionales que se expresan menos (Lykke-Andersen et al., 2021). Gracias a este mecanismo la célula es capaz de “limpiar” toda aquella actividad transcripcional que se encuentre por debajo de un umbral funcionalmente significativo (Lykke-Andersen et al., 2021).

Un escenario similar sería plausible en *Trypanosoma*, donde los transcritos no productivos y de baja expresión, que son producto de la atenuación de la elongación de la transcripción cerca de los TTS o de loci que apenas se transcriben, serían guiados por RBP33 hacia el exosoma para su destrucción (Figura 28). Estos precursores de RNA probablemente sean más susceptibles a la degradación en comparación con los mensajeros que se encuentren por encima de un nivel determinado. Esto está respaldado por nuestra observación de que un transgén *eYFP* insertado en distintos loci cSSRs no se exprese en condiciones fisiológicas pero, en ausencia de RBP33, sí se detecte fluorescencia.

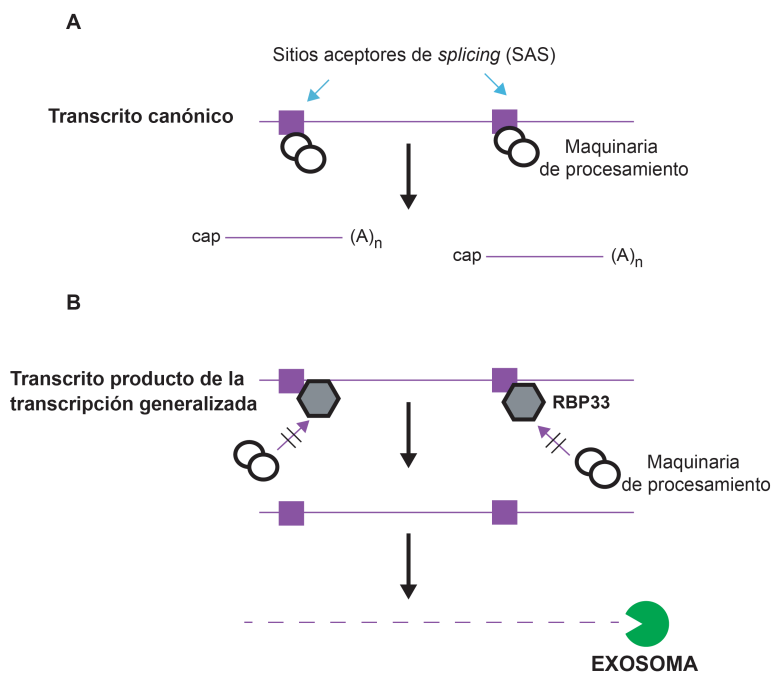


Figura 28. Modelo propuesto como mecanismo de acción de RBP33. I, Condiciones fisiológicas. (A) Los transcritos policistrónicos generados a partir de regiones codificantes se procesan correctamente generando mRNAs maduros que escapan de la degradación y se traducen. (B) En condiciones fisiológicas, RBP33 impide el procesamiento del transcrito aberrante generado a partir de la transcripción generalizada facilitando su degradación vía exosoma. Los sitios aceptores de *splicing* (SAS) se indican con flechas azules.

En vista a los resultados presentados en esta tesis y a otros datos publicados por nuestro grupo y por otros laboratorios, proponemos que RBP33 podría inhibir el *trans-splicing* de estos transcritos, de modo que las moléculas de RNA sin procesar resultantes serían degradadas rápidamente por el exosoma (Figura 28); por el contrario, en ausencia de RBP33, los transcritos no productivos se procesarían correctamente y estarían menos expuestos a la maquinaria de degradación (Figura 29).

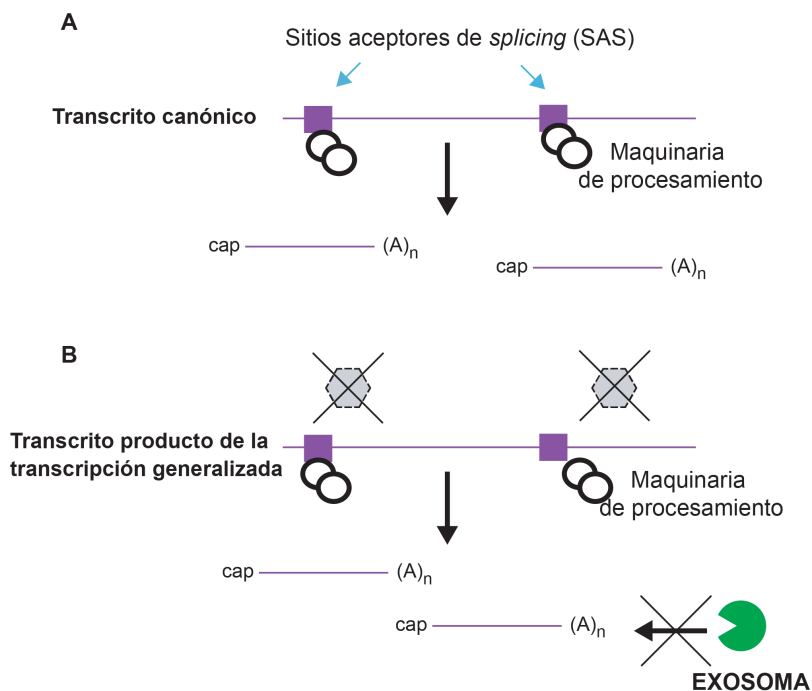


Figura 29. Modelo propuesto como mecanismo de acción de RBP33. II, En ausencia de RBP33. (A) En ausencia de RBP33, los transcritos canónicos continúan procesándose correctamente evitando la degradación. (B) Los transcritos no productivos se procesan correctamente en ausencia de RBP33 y por tanto no se degradan vía exosoma. Los sitios aceptores de *splicing* (SAS) se indican con flechas azules.

DISCUSIÓN

De acuerdo con esto, mostramos que RBP33 interactúa con varios factores de *splicing*, SF1 y U2AF2. Esto también está respaldado por datos previos de nuestro laboratorio que sugieren que el silenciamiento de RBP33 conduce a una mayor maduración de los transcritos derivados de retrotransposones (Fernández-Moya et al., 2014) (Figura 10 en introducción), así como por el hecho de que el exosoma sea responsable de la degradación de mRNAs mal procesados en *T. brucei* (Kramer et al., 2016). En cualquier caso, con nuestros datos no podemos saber si las interacciones observadas entre RBP33 y los factores de *splicing* son directas o mediadas por RNA.

Los niveles de proteína RBP33 deben de ser cruciales para mantener a raya este tipo de transcritos. En un trabajo previo de nuestro laboratorio realizado por Sandra Fernández Moya, en su tesis doctoral titulada 'Caracterización funcional de las proteínas de unión a RNA RBP33 y DRBD3 como reguladoras de la expresión génica de *Trypanosoma brucei*', 2013 (resultado publicado en Gómez-Liñán et al., 2022), se transfectó la línea celular TAP-RBP33 con pGR255. Este vector RNAi, que expresa RNA bicatenario correspondiente a la secuencia 5'UTR de RBP33, permite el silenciamiento (inducible por tetraciclina) de la copia endógena de la proteína. Mediante este sistema la célula solo expresa la copia etiquetada, debido a que contiene una 5'UTR heteróloga no susceptible al RNAi. En presencia de tetraciclina se observa una sobreexpresión de la copia etiquetada, lo que sugiere la necesidad por parte de la célula de compensar los niveles de RBP33 y mantener una determinada cantidad de proteína. Bajo esta hipótesis, unos niveles exagerados de RBP33 tampoco serían tolerados por la célula. De hecho, la sobreexpresión de RBP33 conduce tanto a una parada en el ciclo celular como a una disminución de los niveles celulares totales de mRNA (Cirovic et al., 2017), lo que probablemente sea el resultado de la unión de RBP33 a transcritos a los que normalmente no se une cuando la proteína se expresa a niveles fisiológicos, guiándolos para su degradación (Figura 30). No obstante, la sobreexpresión de la proteína no es letal para el parásito, ya que el fenotipo se revierte a las 36

horas tras la inducción. Los autores especulan que dado que RBP33 se une a su propio mRNA (Fernández-Moya et al., 2014), la proteína podría desestabilizar su propio mensajero provocando niveles de expresión normales pasados unos días (Cirovic et al., 2017).

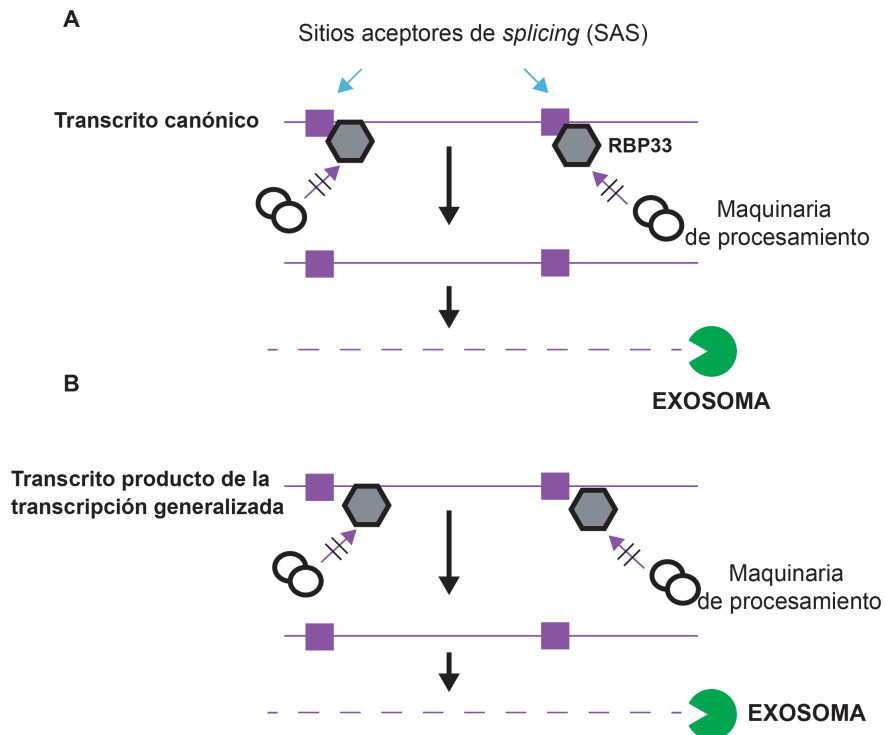


Figura 30. Modelo propuesto como mecanismo de acción de RBP33. III, Cuando se sobreexpresa RBP33. (A) Cuando se incrementan los niveles de RBP33, la proteína es capaz de unirse a transcritos canónicos, marcándolos para su degradación. (B) Los transcritos aberrantes continúan degradándose en presencia de RBP33 a través del exosoma. Los sitios aceptores de *splicing* (SAS) se indican con flechas azules.

En otros eucariotas existe una interacción directa entre las proteínas que dirigen los transcritos no productivos y el complejo exosoma, por ejemplo en el caso del complejo NNS en levaduras y de *Integrator* en mamíferos. Sin embargo, no hemos podido detectar una asociación entre RBP33 y el exosoma. La proteína RBP33 se caracteriza porque sus dos últimos tercios están intrínsecamente desordenados. Se

DISCUSIÓN

sabe que las regiones desordenadas pueden adoptar conformaciones inestables otorgándole a la proteína una gran variedad de posibilidades estructurales, así como la capacidad de unión a otras proteínas y/o a RNA (Balcerak et al., 2019). Esto podría indicar que la asociación de RBP33 con proteínas y/o moléculas de RNA probablemente sea muy dinámica y transitoria. La estructura proteica de RBP33 consta de un motivo RRM de unión a RNA y un motivo SLiM embebido dentro una región desordenada. Se cree que los motivos SLiMs podrían ser el nexo de unión entre la RBP y su efector (He et al., 2022). Los SLiMs tienen una afinidad relativamente baja por las moléculas con las que interactúan, lo que hace que estas interacciones sean transitorias y reversibles y, por lo tanto, ideales para mediar procesos dinámicos en la célula. Además, el motivo puede sufrir modificaciones postraduccionales que alteren sus propiedades fisicoquímicas y estructurales, lo que les confiere una gran capacidad de modulación (Davey et al., 2012). Por otra parte, debido a que los SLiMs son cortos y albergan una pequeña cantidad de residuos esenciales, además de localizarse en regiones desordenadas, pueden evolucionar rápidamente *de novo* a partir de secuencias peptídicas aleatorias y reconfigurar las redes de interacción de proteínas (Buljan et al., 2012; Davey et al., 2015; van Roey et al., 2014).

En conjunto, los resultados de la presente tesis doctoral muestran que la presencia de marcas de histonas específicas en los TTS actuando como “obstáculos físicos”, aunque es importante, no es suficiente para evitar la transcripción generalizada y el silenciamiento subtelomérico. Proponemos que en *T. brucei* existe un mecanismo de vigilancia donde participan RBP33 y el complejo exosoma para degradar los transcritos no productivos, aunque el mecanismo por el cual RBP33 guía a los RNAs para su destrucción queda aún por determinar. Todos los organismos generan continuamente moléculas de RNA antisentido potencialmente tóxicas y poseen proteínas especializadas que los detectan, marcan y destruyen. Nuestro trabajo demuestra que los tripanosomas no son una excepción, y a la vez indica que estos parásitos detectan dichas especies

de RNA de manera distinta al resto de organismos. Nuestros resultados, por tanto, poseen un interés adicional para el estudio de la evolución del metabolismo del RNA en eucariotas y pueden ser también útiles para entender mejor estos procesos en organismos más complejos.

CONCLUSIONES

CONCLUSIONES

1. La proteína RBP33 es una proteína específica de tripanosomátidos que posee un dominio SLiM embebido dentro de una región intrínsecamente desordenada, que abarca los dos últimos tercios de la proteína.
2. El silenciamiento de RBP33 por ribointerferencia conduce a una fuerte acumulación de transcritos no canónicos y principalmente antisentido, procedentes de zonas normalmente silenciadas tales como regiones de cambios de cadena convergentes y subtelómeros.
3. Los sitios de expresión de VSG de formas metacíclicas, pero no los sitios de expresión de VSG de formas sanguíneas, se sobreexpresan en ausencia de RBP33, lo que sugiere que existen distintos mecanismos para silenciar las VSGs en tripanosomas sanguíneos y metacíclicos.
4. La correcta expresión de un transgén *eYFP* insertado en zonas silenciadas en ausencia de RBP33 indica que los transcritos regulados por RBP33 podrían traducirse.
5. El aumento de transcritos no canónicos que se observa tras la pérdida de RBP33 no parece deberse a cambios en la accesibilidad de la cromatina.
6. Los transcritos derivados de tres sitios de terminación de la transcripción son significativamente más estables en tripanosomas carentes de RBP33 y son degradados por el complejo exosoma.
7. RBP33 se asocia con otras proteínas de unión a RNA. Algunas de las proteínas identificadas son factores de *splicing* conocidos, por lo que RBP33 podría jugar un papel a nivel de procesamiento de RNA, probablemente inhibiendo el *trans-splicing* de transcritos no canónicos.

CONCLUSIONES

8. Proponemos que la presencia de marcas de histonas específicas en los sitios de terminación, aunque es importante, no es suficiente para evitar la transcripción generalizada y el silenciamiento subtelomérico. Los tripanosomas poseen un mecanismo de vigilancia donde participan RBP33 y el complejo exosoma para amortiguar la transcripción no productiva.

ANEXOS

ANEXO I: FIGURAS SUPLEMENTARIAS

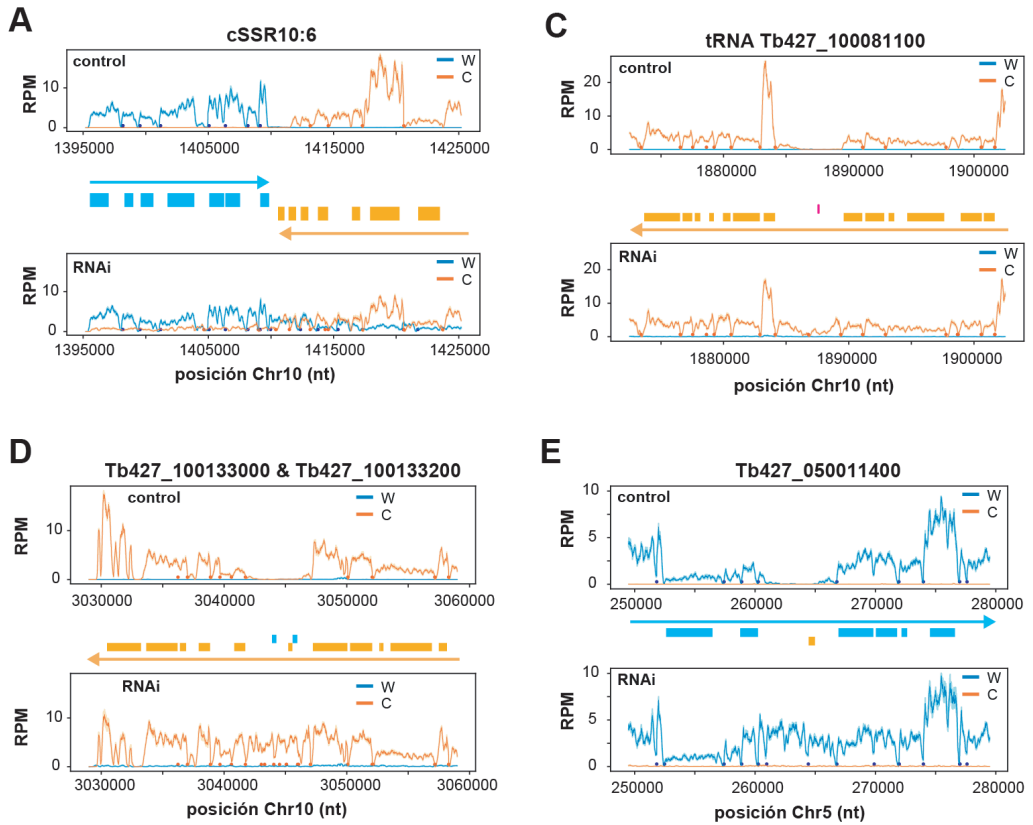


Figura S1. Material suplementario correspondiente a las Figuras 16 y 17. Gráficos de cobertura de lecturas de (A) cSSR10:6; (B) gen de tRNA no agrupado Tb427_100081100; (D y E) genes situados dentro de una PTU en orientación inversa.

ANEXOS

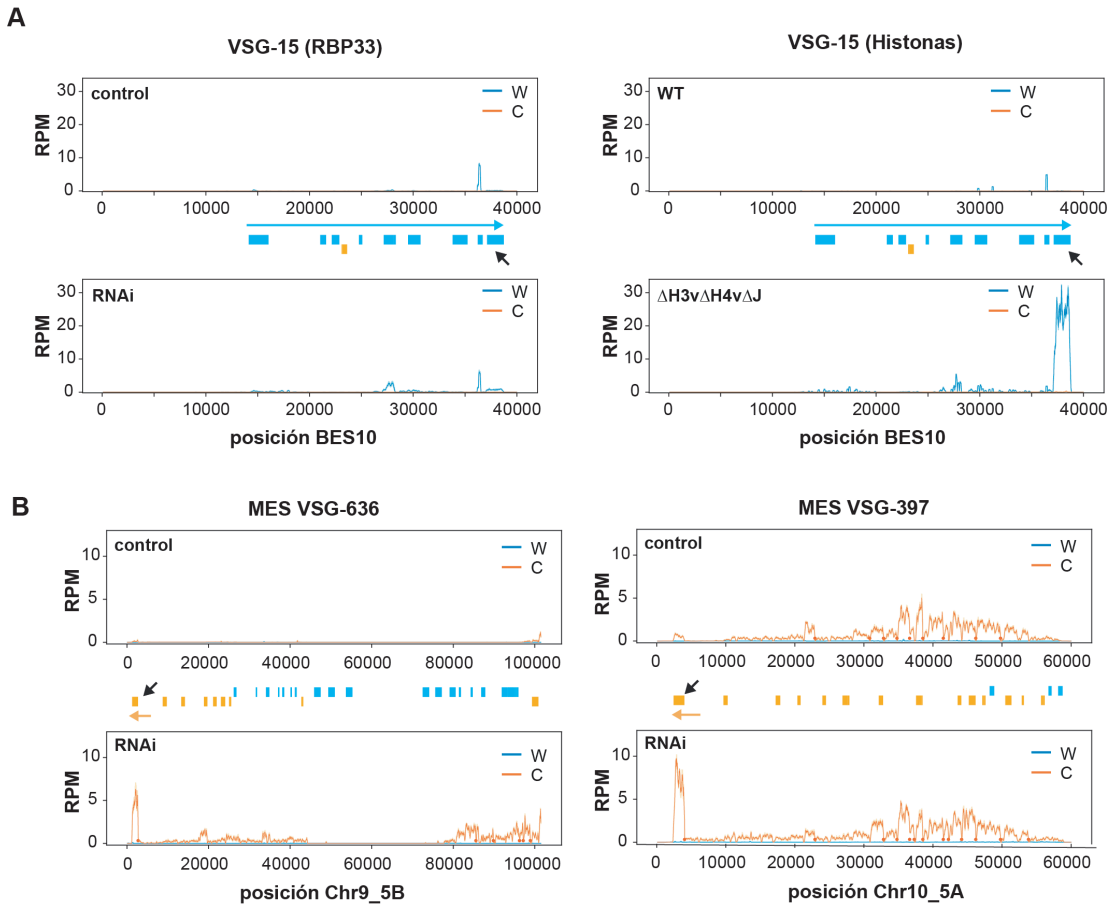


Figura S2. Los transcritos de MES VSG pero no los de BES VSG ven aumentada su expresión en ausencia de RBP33 en tripanosomas sanguíneos. (A) Gráficos de cobertura para el locus BES VSG-15 del control frente al RNAi de RBP33 (izquierda) o del control frente a al mutante triple KO H3vΔH4vΔJ (derecha). El gen VSG-15 se indica con una flecha. (B) Gráficos de cobertura de dos loci MES VSG. Los genes MES-VSG se indican con flechas.

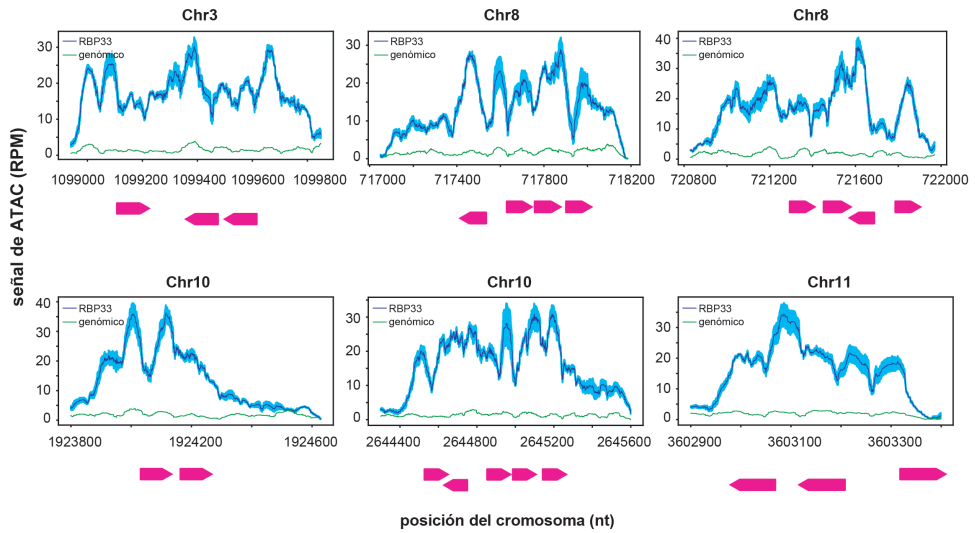


Figura S3. Gráficos de cobertura para la señal de ATAC en los genes de tRNAs agrupados (rectángulos rosas) en tripanosomas sanguíneos control. Los valores se representan como la media (línea azul) \pm sem (área sombreada) de tres réplicas de ATAC-seq. También se muestra la señal de ATAC correspondiente a la muestra control de DNA genómico (línea verde).

ANEXO II: PUBLICACIÓN CIENTÍFICA

Published online 1 December 2022

Nucleic Acids Research, 2022, Vol. 50, No. 21 12251–12265
<https://doi.org/10.1093/nar/gkac1123>

The RNA-binding protein RBP33 dampens non-productive transcription in trypanosomes

Claudia Gómez-Liñán, Elena Gómez-Díaz[✉], Gloria Ceballos-Pérez, Sandra M. Fernández-Moya and Antonio M. Estévez^{✉*}

Instituto de Parasitología y Biomedicina 'López-Neyra' (IPBLN), CSIC, Parque Tecnológico de Ciencias de la Salud, Avda. del Conocimiento 17, 18016, Armilla, Granada, Spain

Received June 02, 2022; Revised November 01, 2022; Editorial Decision November 02, 2022; Accepted November 09, 2022

ABSTRACT

In-depth analysis of the transcriptomes of several model organisms has revealed that genomes are pervasively transcribed, giving rise to an abundance of non-canonical and mainly antisense RNA polymerase II-derived transcripts that are produced from almost any genomic context. Pervasive RNAs are degraded by surveillance mechanisms, but the repertoire of proteins that control the fate of these non-productive transcripts is still incomplete. Trypanosomes are single-celled eukaryotes that show constitutive RNA polymerase II transcription and in which initiation and termination of transcription occur at a limited number of sites per chromosome. It is not known whether pervasive transcription exists in organisms with unregulated RNA polymerase II activity, and which factors could be involved in the process. We show here that depletion of RBP33 results in overexpression of ~40% of all annotated genes in the genome, with a marked accumulation of sense and antisense transcripts derived from silenced regions. RBP33 loss does not result in a significant increase in chromatin accessibility. Finally, we have found that transcripts that increase in abundance upon RBP33 knockdown are significantly more stable in RBP33-depleted trypanosomes, and that the exosome complex is responsible for their degradation. Our results provide strong evidence that RBP33 dampens non-productive transcription in trypanosomes.

INTRODUCTION

Recent advances in transcriptome sequencing technologies have uncovered a plethora of non-canonical, non-coding and mainly antisense transcripts that are gener-

ated by pervasive (widespread) transcription from almost any genomic context in organisms as diverse as bacteria, yeast and mammals (1,2). Pervasive transcription seems to originate mainly from RNA polymerase II (RNA pol II) transcription termination readthrough, and it is coupled to RNA degradation by the nuclear RNA exosome. In yeast, this is carried out by the Nrd1–Nab3–Sen1 complex, which binds to specific sequences within target transcripts and recruits the nuclear exosome for rapid degradation (3). Chromatin remodeling elements which alter chromatin structure at termination sites are also important to suppress pervasive transcription and subtelomeric silencing in yeast (4). Another important factor is the RNA-binding protein Npl3, whose absence results in widespread transcription readthrough with pervasive effects on gene expression (5). In mammals and *Drosophila*, the Integrator protein complex has been shown to be essential to attenuate non-productive transcription. This complex cleaves transcripts co-transcriptionally in a sequence-independent manner, and pervasive transcripts are then rapidly degraded in the nucleus by the exosome which associates with Integrator (6,7).

RNA pol II transcription is apparently unregulated in trypanosomatid protozoa, and initiation and termination events occur only at a limited number of sites per chromosome. In these early-branched eukaryotes, protein-coding genes are arranged into polycistronic transcription units (PTUs) that are constitutively transcribed by RNA pol II (8). Mature mRNAs are generated by *trans*-splicing of a capped spliced leader sequence at the 5' end, and coupled polyadenylation at the 3' end. There are no specific polyadenylation signal motifs; instead, polyadenylation sites are chosen depending on the distance to *trans*-splicing signals present in the cistron immediately downstream (9). Thus, while RNA pol II termination is linked to mRNA polyadenylation in other eukaryotes, this cannot be the case in an organism with polycistronic transcription (10). As mentioned above, pervasive transcription is usually associated with transcription termination, but it is not

*To whom correspondence should be addressed. Tel: +34 958 181652; Fax: +34 958 181632; Email: aestevez@ipb.csic.es

Present address: Sandra M. Fernández-Moya, Gene Regulation of Cell Identity, Regenerative Medicine Program, Bellvitge Institute for Biomedical Research (IDIBELL) and Program for Advancing Clinical Translation of Regenerative Medicine of Catalonia, P-CMR[C], L'Hospitalet del Llobregat, 08908 Barcelona, Spain.

© The Author(s) 2022. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

known whether this process is operative in organisms that regulate gene expression in the absence of RNA pol II transcriptional control.

Trypanosoma brucei is the organism of choice for functional analyses in trypanosomatids. It is responsible for human and animal trypanosomiasis in sub-Saharan Africa. Trypanosomes are transmitted between mammals by tsetse flies, and undergo a profound differentiation process in order to adapt to the different environments they face in the life cycle. Trypanosome cells proliferate as procyclic forms in the insect's midgut, and migrate to the salivary glands to differentiate into quiescent metacyclic forms. Metacyclic trypanosomes are transmitted through tsetse bites to the mammalian bloodstream, where they differentiate first into proliferative bloodstream forms and then into non-dividing stumpy forms that are transferred to the fly during a blood meal and differentiate to procyclic forms, closing the cycle (11). To achieve a successful and persistent infection, bloodstream forms express a dense surface coat consisting of a variant surface glycoprotein (VSG), the basis of antigenic variation that allows the infection to evade the host's immune response (12). Expressed VSG genes are found near telomeres, and a tightly regulated mechanism of monoallelic expression ensures that only one VSG gene is transcribed at a time by RNA pol I (13,14). Metacyclic forms are also covered with a VSG coat that is transcribed from one of several specific metacyclic expression sites (MESs) to pre-adapt parasites for infection of the mammalian host (15). Metacyclic trypanosomes display different VSG variants; however, each individual cell expresses just one specific MES VSG (16). With the exception of the active VSG gene, subtelomeres are mostly silenced in *T. brucei* (17).

Little is known about RNA pol II transcription initiation and termination in trypanosomatids. PTUs can be divergent or convergent, and are separated by strand switch regions (SSRs). Run-on studies have shown that divergent SSRs (dSSRs) associate with RNA pol II transcription start sites (TSSs), whereas transcription termination sites (TTSs) locate at convergent SSRs (cSSRs) or in the vicinity of genes transcribed by different RNA polymerases (8,10,18,19). TSSs are enriched in the essential epigenetic marks H3K4me3, H4K10ac, H2Az, H2Bv and bromodomain factor 3, have an open chromatin structure and contain sequence-specific promoters able to initiate gene transcription and cause histone variant deposition (18,20–23). In contrast to TSSs, TTSs are enriched in three non-essential chromatin marks: base J (β -D-glucosyl-hydroxymethyluracil, a trypanosomatid-specific DNA modification), H3v and H4v (18,20,24). All three marks have been shown to be important for transcription termination, although H4v seems to be the major epigenetic sign for termination at TTSs in *T. brucei* (24–27). H4v null trypanosomes are viable and show increased levels of antisense transcripts derived from TTSs (27). H3v and base J, but not H4v, are enriched at telomeric sites (18,20). VSG monoallelic expression is disrupted in H3v null mutants, giving rise to the accumulation of many BES and MES VSG transcripts (17,27).

The fact that readthrough antisense transcripts are detected in the absence of TTS-specific epigenetic marks suggests that pervasive transcription also exists in try-

panosomes, but factors that control the fate of these non-productive transcripts are not known at present. A promising candidate is the RNA-binding protein RBP33 (28,29). This protein contains an RNA recognition motif (RRM) close to its N-terminus, and the carboxyl two-thirds have no apparent similarity to any known protein outside trypanosomatids. It is exclusively localized in ~60 foci in the nucleus outside the nucleolus (29), and it is expressed at similar levels in bloodstream and procyclic *T. brucei* cells (28). RBP33 depletion leads to rapid cell death in both developmental forms (28), whereas its overexpression results in quick growth arrest in bloodstream trypanosomes (29). RBP33 binds preferentially to RNAs originating from cSSRs and other silent genomic regions in *T. brucei* (28), which hints at a possible role for this protein in the regulation of the abundance of this type of transcripts. In this work, we have analyzed the transcriptomes of RBP33-depleted bloodstream and procyclic *T. brucei* cells, and observed that RBP33 loss leads to a strong accumulation of transcripts derived from TTS, subtelomeric genes and MES VSGs. We show using ATAC-seq that RBP33 depletion does not result in an apparent increase in chromatin accessibility at TTSs or MES VSG loci. We also provide evidence that TTS-derived transcripts are significantly more stable in RBP33-depleted trypanosomes and that they are degraded by the exosome complex.

MATERIALS AND METHODS

Trypanosome culture and RNA interference (RNAi)

Trypanosoma brucei 449 procyclic cells (30) were cultured at 28°C in SDM-79 medium (31) supplemented with 10% fetal bovine serum (FBS). 'Single marker' *T. brucei* Lister 427 bloodstream cell line S16 (32) was maintained in HMI-9 medium (33) containing 10% FBS at 37°C with 5% CO₂. Trypanosomes were transfected following standard procedures (34). RBP33 expression was silenced by RNAi using bloodstream and procyclic cell lines transfected with plasmid pGR70, which expresses double-stranded RNA (dsRNA) corresponding to the carboxyl-half of RBP33 in a tetracycline-inducible manner (28). Likewise, depletion of exosome subunits RRP44 and RRP6 in bloodstream and procyclic cells was achieved using the dsRNA-expressing plasmids pHD1157 and pHD1165, respectively (35). In all RNAi experiments, cells were incubated for 48 h in the presence of 1 μ g/ml tetracycline, and successful depletion of the proteins was confirmed by western blot assays using specific antisera (28,35,36).

RNA-seq

For transcriptome analyses of bloodstream or procyclic RBP33-depleted trypanosomes, total RNA was obtained from either untransfected (parental S16 or 449 cells) or RNAi-induced cells using the RNeasy Mini Kit (Qiagen). RNA libraries were prepared from poly(A)-selected transcripts using the standard TruSeq stranded mRNA sample preparation protocol (Illumina). Biological triplicates were sequenced at the Genomics Unit of the IPBLN-CSIC (Granada, Spain) using a NextSeq 500 platform (Illumina). The resulting 76 nt paired-end sequences were checked

for quality using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Libraries were subsampled using seqtk (version 1.2, <https://github.com/lh3/seqtk>) so the differences among sample sizes were <10%, and then mapped to the *T. brucei* Lister 427 2018 (v4.9) genome using the ‘align’ program of the Subread package [Rsubread version 1.34.7 (37)]; only uniquely mapped reads were considered (option ‘unique’ set to TRUE). Reads assignment to genes was done using the ‘featureCounts’ program of the Subread package [version 1.5.0-p (38)] with options -p -B -C; the annotation file was TriTrypDB-49.TbruceiLister427.2018.gff (downloaded from <https://tritrypdb.org/tritrypdb/app/downloads/>), which was supplemented with 126 manually annotated SSRs assigned arbitrarily to the Watson (+) strand (Supplementary Table S1). Forward read mapping to forward PTUs, or reverse reads mapping to reverse PTUs, were considered to represent antisense transcription. Antisense and sense count profiles were generated independently (option -s set to 1 or 2 in featureCounts, respectively), and then combined and analyzed for differential expression using edgeR [version 3.36.0 (39)]. Only genes containing more than one count per million mapped reads (CPM) in at least three samples were considered. For Pearson correlation and principal component analysis (PCA), CPM were normalized using the TMM method (39) and transformed to log₂ in edgeR. PCA plots were generated using the sklearn.decomposition module [version 0.24.2 (40)]. The breadth of coverage was calculated from sequencing depth values obtained using the ‘depth’ module of samtools [version 1.9 (41)] with options -d 0 -aa. Positional enrichment of RBP33-regulated transcripts was analyzed using right-tailed Fisher’s exact tests; *P*-values were corrected for false discovery rate (FDR) according to the Benjamini–Hochberg method (42). For coverage plots, regions of interest were binned and counted (sliding window, 100 bp; step size, 10 bp) using the ‘countReadsPerBin’ module of the deepTools API [version 3.5.0 (43)], and corrected for library size. To map splice acceptor sites (SASs), forward reads containing at least the last 14 nt of the spliced leader (SL) sequence were selected (44). The SL sequence was removed, reads were aligned to the *T. brucei* 427 genome as described above and SASs were assigned to the first position of all uniquely mapped SL-containing reads. Only those SASs supported by a minimum of five reads in at least two replicates were considered for further analysis.

External data corresponding to a conditional triple knockout cell line lacking chromatin marks H3v, H4v and base J (27), as well as data from the parental cell line control, were downloaded from the European Nucleotide Archive (project PRJNA727846; samples SRR9088967, SRR9088968, SRR9088969, SRR9088997, SRR9088998 and SRR9088999) and subjected to the same pipeline described above.

Quantitative RT-PCR

We first converted 250 ng of RNA to cDNA using 0.5 µg of hexamers (Invitrogen), 0.25 mM of each dNTP, 20 U of RiboLock (Thermo Scientific) and 200 U of Maxima reverse transcriptase (Thermo Scientific) in a final volume of

20 µl. Reactions were allowed to proceed for 30 min at 50°C followed by incubation at 85°C for 5 min. Quantitative RT-PCRs were carried out in 96-well plates (Thermo Scientific) in a BioRad CFX96 thermal cycler using the following thermocycler steps: 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 s alternating with 55°C for 1 min. Melting curve analyses were performed to confirm a single amplicon for each mRNA tested. Reactions were set up in a final volume of 10 µl containing 0.5 µl of cDNA, 1× SYBR Green master mix (Thermo Scientific) and 0.5 µM of each oligodeoxynucleotide. Fold changes in expression were calculated using the 2^{-ΔΔCT} method (45) with *actin* or *ZFP2* mRNAs as references. All quantitative RT-PCR experiments were performed with at least three biological replicates. Oligodeoxynucleotide pairs used are listed in Supplementary Table S2.

Flow cytometry

For enhanced yellow fluorescent protein (eYFP) expression, the eYFP open reading frame (ORF) from p2675 (46) was PCR-amplified and cloned in reverse orientation in the plasmid pGR12 (47) to yield pGR417, where the eYFP ORF is flanked by the *EPI* 5'-untranslated region (5'-UTR) and the *actin* 3'-UTR. The rDNA spacer targeting sequence in pGR417 was replaced with DNA fragments corresponding to genes Tb427.110143800 or Tb427.100047800 to yield pGR431 and pGR433, respectively. Plasmids were linearized with EcoRI (pGR431) or MfeI (pGR433) and transfected to bloodstream S16 cells. Stable cell lines were selected in the presence of 5 µg/ml blasticidin, and proper integration of plasmids was checked by PCR of genomic DNA. Cell lines harboring eYFP transgenes were further transfected with pGR70 for tetracycline-inducible RNAi of RBP33. Flow cytometry was carried out with 1 × 10⁶ trypanosomes from uninduced or RNAi-induced cultures. Cells were centrifuged and resuspended in phosphate-buffered saline (PBS). eYFP fluorescence was measured in a FaCSCalibur cytometer (Becton Dickinson) equipped with a 488 nm laser, and 20 000 events were recorded per sample. Data were acquired with CellQuest software and analyzed with FlowJo (version 10) software.

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

We adapted the method described in (17), as follows: 1 × 10⁷ untransfected (S16) or RBP33-depleted bloodstream trypanosomes were washed in cold PBS and resuspended in 100 µl of permeabilization buffer (10 mM Tris, pH 8.0, 100 mM KCl, 25 mM EDTA and cComplete mini EDTA-free protease inhibitors). Digitonin (Millipore 300410) was added at a final concentration of 40 µM (from a 4 mM stock freshly prepared in permeabilization buffer), and the suspension was incubated for 5 min at room temperature. Cells were then centrifuged at 4°C for 5 min at 1400 g, washed in 600 µl of cold isotonic buffer [10 mM Tris pH 8.0, 100 mM KCl, 5% glycerol, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF)], resuspended in 50 µl of transposition mix (25 µl of 2× TD buffer, 22.5 µl of nuclease-free water, 2.5 µl of Tn5 transposase, Illumina

Nextera kit FC-121-1030) and incubated for 30 min at 37°C. As a control, a sample containing 200 ng of genomic DNA was incubated in transposition mix for 5 min at 55°C in the same conditions, except that 1.25 µl of transposase were used. Reactions were purified using the QuiaGen MiniElute Kit and eluted in 11 µl of 10 mM Tris pH 8.0. Library amplification was carried out in 2× KAPA HiFi mix and 1.25 µM of Nextera primers during seven cycles; this cycle number was determined by quantitative PCR using the conditions described in (48). Fragment selection was performed using the 'NucleoMag kit for clean up and size selection' (Macherey-Nagel 744970) following a double size selection scheme (first round, beads volume = 0.6× sample volume; second round, beads volume = 0.7× sample volume; final elution volume = 17 µl). Size selection for the genomic sample was done in a single step using a volume of beads equal to that of the sample. Biological triplicates were sequenced as described above for RNA-seq. Reads were aligned to the *T. brucei* Lister 427 2018 (v4.9) genome using bowtie 2.0 (49) with the options `-no-unal -no-mixed -5 5 -3 5`. Reads corresponding to nucleosome-free regions were selected using the 'alignmentSieve' module of deeptools (43) with the options `-minMappingQuality 30 -maxFragmentLength 100 -ATACshift -ignoreDuplicates`; the `-bl` option was also included to remove reads corresponding to the maxicircle (mitochondrial) DNA (unitig_851_maxicircle_Tb427v10). For Pearson correlation and PCA, genome-wide (bin size = 100 bp) or gene-wise read coverages were obtained using the 'multiBamSummary' module of the deeptools API (43), normalized as for RNA-seq to yield ATAC signals, and transformed to log₂; for PCA, the ATAC signal was further normalized to that of the genomic DNA control. Single-base resolution plots were generated using the 'countReadsPerBin' module of the deeptools API (43) and normalized for library size. Peak calling on nucleosome-free reads was performed by MACS2 [version 2.2.7.1 (50)] using the module 'callpeak' with the following parameters: `-g 3.5e7 -nomodel -f BAMPE -B -keep-dup all`. Transposed genomic DNA was used as a control for the peak calling. Signal tracks were generated using the module 'bdgcmp' of MACS2 with the option `-m ppois`, and were then converted to bigWig files using the bedGraphToBigWig module of kenUtils (<https://github.com/ENCODE-DCC/kenUtils>).

External data corresponding to bloodstream trypanosomes lacking chromatin marks H3v and H4v (17), as well as data from the parental cell line and transposed genomic DNA samples, were downloaded from the European Nucleotide Archive (project PRJNA487259; samples SRR7739654, SRR7739658, SRR7739651, SRR7739655, SRR7739659 and SRR7739660) and analyzed according to the same pipeline as described above.

RNA half-life assays

Untransfected (449 cell line) or RBP33-depleted procyclic trypanosomes were incubated with 10 µg/ml actinomycin D (from a stock of 5 mg/ml in dimethylsulfoxide) and 2 µg/ml sinefungin (from a stock of 2 mg/ml in water). Sinefungin was added 5 min prior to actinomycin D. Total RNA was obtained from samples taken at different times up to 1 h using the NucleoSpin RNA kit

(Macherey-Nagel). Transcript abundance was measured by quantitative RT-PCR and normalized to that of *ZFP2* (Tb427.110167900/Tb927.11.14950) mRNA, which is stable over the time course of the experiments (51). The rate constant for decay, *k*, was obtained after fitting data to first-order kinetics using the equation $[RNA]_t = [RNA]_0 \times e^{-kt}$; half-life values were then calculated as $\ln(2)/k$ (52).

Tandem affinity purification (TAP)

A procyclic cell line expressing TAP-RBP33 from the endogenous locus was generated using the strategy described in (53). Protein complexes were purified from $1-2 \times 10^{10}$ procyclic trypanosomes using the TAP method (54) with the modifications described in (53). TAP-purified material was subjected to liquid chromatography with tandem mass spectrometry (LC-MS/MS) in EASY-nLC equipment (Proxeon) coupled to an amaZon speed ion ETD trap (Bruker). Proteins were identified using ProteinScape (Bruker) and MASCOT (Matrix Science) software. We considered only those proteins with a score >60 and with two or more detected peptides.

RESULTS

RBP33 depletion leads to a marked increase in the level of antisense transcripts

RBP33 is a nuclear and essential RNA-binding protein (28). The carboxyl two-thirds of the protein are mostly intrinsically disordered, and show no apparent similarity to any known protein outside trypanosomatids [(28), Supplementary Figure S1]. An RBP33 ortholog was readily detected in *Bodo saltans*, a free-living outgroup species belonging to the Eubodonida order, which is closely related to parasitic trypanosomatids (Supplementary Figure S1), but not in other species from the orders Prokinetoplastida, Parabodonida or Neobodonida. Thus, RBP33 seems to be a trypanosomatid-specific protein with no homologs present even in the most closely related kinetoplastid species.

To gain insight into the function of RBP33, the transcriptomes of bloodstream and procyclic trypanosomes were analyzed upon depletion of the protein by RNAi. Stranded RNA-seq reads were aligned to the Lister 427 (version 2018) genome, and assignment to genes was done separately for sense and antisense reads. Pearson correlation coefficients were >0.980, indicating good reproducibility between replicates (Supplementary Figure S2A). We could detect a notable increase in the abundance of antisense transcripts in both developmental forms when RBP33 expression was silenced, which was readily visible as a 'hump' in scatter plots of reads per kilobase of transcript per million mapped reads (RPKM) values for all genes in control versus RNAi-induced cells (Figure 1A). Increased levels of antisense transcripts have also been reported in trypanosomes lacking base J and histone variant H3v (24), in a knockout mutant for histone variant H4v (27) and in a triple knockout (KO) $\Delta H3v\Delta H4v\Delta J$ cell line (27). We analyzed the RNA-seq dataset corresponding to the transcriptome of the $\Delta H3v\Delta H4v\Delta J$ triple KO mutant (27) to compare the degree of accumulation of antisense transcripts in cells lacking all three histone marks with that ob-

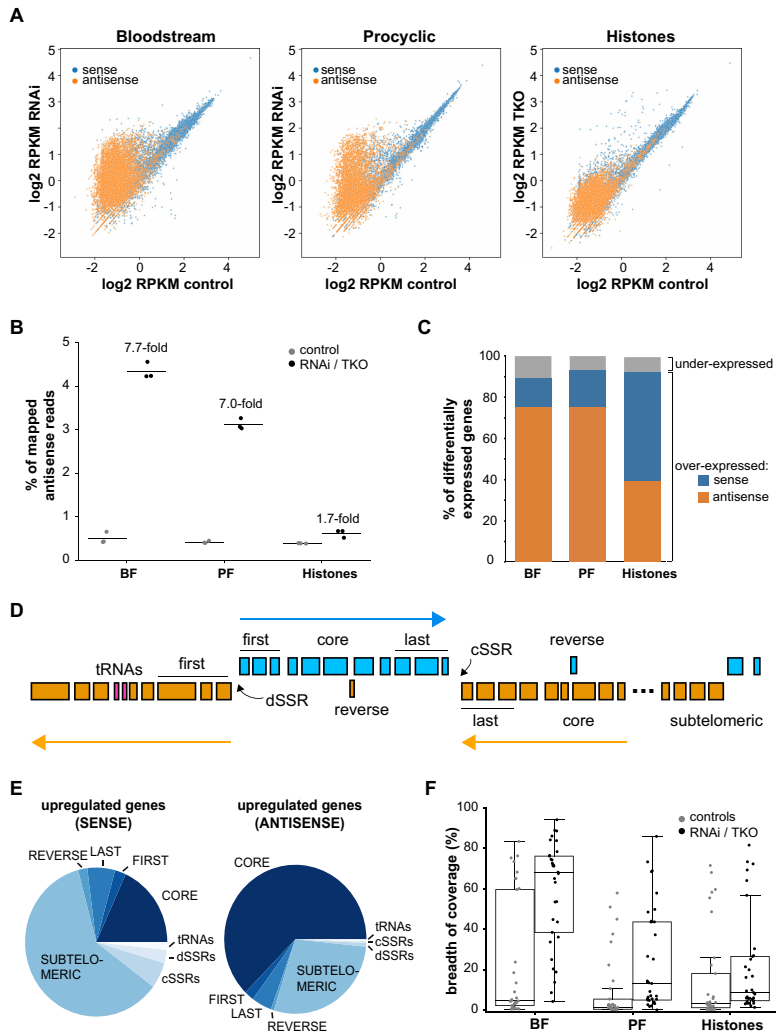


Figure 1. Effect of RBP33 depletion on the transcriptome of *T. brucei*. (A) Scatter plots comparing expression values for control versus RBP33-depleted cells or wild-type versus $\Delta H3v\Delta H4v\Delta J$ triple KO (TKO) mutants. Average read counts across replicates were normalized to RPKM and transformed to \log_2 . Each dot represents a transcript. (B) Quantification of antisense reads in RBP33 and histones datasets. The percentage of mapped antisense reads obtained after read assignment to genes relative to total mapped reads is represented as the mean (line) of three RNA-seq biological replicates (dots). BF, bloodstream forms; PF, procyclic forms. (C) Percentage of genes differentially expressed in sense or antisense fashion in RBP33 and histones datasets. (D) Schematic representation of a *T. brucei* chromosome indicating the different localizations used in this study. (E) Pie chart indicating the relative proportions of up-regulated genes within each localization category in bloodstream trypanosomes. (F) Breadth of coverage of subtelomeric regions in RBP33 and $H3v\Delta H4v\Delta J$ triple KO (Histones) datasets. Box plots represent the percentage of bases in subtelomeres sequenced at $1\times$ depth (bases covered by at least one uniquely mapped read in at least two replicates). Boxes indicate the interquartile range (IQR); whiskers, ± 1.5 IQR; waists, medians; dots, individual subtelomeres.

served in RBP33-depleted trypanosomes. As seen in Figure 1A, the increase in abundance of antisense transcripts was more pronounced in the RBP33 RNAi cell line compared with the triple KO mutant. Indeed, antisense mapped reads in RBP33-depleted trypanosomes were 7- to 8-fold higher than those detected in control cells, whereas an increase of just 1.8-fold was observed in the triple KO cell line [Figure 1B and (27)]. Moreover, a PCA of RBP33 and $\Delta H3v\Delta H4v\Delta J$ datasets showed that RBP33-depleted replicates cluster more distantly from their corresponding control replicates than triple KO samples do (Supplementary Figure S2B), thus suggesting a larger alteration in gene expression upon RBP33 knockdown.

Next, we performed differential expression analysis to investigate which genes are regulated by RBP33. Sense and antisense read count profiles were combined and analyzed using edgeR, and genes with \log_2FCI values >1.0 and FDRs <0.01 were considered for further study. We found that 40% (bloodstream forms) or 20% (procyclic forms) of the 17 243 genes annotated in the Lister 427 2018 genome were differentially expressed in RBP33-depleted trypanosomes (Supplementary Tables S3–S8). Among these, the vast majority ($>94\%$) had increased expression, mainly in an antisense fashion (Figure 1C). In $\Delta H3v\Delta H4v\Delta J$ trypanosomes, on the other hand, only 6% of genes were differentially expressed (Supplementary Table S9), 96% of which were overexpressed in roughly an equal sense and antisense manner (Figure 1C).

RBP33-regulated transcripts are derived mainly from silenced genomic regions and antisense transcription of core genes

Chromosomes in *T. brucei* partition into transcribed ‘cores’ and subtelomeric regions, which are thought to be non-expressed (17). A simple glance at the list of differentially expressed genes in RBP33-depleted bloodstream and procyclic trypanosomes revealed a notable abundance of transcripts originating from genomic loci from which no transcripts are detected under physiological conditions, such as SSRs or subtelomeres, and also from antisense reads derived from housekeeping core genes (Supplementary Tables S3 and S6). To quantify this more precisely, we classified *T. brucei* genes into the following categories according to their localization in the genome: ‘core’, housekeeping protein-coding genes within core regions; ‘subtelomeric’, non-expressed genes located at subtelomeric regions; ‘tRNAs’, tRNA genes within core regions; ‘cSSR’ and ‘dSSR’, convergent and divergent SSRs within core regions; ‘first’, first three genes within a PTU (flanking dSSRs); ‘last’, last three genes within a PTU (flanking cSSRs); and ‘reverse’, non-expressed genes in reverse orientation within a PTU (Figure 1D; Supplementary Tables S3 and S6). We could observe that most RBP33-up-regulated RNAs derive from subtelomeres (sense) or core regions (antisense transcription), as shown in Figure 1E (bloodstream forms) and Supplementary Figure S3A (procyclic forms). We next evaluated the statistical enrichment of RBP33-regulated transcripts within each of the categories using a Fisher’s exact test, and observed a significant enrichment (FDR <0.05) of sense transcripts derived from subtelomeres, cSSRs, dSSRs,

‘last’, tRNAs and ‘reverse’ loci, and of antisense transcripts originating from ‘core’, ‘last’, ‘first’, ‘reverse’ and ‘cSSRs’ regions (Supplementary Figure S3B, C). The enrichment in cSSRs was particularly striking, as transcripts arising from 58 (bloodstream forms) or 56 (procyclic forms) of a total of 59 annotated cSSRs were significantly overexpressed in RBP33-depleted trypanosomes (Supplementary Tables S3 and S6).

As mentioned above, most subtelomeres are thought to be transcriptionally silent in physiological conditions, and subtelomeric genes were particularly over-represented in RBP33-depleted datasets, especially in bloodstream forms. To visualize the extent of this phenomenon, we calculated the breadth of coverage for each subtelomere as the percentage of bases that were sequenced at least once in at least two RNA-seq replicates (depth = $1\times$). Figure 1F shows breadth values for all 33 subtelomeres in control versus RBP33-depleted bloodstream and procyclic trypanosomes; the $\Delta H3v\Delta H4v\Delta J$ dataset was also included for comparison (raw data are provided in Supplementary Table S10). A pronounced increase in breadth of coverage was observed upon depletion of RBP33 in bloodstream forms: on average, only 6% of subtelomeric bases were sequenced at least once in control cells, whereas this value increased to 43% in RBP33-depleted bloodstream forms, yielding a 7-fold increase in breadth. Although not as marked, breadth values were also augmented in procyclic and histones datasets (4.5-fold and 2.5-fold, respectively). Distinct increases in breadth values were also observed at different sequencing depths (Supplementary Figure S3D).

The overexpression of transcripts derived from cSSRs and neighbor genes prompted us to study in more detail the distribution of mapped reads around cSSRs. We used sliding windows to analyze genomic regions centered at cSSRs and extended 15–20 kbp on each side. Figure 2A and B shows read coverage profiles corresponding to two representative cSSRs (additional examples are presented in Supplementary Figure S4A to S4C; see also below). Reads per million mapped reads (RPM) values were generated separately for Watson and Crick strands. In control samples, transcripts arising from cSSRs and adjacent regions were barely detected. However, depletion of RBP33 resulted in a clear increase of antisense transcripts that could be readily identified as orange peaks in the upstream (‘left’) PTU or blue peaks in the downstream (‘right’) PTU. This phenomenon was better visualized when plotting the average of RPM across 30 different cSSRs flanked by PTUs at least 50 kbp long (Figure 2C), where it can be observed that antisense transcripts are generated up to 40 kbp away from cSSRs. This phenomenon was also observed in $\Delta H3v\Delta H4v\Delta J$ trypanosomes (24,27), although the effect was much less pronounced (Supplementary Figure S4D).

We showed above that tRNAs and transcripts corresponding to genes located in reverse orientation within a PTU were significantly up-regulated upon depletion of RBP33. Most tRNA genes are located at cSSRs, an explanation for their overexpression in RBP33-depleted conditions (see for example Supplementary Figure S4B and C); some others residing within PTUs also gave rise to increased levels of transcripts upon RBP33 silencing (Supplementary Figure S4E). A similar phenomenon was observed

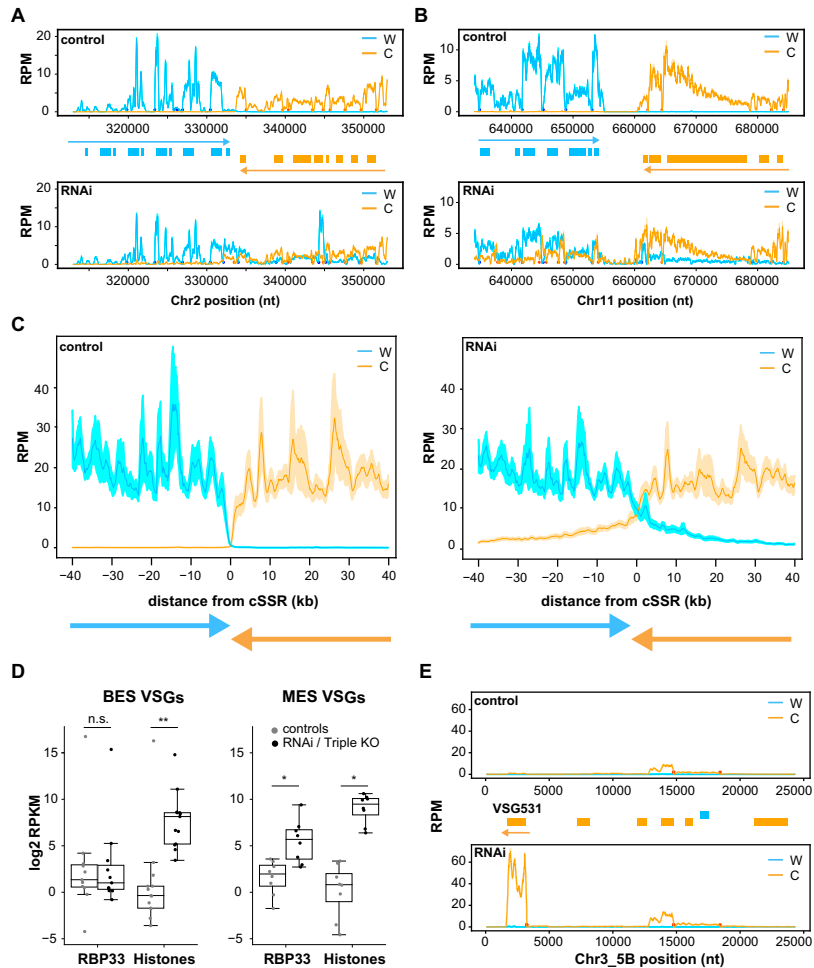


Figure 2. Coverage plots of representative genomic regions, and effect of RBP33 depletion on VSG transcripts levels. Read counts were obtained separately from Watson (W, blue lines) or Crick (C, orange lines) strands in bloodstream forms with sliding windows (100 bp bins, 10 bp step in A, B and E; 1000 bp bins and steps of 100 bp in C) and normalized by library size (RPM). Values from control were compared with those from RBP33-depleted samples (RNAi), and represented as the mean (solid lines) \pm SEM (shaded area) of the coverage across 30 cSSRs (C) or the mean \pm SEM of three RNA-seq biological replicates (A, B and E). Genes are represented as boxes (blue, protein-coding genes in the W strand; orange, protein-coding genes in the C strand), and colored arrows denote the direction of transcription in each PTU. Blue and orange circles indicate *trans*-splicing acceptor sites assigned to W and C strands, respectively. (A and B) Examples of convergent SSRs, cSSR2:2 and cSSR11:2 (see Supplementary Table S1 for SSR nomenclature details). (C) Mean coverage across 30 cSSRs. (D) Effect of RBP33 depletion and Δ H3v Δ H4v Δ J triple KO on the expression of metacyclic (MES) and bloodstream (BES) VSG genes. Read counts were normalized to CPM in edgeR, converted to RPKM and transformed to log₂. Average RPKM values across RNA-seq replicates were calculated for 17 BES VSG and 8 MES VSG transcripts and represented in box plots (boxes represent the IQR; whiskers, ± 1.5 IQR; waists, medians; dots, individual VSG transcripts). Only VSG transcripts showing values of RPKM ≥ 1 in at least three replicates were considered. Two-sided Mann-Whitney U-tests were used to assess whether there was significant differential expression between RBP33-depleted or triple KO and their respective control samples; n.s., not significant ($P > 0.5$); * $P < 0.005$; ** $P < 0.0005$ ($n = 11/13$ BES VSGs; $n = 8$ MES VSGs). (E) Coverage plot corresponding to the metacyclic VSG531 locus.

for several protein-coding genes embedded in reverse orientation within PTUs (Supplementary Figure S4F, G), and in boundaries between RNA Pol I and Pol II transcription units (*EPI* locus, Supplementary Figure S4H).

Only a small fraction of the differentially regulated transcript set (6% in bloodstream forms) decreased in abundance upon RBP33 depletion. Almost all (98%) were down-regulated in a sense fashion (Supplementary Tables S3 and S6). This could be the result of RNAi triggered by dsRNA formation due to increased levels of antisense transcripts. If that were the case, one would expect down-regulated genes to be located preferentially near cSSRs, where antisense transcripts are prominently produced. To test this, we analyzed the genomic localization of down-regulated genes relative to cSSRs, and observed that indeed most genes were found close to these regions (Supplementary Figure S4I). Interestingly, at least 141 down-regulated genes in bloodstream and 11 in procyclic forms are thought to be essential for trypanosome growth according to genome-wide RNAi screenings [55], Supplementary Tables S3 and S6].

RBP33-regulated transcripts are trans-spliced

To assess whether RBP33-regulated transcripts were properly processed, we made a global survey of *trans*-SASs in our RNA-seq dataset. We could identify 8553 SASs in control and 10 426 SASs in RBP33-depleted trypanosomes (Supplementary Tables S11 and S12). A canonical AG acceptor dinucleotide was found at 98% of SASs in both conditions, in agreement with previous reports [56,57]. Mapped SASs are indicated as blue (W strand) or orange (C strand) circles in Figure 2A, B and E, and in Supplementary Figure S4, where it is observed that many RBP33-specific peaks are supported by SASs. We detected 42 SASs in control and 707 SASs in RBP33-depleted cells for which the nearest downstream gene was located in the opposite strand, thus being indicative of potential antisense transcripts (Supplementary Tables S11 and S12). We could not address whether these transcripts were also polyadenylated due to the library preparation method used. However, since RNA samples were processed using oligo(dT) chromatography before library construction, it is reasonable to assume that RBP33-regulated transcripts do contain poly(A) tails. In fact, transcripts derived from cSSRs in RBP33-depleted trypanosomes were shown to be polyadenylated by RNase H digestion assays [28].

These results suggest that at least some of the novel transcripts arising from RBP33 depletion are correctly processed and are therefore potentially translatable. Indeed, some of the antisense cSSR-derived transcripts encode peptides with high coding potential (Supplementary Figure S5A). This was directly addressed by inserting a gene encoding eYFP in an antisense orientation at two different cSSR loci in bloodstream trypanosomes (Supplementary Figure S5B; see the Materials and Methods for information on expression constructs). Proper integration of plasmids was assessed by PCR of genomic DNA (Supplementary Figure S5C). The resulting cell lines were further transfected with plasmid pGR70 [28] for tetracycline-inducible RNAi of RBP33, and eYFP expression was measured by flow cytometry. As expected, little fluorescence could be detected in

uninduced cultures. However, eYFP expression increased 3- to 10-fold upon depletion of RBP33 (Supplementary Figure S5D, E).

Transcripts encoding metacyclic VSGs are overexpressed in RBP33-depleted trypanosomes

Expression of telomeric regions is strongly repressed in *T. brucei* [17,58]. These loci usually harbor genes encoding VSGs which are organized in BES or MES (see the Introduction). H3v and base J are found at telomeric regions in *T. brucei* [18,59], and both BES and MES VSGs are overexpressed in Δ H3v Δ H4v Δ J trypanosomes [17,27]. We next addressed whether depletion of RBP33 had any effect on VSG expression. Normalized read count values corresponding to 17 BES VSG and 8 MES VSG genes were compared between control and RNAi samples. As a reference, the Δ H3v Δ H4v Δ J dataset was analyzed in the same manner. All MES VSG transcripts had increased expression in bloodstream RBP33-depleted cells, whereas the levels of mRNAs encoding BES VSGs were barely affected (Figure 2D; Supplementary Figure S6A; Supplementary Tables S13 and S14). Coverage plots of some telomeric regions are shown to illustrate these observations (Figure 2E; Supplementary Figure S6B–D). This was in apparent contrast to the Δ H3v Δ H4v Δ J triple KO mutant, where both BES and MES VSG transcripts were strongly up-regulated [27; see also Figure 2D, Supplementary Figure S6A and Supplementary Tables S13 and S14]. Changes in the expression of two MES VSGs upon RBP33 knockdown were confirmed by quantitative RT-PCR (Supplementary Figure S6E). Some MES VSG transcripts also increased in abundance upon RBP33 depletion in procyclic forms, although their expression levels were generally low compared with those observed in bloodstream forms (Supplementary Tables S13 and S14).

Analysis of chromatin accessibility in RBP33-depleted cells

The up-regulation of transcripts originating from cSSRs and MES VSGs could be explained by an increase in chromatin accessibility caused by RBP33 depletion. In fact, the absence of H3v, H4v and base J, which are enriched in these genomic loci, leads to a somewhat similar phenotype to that observed for RBP33 depletion [27], see above]. To test this hypothesis, we profiled chromatin accessibility by the ATAC-seq in control and RBP33-depleted trypanosomes, and compared the results with those obtained in an ATAC-seq analysis performed in a Δ H3v Δ H4v double KO mutant [17]. Nucleosome-free fragments were selected after aligning sequencing reads to the genome (size threshold = 100 bp, Supplementary Figure S7A), and used for further analysis. Pearson correlation coefficients were >0.940, indicating good reproducibility between replicates (Supplementary Figure S7B).

A standard quality control in ATAC-seq experiments involves the analysis of read coverage around known nucleosome-free regions (NFRs) such as TSSs of protein-coding genes [60]. However, no defined promoters are associated with individual genes in trypanosomatids, since genes are organized in PTUs. Instead, we measured read coverage

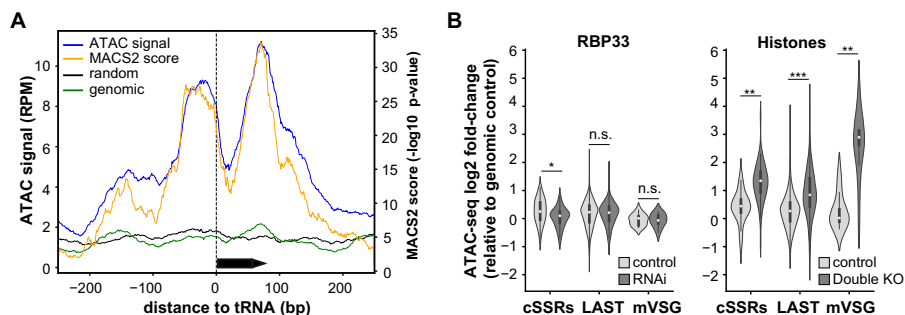


Figure 3. ATAC-seq analysis. (A) Single base coverage plot for ATAC signal at non-clustered tRNA genes in control bloodstream trypanosomes. Black arrow box represents a tRNA gene. Read counts were obtained for genomic regions centered at non-clustered tRNA genes and extended 250 bp on each side, normalized by library size (RPM), and represented as the average of RPM values (blue line) or MACS2 scores (orange line) across 10 different tRNA loci. ATAC signals corresponding to the average of 10 randomly chosen regions of the same size (black line), and to a transposed genomic DNA control (green line) are also shown. (B) Violin plots showing chromatin accessibility in controls (light gray violins), RBP33-depleted and $\Delta H3v\Delta H4v$ double KO trypanosomes (dark gray violins). Average ATAC signal values across ATAC-seq replicates were obtained for cSSRs, 'last' and metacyclic VSG genes, and expressed as the log₂ ratio relative to transposed genomic DNA control; only loci with CPM values ≥ 1 in at least two replicates were considered. Miniature box plots are shown inside each violin (boxes represent the IQR; whiskers, ± 1.5 IQR; white circles, medians). Two-sided unpaired Student's *t*-tests were used to assess whether chromatin accessibility was significantly altered in RBP33-depleted or $\Delta H3v\Delta H4v$ double KO samples relative to their respective control samples. n.s., not significant ($P > 0.5$); * $P < 5 \times 10^{-3}$; ** $P < 5 \times 10^{-6}$; *** $P < 5 \times 10^{-20}$ ($n = 58$, $n = 279$ and $n = 8$ for cSSRs, 'last' genes and metacyclic VSGs, respectively).

at tRNA loci, which are generally nucleosome free in *T. brucei* and other eukaryotes (61,62). Indeed, a clear NFR was observed when plotting the average of RPM across 10 different isolated (i.e. not clustered) tRNA genes (Figure 3A, blue line; see Supplementary Table S15 for a list of tRNA genes). Identification of open chromatin using the peak caller MACS2 gave essentially the same result (Figure 3A, orange line), whereas neither a transposed genomic DNA control sample (green line) nor the average of 10 randomly selected genomic loci of the same length (black line) showed a noticeable increase in ATAC-seq signal. Similar profiles of open chromatin at tRNA loci were obtained when RBP33 RNAi and $\Delta H3v\Delta H4v$ datasets were analyzed (Supplementary Figure S7C, D). NFRs were also observed at clustered tRNAs loci (Supplementary Figure S7E). A double peak pattern in ATAC signal was consistently observed in most tRNA genes with a valley immediately downstream of the TSS (Figure 3A; Supplementary Figure S7E). This could be interpreted as a footprint caused by an active RNA pol III transcription factor bound to the box A element of the tRNA gene promoter.

We next compared the RBP33 and $\Delta H3v\Delta H4v$ datasets using PCA. Read coverage ratios relative to transposed genomic DNA controls were obtained either gene-wise for all annotated genes in the *T. brucei* genome, or genome-wide after dividing the whole genome into 100 bp non-overlapping bins. RBP33 control and RNAi replicates clustered together in both analyses, whereas a marked separation was observed between histone controls (wild-type) and $\Delta H3v\Delta H4v$ samples (Supplementary Figure S8A, B). Moreover, strip plots representing coverage ratios between RBP33-depleted and control trypanosomes, or between $\Delta H3v\Delta H4v$ and wild-type cells, showed a much larger effect on chromatin accessibility in the latter dataset (Supple-

mentary Figure S8C, D). These observations already indicate that there are fewer changes in chromatin accessibility upon RBP33 depletion compared with those observed in $\Delta H3v\Delta H4v$ mutants. Indeed, when we looked more specifically at MES VSGs, cSSRs and genes flanking cSSRs ('last' genes), where H3v and H4v variants are enriched, we observed that chromatin was significantly more open in $\Delta H3v\Delta H4v$ trypanosomes, as expected, whereas RBP33 depletion did not result in a significant increase in chromatin accessibility (Figure 3B). Coverage plots representing the average of ATAC signals across 47 different cSSRs longer than 1.2 kb confirmed these observations (Supplementary Figure S8E). These results indicate that the observed up-regulation of transcripts derived from these regions does not seem to be due to increased chromatin accessibility in RBP33-depleted cells. In fact, chromatin seemed to be significantly less open at cSSRs in RBP33-depleted cells (Figure 3B), as well as in tRNAs, dSSRs and 'first' loci (Supplementary Figure S8F). On the other hand, BES VSG loci did show a small but significant ($P = 0.03$) increase in accessibility (Supplementary Figure S8F); however, these apparent changes in local chromatin structure did not result in up-regulation of BES VSG transcripts, as shown above.

RBP33 depletion results in the stabilization of regulated transcripts

To assess whether the observed changes in RNA abundance were due to alterations in RNA degradation rates, we analyzed in procyclic trypanosomes the decay of transcripts derived from two different cSSR loci and from a 'last' gene, Tb427.060007400, which flanks a third cSSR (see Supplementary Figure S9 for genomic localization and coverage plots). As controls, we measured

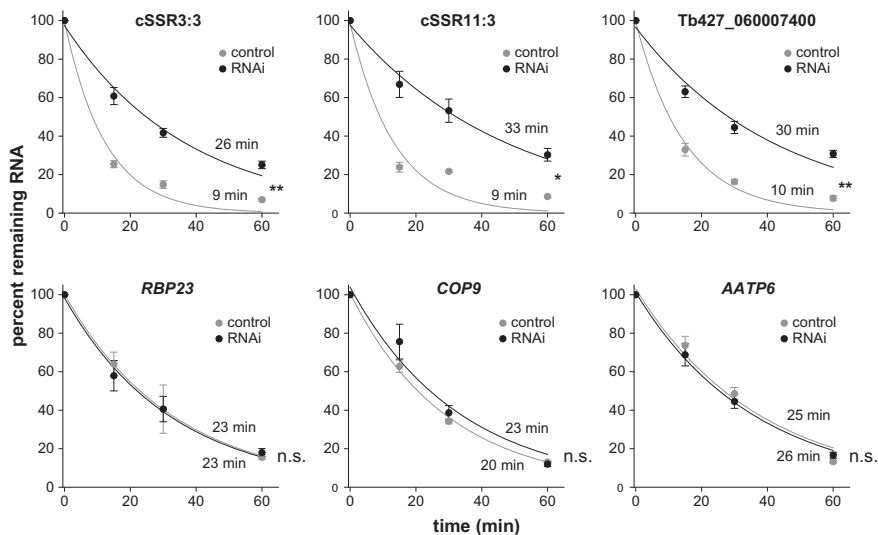


Figure 4. Effect of RBP33 depletion on transcript stability. Actinomycin D and sinefungin were added to control (gray) or RBP33-depleted (black) procyclic trypanosomes. RNA samples were collected at various time points and analyzed by quantitative RT-PCR. Values in each point represent the mean \pm SEM of three independent RNAi inductions. Estimated half-life values are indicated (top, RBP33-depleted; bottom, control). Two-sided unpaired Student's *t*-tests were used to assess whether half-lives were significantly altered upon RBP33 ablation. n.s., not significant ($P > 0.4$); * $P < 0.05$; ** $P < 0.005$.

the degradation of three transcripts encoding housekeeping proteins: *AATP6* (Tb427_080052600/Tb927.8.4700), *COP9* (Tb427_100084600/Tb927.10.7800) and *RBP23* (Tb427_100119800/Tb927.10.11270). *AATP6* and *COP9* are 'core' genes located in the middle of two PTUs, whereas *RBP23* is the first gene transcribed within a third PTU. Cells were incubated with actinomycin D and sinefungin to inhibit the synthesis and the maturation of mRNA simultaneously (63), and samples were collected over a period of 1 h. All six transcripts seemed to follow exponential decay patterns, as seen in Figure 4 ($R^2 > 0.95$ for all curve fits). We could observe that the half-lives of cSSR- and Tb427_060007400-derived transcripts increased ~ 3 -fold in RBP33-depleted cells, whereas the decay of housekeeping mRNAs was not significantly altered.

Interaction of RBP33 with other proteins

We used the TAP method coupled to mass spectrometry to analyze whether RBP33 binds to other proteins in the cell. We first generated a procyclic *T. brucei* cell line that expressed an N-terminal TAP-tagged version of RBP33 from the endogenous locus (Supplementary Figure S10A). TAP-RBP33 expression was localized to the nucleus (Supplementary Figure S10B). To check whether the fusion protein was functional, we transfected the cell line expressing TAP-RBP33 with an RNAi vector that produced dsRNA corresponding to the endogenous RBP33 5'-UTR. Since the

TAP-RBP33 gene contains a heterologous 5'-UTR, the expression of endogenous RBP33, but not TAP-RBP33, was efficiently silenced by RNAi using this strategy (Supplementary Figure S10C). As a control for RNAi efficiency, the RNAi plasmid was transfected in the parental cell line. The viability of the control cell line was severely impaired upon induction of RNAi, as expected, whereas the growth of trypanosomes expressing TAP-RBP33 was unaffected (Supplementary Figure S10D). This indicates that TAP-RBP33 fusion protein is functional.

TAP-purified material from two independent purifications was subjected to LC-MS/MS, and proteins were identified as described in the Materials and Methods. As controls, TAPs were carried out in parallel using protein extracts obtained from cell lines expressing the TAP tag alone (Supplementary Table S16) or the ribonucleoprotein complex PuREBP1/2 (47) using either PuREBP1-TAP or TAP-PuREBP2 as baits (Supplementary Table S17). We also compared the identified RBP33-associated proteins with a list of putative contaminants commonly observed in TAP purifications (64). Proteins were considered as RBP33 partners if they fulfilled the following criteria: (i) were detected in both TAP-RBP33-independent purifications; (ii) were not present more than once in the list of putative contaminants, TAP alone and PuREBP1/2 purifications; and (iii) are considered nuclear proteins based on database annotations or proteomic surveys (65). This gave as a list of six proteins (Table 1; see Supplementary Table S18 for all

Table 1. Proteins associated with RBP33 in procyclic *T. brucei* cells

Gene ID	Description
Tb927.10.13720	RNA-binding protein RBP29
Tb927.6.1470	Hypothetical protein, conserved
Tb927.9.6870	RNA-binding protein RBSR1
Tb927.10.9400	Splicing factor SF1
Tb927.2.4710	RNA-binding protein RRM1
Tb927.10.3500	Splicing factor U2AF2 (RBSR4)

identified proteins). RBP33 was found associated with other RNA-binding proteins, three of which are known splicing factors: RBSR1 [orthologous to the human splicing factor SRSF7 (66)], the basal splicing factor SF1 (67) and U2AF2/RBSR4 (68). The observed interactions with splicing factors are probably RNA mediated, as has been described for other RNA-binding proteins in *T. brucei* (53). RBSR1 counterparts are apparently present only in *T. cruzi*, *Bodo saltans* and *Paratrypanosoma confusum*, whereas orthologs corresponding to SF1 and U2AF2/RBSR4 are readily identified in all trypanosomatids for which genome sequences are available (<https://tritypdb.org>).

The *T. brucei* exosome is involved in the degradation of cSSR-derived transcripts

We showed above that transcripts originating from cSSR loci become stabilized upon depletion of RBP33, which indicates that they are normally degraded in physiological conditions. In other eukaryotes, the nuclear exosome complex plays a central role in RNA surveillance by degrading transcripts produced as a consequence of non-productive transcription (see the Introduction). Trypanosomes also have an exosome complex (35,69) which is localized mainly in the nucleus (36,70). To assess whether the exosome is responsible for the degradation of cSSR-derived transcripts, we generated bloodstream and procyclic cell lines that expressed dsRNA against the exosome catalytic subunits RRP44 and RRP6 [as described in (35)]. Depletion of either subunit was confirmed by immunoblot (Figure 5A). We next analyzed the levels of transcripts derived from four different cSSR loci in control versus exosome-depleted trypanosomes or in control versus RBP33-depleted cells. A marked accumulation of cSSR-derived transcripts was observed for all four loci in both bloodstream (Figure 5B) and procyclic (Figure 5C) cell lines upon exosome knock-down. In procyclic trypanosomes, exosome depletion resulted in even higher levels of cSSR transcripts than those detected upon depletion of RBP33, whereas the effect observed in bloodstream forms was similar in both exosome- and RBP33-depleted trypanosomes. This could be due to a lower RNAi efficiency in the latter case, as some protein could be observed in RNAi-induced bloodstream samples, especially in RRP6-depleted cells (Figure 5A).

DISCUSSION

Advances in RNA sequencing technologies have revealed that RNA pol II transcription is pervasive, and that RNA pol II-derived transcripts cover almost the entire eukaryotic genome (1). Pervasive transcription gives rise to an abundance of non-coding and apparently non-productive RNAs

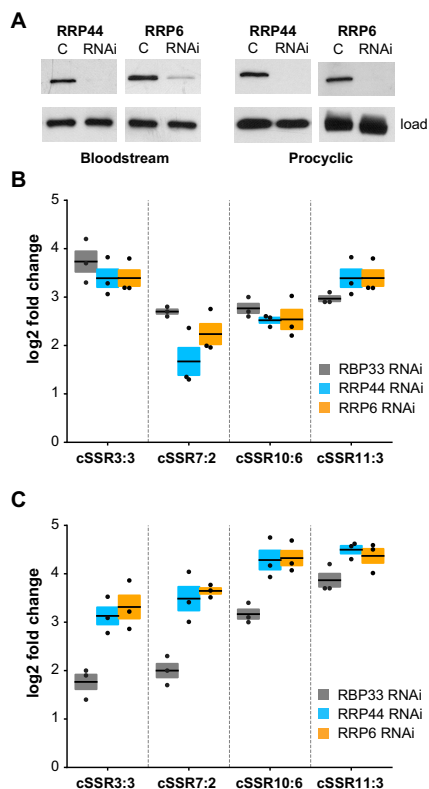


Figure 5. Effect of exosome depletion on the abundance of cSSR-derived transcripts. (A) Bloodstream and procyclic cell lines were generated that expressed dsRNA corresponding to RRP44 or RRP6 exosome subunits in a tetracycline-inducible manner. Depletion of each protein was monitored by immunoblot after 48 h of tetracycline induction (RNAi). C, control (parental) cell lines. RBP33 served as a loading control in all samples except in procyclic blots, where α -tubulin was used instead. (B and C) Quantitative RT-PCR analysis to measure the levels of transcripts derived from four convergent SSRs. Fold changes (\log_2 converted) relative to control cells are expressed as the mean (horizontal lines) \pm SEM (shaded areas) of three independent RNAi inductions in bloodstream (B) or procyclic (C) trypanosomes.

that are generated mainly from transcriptional readthrough at termination sites, and are degraded by surveillance pathways involving the RNA exosome (6,71). Transcription termination in animals and fungi is linked to RNA polyadenylation, but this cannot be the case in trypanosomatid protozoa. In these organisms, transcription is polycistronic, and individual mRNAs are generated by coupled *trans*-splicing and polyadenylation reactions (9). Interactions be-

tween splicing/polyadenylation and RNA pol II machineries have not been described in trypanosomes; in fact, mRNAs are probably processed once the RNA polymerase is 1–2 kb further downstream (72). RNA pol II transcription appears to be constitutive, and there are few initiation and termination sites per chromosome that are enriched in distinct histone variants (8,18). It is generally assumed that the presence of specific epigenetic marks (H3v, H4v and base J) is enough for efficient transcription termination in trypanosomatids, and that subtelomeres and cSSRs are kept silent simply because they are not transcribed (17,27).

A previous study from our laboratory showed that the RNA-binding protein RBP33 was able to bind specifically to transcripts derived from subtelomeres and cSSRs, and could therefore be an important regulator of the abundance of this type of transcripts (10,28). In this work we have explored this possibility and shown that depletion of RBP33 leads to a 7- to 8-fold increase in the abundance of antisense transcripts arising from TTSs and other silent genomic loci in both bloodstream and procyclic trypanosomes. RBP33 loss also caused overexpression of subtelomeric genes, especially in bloodstream cells; we do not know whether this is due to different RNAi efficiencies or reflects distinct roles of RBP33 in each life form. These changes are not due to cell death associated with RBP33 RNAi, as accumulation of cSSR-derived transcripts was not observed upon knockdown of DRBD3, an unrelated and essential RNA-binding protein (28). Moreover, the trypanosome transcriptome was minimally altered in cells depleted of another essential RNA-binding protein, PuREBP1 (47).

Although apparently similar, the phenotype observed in RBP33-depleted cells presents important differences relative to that exhibited by Δ H3v Δ H4v(Δ J) mutants. RBP33 is essential (28), whereas neither chromatin mark is essential by itself even though their absence results in the accumulation of TSS-derived transcripts (25,27). Importantly, procyclic trypanosomes lack base J, and we could observe a marked accumulation of TSS-derived transcripts in this life form. Moreover, the abundance of BES VSG transcripts was little affected in RBP33-depleted trypanosomes, whereas these RNAs were strongly up-regulated in Δ H3v Δ H4v(Δ J) mutants (17,27) probably due to changes in interchromosomal interactions and to detachment of BESs from specific nuclear sites in the latter case (17). In addition, chromatin accessibility was little altered in RBP33-depleted trypanosomes relative to Δ H3v Δ H4v mutants. Lastly, we could observe a much larger effect on antisense transcript abundance and overexpression of subtelomeric genes in RBP33-depleted cells. Even though the Δ H3v Δ H4v Δ J RNA-seq samples were obtained from conditional KO cell lines immediately after the knockout, these cell lines are based on Δ H4v mutants (27), and therefore we cannot rule out that the smaller effects seen in Δ H3v Δ H4v Δ J mutants are due to unknown events taking place during long periods of culturing that allow trypanosomes to adapt and adjust to readthrough transcription.

Transcripts encoding MES VSGs were overexpressed in RBP33-depleted trypanosomes, while those corresponding to BES VSGs remained unaltered. This suggests that the mechanisms involved in VSG repression are, at least in part,

different in MESs and BESs, as proposed previously (73). Intriguingly, the *RBP33* transcript was found to be up-regulated in metacyclic cells generated *in vitro* by inducible expression of a mutant version of the RNA-binding protein RBP6 (74). However, neither RBP33 protein nor transcript levels were altered in metacyclic trypanosomes obtained by the expression of wild-type RBP6 (73). Very little is known about how all MESs are kept silent in bloodstream form *T. brucei*; whether RBP33 has a role in the establishment of metacyclic forms is an interesting question that awaits further study, and could benefit from the use of single-cell RNA-seq of individual trypanosomes (17).

RBP33 loss also resulted in the down-regulation of many essential genes, probably due to RNAi or antisense silencing. This could readily explain why RBP33 is essential. In fact, antisense transcripts have been shown to alter the expression of nearby genes in other organisms (75,76). Moreover, we have shown that at least some of the antisense transcripts arising from RBP33 depletion seem to be correctly processed, and are therefore potentially translatable. Indeed, an eYFP transgene inserted in the opposite orientation at two different silent loci could be efficiently expressed in an RBP33-dependent fashion. Therefore, RBP33 depletion could lead not only to severe misregulation of numerous essential factors, but also to the appearance of potentially harmful proteins and peptides.

Trypanosomes lacking H3v, H4v and base J overexpress MES and BES VSG transcripts, show high DNA accessibility at repressed BESs and increased levels of transcripts derived from TTSs (17,27). It is known that chromatin opening can result in RNA pol II transcription initiation in trypanosomes (77). It is therefore conceivable that RBP33 depletion could result in chromatin opening at silent regions, which in turn would allow the expression of these silent chromosomal loci. However, no significant increase in ATAC-seq signal was observed at cSSRs and MES VSG loci in RBP33-depleted cells, whereas these regions were significantly more open in Δ H3v Δ H4v mutants. Thus, changes in chromatin accessibility do not seem to be responsible for the increased levels of transcripts that are observed upon RBP33 loss.

Importantly, we have shown that transcripts derived from three different cSSR loci become significantly stabilized upon RBP33 depletion, whereas the half-lives of three other, non-regulated transcripts remained unaltered, ruling out a general inhibition of the RNA degradation machinery caused by RBP33 knockdown or cell death. Moreover, we also shown that the exosome seems to be responsible for the degradation of cSSR-derived transcripts. Since RBP33 binds to RNAs derived from silent loci (28), our results strongly suggest that RBP33 is involved in targeting these type of transcripts for selective degradation by the exosome. This is further supported by the observation that overexpression of RBP33 results in an overall decrease of mRNA abundance (29). We cannot rule out, however, that RBP33 could have additional indirect effects on transcription, for example decreasing elongation rates near TTSs. The profound effects of RBP33 depletion on the transcriptome provide evidence for the existence of pervasive transcription in trypanosomes, as has been observed in other organisms. Thus, many genomic loci traditionally considered as 'non-

expressed' in trypanosomes, such as cSSRs and subtelomeres, would be silent not because they are not transcribed, but because the derived pervasive transcripts are targeted for degradation.

How does RBP33 function in transcript decay? Given the vast repertoire of RBP33-regulated transcripts, it is difficult to envisage a mechanism involving RBP33 targeting mediated by specific RNA sequence motifs. In fact, we could not identify them in the cohort of RBP33-bound transcripts (28), and RBP33 has been shown to have little specificity for RNA sequence *in vitro* (29). This is in agreement with the fact that, as shown above, RBP33 is able to regulate the expression of a heterologous gene, eYFP. Moreover, nuclear RNA degradation by the exosome does not seem to be selective but rather the default pathway in the competition between RNA destruction and processing in eukaryotes (78). This also holds true for trypanosomes, where it has been shown that competition between mRNA processing and co-transcriptional RNA degradation has an important role in determining steady-state RNA levels (51). In mammalian cells, Integrator-mediated destruction of pervasive transcripts does not seem to be sequence specific, but dependent on RNA pol II velocity instead. As a consequence, RNA degradation is more efficient when RNA pol II is stalled or slowed down, and it is especially noticeable on poorly expressed transcriptional units (6). This mechanism aids in cleansing a considerable fraction of background transcriptional activity that is below a functionally meaningful threshold (6). A similar scenario could occur in trypanosomes, with RBP33 targeting non-productive transcripts expressed at a low level that arise as a consequence of the attenuation of transcription elongation near TTSs or poorly transcribed loci, where RNA precursors are more susceptible to degradation relative to transcripts generated from ongoing transcription. This is supported by the fact that an eYFP transgene could not be expressed under physiological conditions from cSSR loci, but fluorescence was readily detected in RBP33-depleted cells. It is also supported, as mentioned above, by the observation that overall mRNA abundance is decreased in trypanosomes overexpressing RBP33 (29), which is probably the result of RBP33 binding to regular housekeeping transcripts not normally bound when the protein is expressed at physiological levels, targeting them for degradation. Our results are consistent with a scenario in which specific chromatin marks at termination sites slow down RNA pol II but cannot completely prevent readthrough transcription. The resulting antisense transcripts are efficiently turned over by RBP33 and the exosome. Readthrough transcription is increased in the absence of histone marks; as a consequence, higher levels of antisense transcripts are produced, a small fraction of which escape from degradation due to saturation of the surveillance machinery. Upon RBP33 loss, however, antisense transcripts accumulate to toxic levels because they cannot be targeted for degradation. This would imply that RBP33 dosage is very important, as has already been proposed (10).

We hypothesize that RBP33 could act by inhibiting *trans*-splicing of pervasive transcripts, resulting in unprocessed RNAs that are rapidly degraded by the exosome; conversely, in the absence of RBP33, non-productive transcripts would

be properly processed and thus be less prone to degradation. Consistent with this, we showed that RBP33 interacts with several splicing factors. Two of them, SF1 and U2AF2, are essential for early spliceosome formation, suggesting that RBP33 interferes with *trans*-splicing at this early stage. A role for RBP33 inhibiting *trans*-splicing is also supported by previous observations suggesting that RBP33 depletion leads to increased maturation or retroposon-derived transcripts (28), and also by the fact that the exosome is responsible for the degradation of misprocessed mRNAs in *T. brucei* (70). In other eukaryotes there is a direct interaction between proteins that target non-productive transcripts and the exosome complex. However, we could not detect any of the exosome subunits associated with RBP33. Although this could be explained by loose or transient protein-protein interactions, it could also be that trypanosomes possess unique mechanisms to deal with pervasive transcription as a consequence of the peculiar way in which they regulate gene expression. In fact, RBP33 seems to be a protein exclusive to trypanosomatids and, as mentioned above, mRNA processing and RNA pol II termination are not linked in these organisms. The actual mechanism by which RBP33 targets RNAs for destruction remains to be determined.

Our results show that the presence of specific histone marks at TTSs acting as 'roadblocks', although important, is not enough to prevent pervasive transcription and subtelomeric silencing, and that a surveillance mechanism involving RBP33 and the exosome exists in trypanosomes to cleanse non-productive transcription.

DATA AVAILABILITY

The RNA-seq and ATAC-seq data discussed in this publication are accessible through GEO Series accession number GSE186570. Flow cytometry raw data are available at FlowRepository with ID FR-FCM-Z55D.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Mark Carrington for eYFP plasmids and critical reading, and Eduardo Andrés, Laura Terrón and José L. Ruiz for helpful advice on bioinformatics and statistics.

FUNDING

This work was supported by the Spanish National Research Council (CSIC) [201920E114 to A.M.E.]; the Spanish Ministry of Science and Innovation [PID2019-111109RB-I00 to E.G.D.]; and La Caixa Foundation-Health Research 2020 Call [HR20-00635 to E.G.D.]. Funding for open access charge: Spanish Ministry of Science and Innovation. *Conflict of interest statement.* None declared.

REFERENCES

- Jensen, T.H., Jacquier, A. and Libri, D. (2013) Dealing with pervasive transcription. *Mol. Cell*, **52**, 473–484.

2. Wade,J.T. and Grainger,D.C. (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.*, **12**, 647–653.
3. Tudek,A., Porrua,O., Kabzinski,T., Lidschreiber,M., Kubicek,K., Fortova,A., Lacroute,F., Vanacova,S., Cramer,P., Steff,R. *et al.* (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell*, **55**, 467–481.
4. Xue,Y., Pradhan,S.K., Sun,F., Chronis,C., Tran,N., Su,T., Van,C., Vashisht,A., Wohlschlegel,J., Peterson,C.L. *et al.* (2017) Mot1, Ino80C, and NC2 function coordinately to regulate pervasive transcription in yeast and mammals. *Mol. Cell*, **67**, 594–607.
5. Holmes,R.K., Tuck,A.C., Zhu,C., Dunn-Davies,H.R., Kudla,G., Clauuder-Munster,S., Granneman,S., Steinmetz,L.M., Guthrie,C. and Tollervey,D. (2015) Loss of the yeast SR protein np13 alters gene expression due to transcription readthrough. *PLoS Genet.*, **11**, e1005735.
6. Lykke-Andersen,S., Žumer,K., Molska,E.Š., Rouvière,J.O., Wu,G., Demel,C., Schwab,B., Schmid,M., Cramer,P. and Jensen,T.H. (2021) Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell*, **81**, 514–529.
7. Tatomer,D.C., Elrod,N.D., Liang,D., Xiao,M.S., Jiang,J.Z., Jonathan,M., Huang,K.L., Wagner,E.J., Cherry,S. and Wilusz,J.E. (2019) The integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.*, **33**, 1525–1538.
8. Martínez-Calvillo,S., Yan,S., Nguyen,D., Fox,M., Stuart,K. and Myler,P.J. (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol. Cell*, **11**, 1291–1299.
9. Liang,X.H., Haritan,A., Uliel,S. and Michaeli,S. (2003) *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell*, **2**, 830–840.
10. Clayton,C. (2019) Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.*, **9**, 190072.
11. Matthews,K.R. (2005) The developmental cell biology of *Trypanosoma brucei*. *J. Cell Sci.*, **118**, 283–290.
12. Horn,D. (2014) Antigenic variation in African trypanosomes. *Mol. Biochem. Parasitol.*, **195**, 123–129.
13. Figueiredo,L.M., Cross,G.A.M. and Janzen,C.J. (2009) Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat. Rev. Microbiol.*, **7**, 504–513.
14. Günzl,A., Bruderer,T., Laufer,G., Schimanski,B., Tu,L.C., Chung,H.M., Lee,P.T. and Lee,M.G.S. (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot. Cell*, **2**, 542–551.
15. Graham,S.V. and Barry,J.D. (1995) Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of *Trypanosoma brucei*. *Mol. Cell Biol.*, **15**, 5945–5956.
16. Tetley,L., Turner,C.M.R., Barry,J.D., Crowe,J.S. and Vickerman,K. (1987) Onset of expression of the variant surface glycoproteins of *Trypanosoma brucei* in the tsetse fly studied using immunoelectron microscopy. *J. Cell Sci.*, **87**, 363–372.
17. Müller,L.S.M., Cosentino,R.O., Förstner,K.U., Guizetti,J., Wedel,C., Kaplan,N., Janzen,C.J., Arampatzis,P., Vogel,J., Steinbiss,S. *et al.* (2018) Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*, **563**, 121–125.
18. Siegel,T.N., Hekstra,D.R., Kempf,L.E., Figueiredo,L.M., Lowell,J.E., Fenyo,D., Wang,X., Dewell,S. and Cross,G.A.M. (2009) Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.*, **23**, 1063–1076.
19. Martínez-Calvillo,S., Vizuet-De-Rueda,J.C., Florencio-Martínez,L.E., Manning-Cela,R.G. and Figueroa-Angulo,E.E. (2010) Gene expression in trypanosomatid parasites. *J. Biomed. Biotechnol.*, **2010**, 525241.
20. Cliffe,L.J., Siegel,T.N., Marshall,M., Cross,G.A.M. and Sabatini,R. (2010) Two tyminidase hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*. *Nucleic Acids Res.*, **38**, 3923–3935.
21. Cordon-Obraes,C., Gomez-Liñan,C., Torres-Rusillo,S., Vidal-Cobo,I., Lopez-Farfán,D., Barroso-del-Jesus,A., Rojas-Barros,D., Carrington,M. and Navarro,M. (2022) Identification of sequence-specific promoters driving polycistronic transcription initiation by RNA polymerase II in trypanosomes. *Cell Rep.*, **38**, 110221.
22. Wedel,C., Förstner,K.U., Derr,R. and Siegel,T.N. (2017) GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J.*, **36**, 2581–2594.
23. Wright,J.R., Siegel,T.N. and Cross,G.A.M. (2010) Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **172**, 141–144.
24. Schulz,D., Zaringhalam,M., Papavasiliou,F.N. and Kim,H.S. (2016) Base j and H3.V regulate transcriptional termination in *Trypanosoma brucei*. *PLoS Genet.*, **12**, e100576.
25. Reynolds,D., Cliffe,L., Förstner,K.U., Hon,C.C., Siegel,T.N. and Sabatini,R. (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base j, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Res.*, **42**, 9717–9729.
26. Van Luenen,H.G.A.M., Farris,C., Jan,S., Genest,P.A., Tripathi,P., Velds,A., Kerkhoven,R.M., Nieuwland,M., Haydock,A., Ramasamy,G. *et al.* (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell*, **150**, 909–921.
27. Kim,H.-S. (2021) Genetic interaction between site-specific epigenetic marks and roles of H4v in transcription termination in trypanosoma brucei. *Front. Cell Dev. Biol.*, **9**, 744878.
28. Fernández-Moya,S.M., Carrington,M. and Estévez,A.M. (2014) Depletion of the RNA-binding protein RBP33 results in increased expression of silenced RNA polymerase II transcripts in *Trypanosoma brucei*. *PLoS One*, **9**, e107608.
29. Cirovic,O., Trikin,R., Hoffmann,A., Doiron,N., Jakob,M. and Ochsenreiter,T. (2017) The nuclear RNA binding protein RBP33 influences mRNA and spliced leader RNA abundance in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **212**, 16–20.
30. Biebinger,S., Wirtz,L.E., Lorenz,P. and Clayton,C.E. (1997) Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **85**, 99–112.
31. Brun,R., Jenni,L., Tanner,M., Schoenberger,M. and Schell,K.F. (1979) Cultivation of vertebrate infective forms derived from metacyclic forms of pleomorphic *Trypanosoma brucei* stocks. *Acta Trop.*, **36**, 387–390.
32. Wirtz,E., Leal,S., Ochatt,C. and Cross,G.A.M. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **99**, 89–102.
33. Hirumi,H. and Hirumi,K. (1989) Continuous cultivation of *Trypanosoma brucei* bloodstream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.*, **75**, 985–989.
34. Clayton,C., Estevez,A.M., Hartmann,C., Alibu,V.P., Field,M.C. and Horn,D. (2005) Down-regulating gene expression by RNA interference in *Trypanosoma brucei*. *Methods Mol Biol.*, **309**, 39–60.
35. Estévez,A.M., Kempf,T. and Clayton,C. (2001) The exosome of *Trypanosoma brucei*. *EMBO J.*, **20**, 3831–3839.
36. Haile,S., Cristodero,M., Clayton,C. and Estévez,A.M.A.M. (2007) The subcellular localisation of trypanosome RRP6 and its association with the exosome. *Mol. Biochem. Parasitol.*, **151**, 52–58.
37. Liao,Y., Smyth,G.K. and Shi,W. (2013) The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.*, **41**, e108.
38. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
39. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2009) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
40. Pedregosa,F., Varoquaux,G., Gramfort,A., Michel,V., Thirion,B., Grisel,O., Blondel,M., Prettenhofer,P., Weiss,R., Dubourg,V. *et al.* (2011) Scikit-learn: machine learning in {P}ython. *J. Mach. Learn. Res.*, **12**, 2825–2830.
41. Danecsek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A., Keane,T., McCarthy,S.A., Davies,R.M. *et al.* (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, **10**, giab008.
42. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.*, **57**, 289–300.
43. Ramirez,F., Ryan,D.P., Grünig,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2:

- a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
44. Siegel, T.N., Hekstra, D.R., Wang, X., Dewell, S. and Cross, G.A.M. (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.*, **38**, 4946–4957.
 45. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$. *Methods*, **25**, 402–408.
 46. Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B. et al. (2007) Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol. Biochem. Parasitol.*, **154**, 103–109.
 47. Rico-Jiménez, M., Ceballos-Pérez, G., Gómez-Liñán, C. and Estévez, A.M. (2021) An RNA-binding protein complex regulates the purine-dependent expression of a nucleobase transporter in trypanosomes. *Nucleic Acids Res.*, **49**, 3814–3825.
 48. Buenostro, J.D., Wu, B., Chang, H.Y. and Greenleaf, W.J. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.*, **109**, 21.29.1–21.29.9.
 49. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nat. Methods*, **9**, 357–359.
 50. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of chip-Seq (MACS). *Genome Biol.*, **9**, R137.
 51. Fadda, A., Ryten, M., Droll, D., Rojas, F., Färber, 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**, 307–326.
 52. Chen, C.Y.A., Ezzeddine, N. and Shyu, A.B. (2008) Messenger RNA half-life measurements in mammalian cells. *Methods Enzymol.*, **448**, 335–357.
 53. Fernández-Moya, S.M., García-Pérez, A., Kramer, S., Carrington, M. and Estévez, A.M. (2012) Alterations in DRBD3 ribonucleoprotein complexes in response to stress in *Trypanosoma brucei*. *PLoS One*, **7**, e48870.
 54. Puig, O., Caspar, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Séraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.
 55. Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C. and Horn, D. (2011) High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.*, **21**, 915–924.
 56. 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 Pathog.*, **6**, e1001090.
 57. 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**, 223.
 58. Cestari, I. and Stuart, K. (2017) Transcriptional regulation of telomeric expression sites and antigenic variation in trypanosomes. *Curr. Genomics*, **19**, 119–132.
 59. Cliffe, L.J., Kieft, R., Southern, T., Birkeland, S.R., Marshall, M., Sweeney, K. and Sabatini, R. (2009) JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res.*, **37**, 1452–1462.
 60. Yan, F., Powell, D.R., Curtis, D.J. and Wong, N.C. (2020) From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.*, **21**, 22.
 61. Maree, J.P., Povelones, M.L., Clark, D.J., Rudenko, G. and Patterson, H.G. (2017) Well-positioned nucleosomes punctuate polycistronic pol II transcription units and flank silent VSG gene arrays in *Trypanosoma brucei*. *Epigenetics Chromatin*, **10**, 14.
 62. Nagarajavel, V., Iben, J.R., Howard, B.H., Marais, R.J. and Clark, D.J. (2013) Global 'bootprinting' reveals the elastic architecture of the yeast TFIIB-TFIIC transcription complex in vivo. *Nucleic Acids Res.*, **41**, 8135–8143.
 63. Li, C.H., Irmer, H., Gudjonsdottir-Planck, D., Freese, S., Salm, H., Haile, S., Estévez, A.M. and Clayton, C. (2006) Roles of a *Trypanosoma brucei* 5'/3' exoribonuclease homolog in mRNA degradation. *RNA*, **12**, 2171–2186.
 64. Ouna, B.A., Stewart, M., Helbig, C. and Clayton, C. (2012) The trypanosome brucei CCCH zinc finger proteins ZC3H12 and ZC3H13. *Mol. Biochem. Parasitol.*, **183**, 184–188.
 65. Goos, C., Dejung, M., Janzen, C.J., Butter, F. and Kramer, S. (2017) The nuclear proteome of *Trypanosoma brucei*. *PLoS One*, **12**, e0181884.
 66. Wippel, H.H., Malgarin, J.S., Martins, S., de T., Vidal, N.M., Marcon, B.H., Miot, H.T., Marchini, F.K., Goldenberg, S. and Alves, L.R. (2019) The nuclear RNA-binding protein RBSR1 interactome in *Trypanosoma cruzi*. *J. Eukaryot. Microbiol.*, **66**, 244–253.
 67. Gupta, S.K., Carmi, S., Ben-Asher, H.W., Tkacz, I.D., Naboischikov, I. and Michaeli, S. (2013) Basal splicing factors regulate the stability of mature mRNAs in trypanosomes. *J. Biol. Chem.*, **288**, 4991–5006.
 68. Vazquez, M.P., Muallem, D., Bercovich, N., Stern, M.Z., Nyambega, B., Barda, O., Masiga, D., Gupta, S.K., Michaeli, S. and Levin, M.J. (2009) Functional characterization and protein-protein interactions of trypanosome splicing factors U2AF35, U2AF65 and SF1. *Mol. Biochem. Parasitol.*, **164**, 137–146.
 69. Estévez, A.M., Lehner, B., Sanderson, C.M., Ruppert, T., Clayton, C., Estévez, A.M., Lehner, B., Sanderson, C.M., Ruppert, T. and Clayton, C. (2003) The roles of intersubunit interactions in exosome stability. *J. Biol. Chem.*, **278**, 34943–34951.
 70. Kramer, S., Piper, S., Estévez, A. and Carrington, M. (2016) Polycistronic trypanosome mRNAs are a target for the exosome. *Mol. Biochem. Parasitol.*, **205**, 1–5.
 71. Villa, T. and Porrua, O. (2022) Pervasive transcription: a controlled risk. *FEBS J.*, <https://doi.org/10.1111/febs.16530>.
 72. Haanstra, J.R., Stewart, M., Luu, V.D., Van Tuijl, A., Westerhoff, H.V., Clayton, C. and Bakker, B.M. (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.
 73. Christiano, R., Kolev, N.G., Shi, H., Ullu, E., Walther, T.C. and Tschudi, C. (2017) The proteome and transcriptome of the infectious metacyclic form of *Trypanosoma brucei* define quiescent cells primed for mammalian invasion. *Mol. Microbiol.*, **106**, 74–92.
 74. Shi, H., Butler, K. and Tschudi, C. (2018) A single-point mutation in the RNA-binding protein 6 generates *Trypanosoma brucei* metacyclics that are able to progress to bloodstream forms in vitro. *Mol. Biochem. Parasitol.*, **224**, 50–56.
 75. Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, L.M. and Knop, M. (2016) Protein abundance control by non-coding antisense transcription. *Cell Rep.*, **15**, 2625–2636.
 76. Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W. and Steinmetz, L.M. (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol. Syst. Biol.*, **7**, 468.
 77. McAndrew, M., Graham, S., Hartmann, C. and Clayton, C. (1998) Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp. Parasitol.*, **90**, 65–76.
 78. Schmid, M. and Jensen, T.H. (2018) Controlling nuclear RNA levels. *Nat. Rev. Genet.*, **19**, 518–529.

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Afrin, M., Gaurav, A. K., Yang, X., Pan, X., Zhao, Y., & Li, B. (2020). TbRAP1 has an unusual duplex DNA binding activity required for its telomere localization and VSG silencing. *Science Advances*, 6(38). <https://doi.org/10.1126/sciadv.abc4065>
- Akama, T., Zhang, Y.-K., Freund, Y. R., Berry, P., Lee, J., Easom, E. E., Jacobs, R. T., Plattner, J. J., Witty, M. J., Peter, R., Rowan, T. G., Gillingwater, K., Brun, R., Nare, B., Mercer, L., Xu, M., Wang, J., & Liang, H. (2018). Identification of a 4-fluorobenzyl L-valinate amide benzoxaborole (AN11736) as a potential development candidate for the treatment of Animal African Trypanosomiasis (AAT). *Bioorganic & Medicinal Chemistry Letters*, 28(1), 6-10. <https://doi.org/10.1016/j.bmcl.2017.11.028>
- Akiyoshi, B., & Gull Sir William, K. (2013). Evolutionary cell biology of chromosome segregation: insights from trypanosomes. *Open Biology*, 3(5). <https://doi.org/10.1098/rsob.130023>
- Allmang, C. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO Journal*, 18(19), 5399-5410. <https://doi.org/10.1093/emboj/18.19.5399>
- Alsford, S., Turner, D. J., Obado, S. O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., & Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Research*, 21(6), 915-924. <https://doi.org/10.1101/gr.115089.110>
- Baalsrud, H. T., Tørresen, O. K., Solbakken, M. H., Salzburger, W., Hanel, R., Jakobsen, K. S., & Jentoft, S. (2018). *De novo* gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data. *Molecular Biology and Evolution*, 35(3), 593-606. <https://doi.org/10.1093/molbev/msx311>
- Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., & Grzybowska, E. A. (2019). RNA-protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biology*, 9(6). <https://doi.org/10.1098/rsob.190096>
- Barman, P., Reddy, D., & Bhaumik, S. (2019). Mechanisms of antisense transcription initiation with implications in gene expression, genomic integrity and disease pathogenesis. *Non-Coding RNA*, 5(1), 11. <https://doi.org/10.3390/ncrna5010011>
- Barnes, D. A., Mottram, J. C., & Agabian, N. (1990). Bloodstream and metacyclic variant surface glycoprotein gene expression sites of *Trypanosoma brucei* gambiense. *Molecular and Biochemical Parasitology*, 41(1), 101-114. [https://doi.org/10.1016/0166-6851\(90\)90101-Q](https://doi.org/10.1016/0166-6851(90)90101-Q)
- Barry, J. D. (1997). The relative significance of mechanisms of antigenic variation in African trypanosomes. *Parasitology Today*, 13(6), 212-218. [https://doi.org/10.1016/S0169-4758\(97\)01039-9](https://doi.org/10.1016/S0169-4758(97)01039-9)

BIBLIOGRAFÍA

- Benabdellah, K., González-Rey, E., & González, A. (2007). Alternative trans-splicing of the *Trypanosoma cruzi* *LYT1* gene transcript results in compartmental and functional switch for the encoded protein. *Molecular Microbiology*, 65(6), 1559-1567. <https://doi.org/10.1111/j.1365-2958.2007.05892.x>
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, 125(1-2), 279-284. [https://doi.org/10.1016/S0166-4328\(01\)00297-2](https://doi.org/10.1016/S0166-4328(01)00297-2)
- Benne, R. (1994). RNA editing in trypanosomes. *European Journal of Biochemistry*, 221(1), 9-23. <https://doi.org/10.1111/j.1432-1033.1994.tb18710.x>
- Berretta, J., & Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Reports* (Vol. 10, Issue 9, pp. 973-982). <https://doi.org/10.1038/embor.2009.181>
- Berretta, J., Pinskaya, M., & Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes & Development*, 22(5), 615-626. <https://doi.org/10.1101/gad.458008>
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., ... El-Sayed, N. M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, 309(5733), 416-422. <https://doi.org/10.1126/science.1112642>
- Biebinger, S., Elizabeth Wirtz, L., Lorenz, P., & Christine Clayton. (1997). Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 85(1), 99-112. [https://doi.org/10.1016/S0166-6851\(96\)02815-0](https://doi.org/10.1016/S0166-6851(96)02815-0)
- Biebinger, S., Rettenmaier, S., Flaspohler, J., Hartmann, C., Pena-Diaz, J., Wirtz, L. E., Hotz, H.-R., Barry, J. D., & Clayton, C. (1996). The PARP promoter of *Trypanosoma Brucei* is developmentally regulated in a chromosomal context. *Nucleic Acids Research*, 24(7), 1202-1211. <https://doi.org/10.1093/nar/24.7.1202>
- Borst, P. (2002). Antigenic variation and allelic exclusion. *Cell*, 109(1), 5-8. [https://doi.org/10.1016/S0092-8674\(02\)00711-0](https://doi.org/10.1016/S0092-8674(02)00711-0)
- Briggs, E., Hamilton, G., Crouch, K., Lapsley, C., & McCulloch, R. (2018). Genome-wide mapping reveals conserved and diverged R-loop activities in the unusual genetic landscape of the African trypanosome genome. *Nucleic Acids Research*, 46(22), 11789-11805. <https://doi.org/10.1093/nar/gky928>
- Brun, R., & Blum, J. (2012). Human African Trypanosomiasis. *Infectious Disease Clinics of North America*, 26(2), 261-273. <https://doi.org/10.1016/j.idc.2012.03.003>

- Brun, R., Blum, J., Chappuis, F., & Burri, C. (2010). Human African trypanosomiasis. *The Lancet*, 375(9709), 148-159. [https://doi.org/10.1016/S0140-6736\(09\)60829-1](https://doi.org/10.1016/S0140-6736(09)60829-1)
- Brun, R., Jenni, L., Tanner, M., Schönenberger, M., & Schell, K. F. (1979). Cultivation of vertebrate infective forms derived from metacyclic forms of pleomorphic *Trypanosoma brucei* stocks. Short communication. *Acta Tropica*, 36(4), 387-390.
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 109(1). <https://doi.org/10.1002/0471142727.mb2129s109>
- Bugge, K., Brakti, I., Fernandes, C. B., Dreier, J. E., Lundsgaard, J. E., Olsen, J. G., Skriver, K., & Kragelund, B. B. (2020). Interactions by disorder – A matter of context. *Frontiers in Molecular Biosciences*, 7. <https://doi.org/10.3389/fmolb.2020.00110>
- Bühlmann, M., Walrad, P., Rico, E., Ivens, A., Capewell, P., Naguleswaran, A., Roditi, I., & Matthews, K. R. (2015). NMD3 regulates both mRNA and rRNA nuclear export in African trypanosomes via an XPO1-linked pathway. *Nucleic Acids Research*, 43(9), 4491-4504. <https://doi.org/10.1093/nar/gkv330>
- Buljan, M., Chalancon, G., Eustermann, S., Wagner, G. P., Fuxreiter, M., Bateman, A., & Babu, M. M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Molecular Cell*, 46(6), 871-883. <https://doi.org/10.1016/j.molcel.2012.05.039>
- Burkard, G., Fragoso, C. M., & Roditi, I. (2007). Highly efficient stable transformation of bloodstream forms of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 153(2), 220-223. <https://doi.org/10.1016/j.molbiopara.2007.02.008>
- Buttner, K., Wenig, K., & Hopfner, K.-P. (2006). The exosome: a macromolecular cage for controlled RNA degradation. *Molecular Microbiology*, 61(6), 1372-1379. <https://doi.org/10.1111/j.1365-2958.2006.05331.x>
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppo, G., & Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell*, 131(4), 706-717. <https://doi.org/10.1016/j.cell.2007.09.014>
- Campbell, D. A., van Bree, M. P., & Boothroyd, J. C. (1984). The 5'-limit of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base-pair repeats flanking (TAA)₉₀. *Nucleic Acids Research*, 12(6), 2759-2774. <https://doi.org/10.1093/nar/12.6.2759>
- Campos, P. C., Bartholomeu, D. C., DaRocha, W. D., Cerqueira, G. C., & Teixeira, S. M. R. (2008). Sequences involved in mRNA processing in *Trypanosoma cruzi*. *International Journal for Parasitology*, 38(12), 1383-1389. <https://doi.org/10.1016/j.ijpara.2008.07.001>

BIBLIOGRAFÍA

- Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A. R. R., Carninci, P., Biffo, S., Stupka, E., & Gustincich, S. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*, *491*(7424), 454-457. <https://doi.org/10.1038/nature11508>
- Carvunis, A. R., Rolland, T., Wapinski, I., Calderwood, M. A., Yildirim, M. A., Simonis, N., Charlotheaux, B., Hidalgo, C. A., Barbette, J., Santhanam, B., Brar, G. A., Weissman, J. S., Regev, A., Thierry-Mieg, N., Cusick, M. E., & Vidal, M. (2012). Proto-genes and *de novo* gene birth. *Nature*, *487*(7407), 370-374. <https://doi.org/10.1038/nature11184>
- Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., & Zenklusen, D. (2013). Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nature Structural & Molecular Biology*, *20*(7), 851-858. <https://doi.org/10.1038/nsmb.2598>
- Cestari, I., & Stuart, K. (2018). Transcriptional regulation of telomeric expression sites and antigenic variation in trypanosomes. *Current Genomics*, *19*(2). <https://doi.org/10.2174/1389202918666170911161831>
- Chen, C. A., Ezzeddine, N., & Shyu, A. (2008). *Chapter 17 Messenger RNA Half-Life Measurements in Mammalian Cells* (pp. 335-357). [https://doi.org/10.1016/S0076-6879\(08\)02617-7](https://doi.org/10.1016/S0076-6879(08)02617-7)
- Christiano, R., Kolev, N. G., Shi, H., Ullu, E., Walther, T. C., & Tschudi, C. (2017). The proteome and transcriptome of the infectious metacyclic form of *Trypanosoma brucei* define quiescent cells primed for mammalian invasion. *Molecular Microbiology*, *106*(1), 74-92. <https://doi.org/10.1111/mmi.13754>
- Cirovic, O., Trikin, R., Hoffmann, A., Doiron, N., Jakob, M., & Ochsenreiter, T. (2017). The nuclear RNA binding protein RBP33 influences mRNA and spliced leader RNA abundance in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, *212*, 16-20. <https://doi.org/10.1016/j.molbiopara.2016.12.008>
- Clayton, C. (2002). Life without transcriptional control? From fly to man and back again. *The EMBO Journal*, *21*(8), 1881-1888. <https://doi.org/10.1093/emboj/21.8.1881>
- Clayton, C. (2019). Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. *Open Biology*, *9*(6). <https://doi.org/10.1098/rsob.190072>
- Clayton, C., Fueri, J. P., Itzhaki, J. E., Bellofatto, V., Sherman, D. R., Wisdom, G. S., Vijayarathy, S., & Mowatt, M. R. (1990). Transcription of the procyclic acidic repetitive protein genes of *Trypanosoma brucei*. *Molecular and Cellular Biology* (Vol. 10, Issue 6). <https://doi.org/10.1128/mcb.10.6.3036-3047.1990>
- Clayton, C., Estévez, A. M., Hartmann, C., Alibu, V. P., Field, M., & Horn, D. (2005). Down-Regulating Gene Expression by RNA Interference in *Trypanosoma brucei*. *Methods in molecular biology (Clifton, N.J.)*, *309*, 39-60. <https://doi.org/10.1385/1-59259-935-4:039>

- Clayton, C., & Michaeli, S. (2011). 3' processing in protists. *Wiley Interdisciplinary Reviews: RNA*, 2(2), 247-255. <https://doi.org/10.1002/wrna.49>
- Clayton, C., & Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and Biochemical Parasitology* (Vol. 156, Issue 2, pp. 93-101). <https://doi.org/10.1016/j.molbiopara.2007.07.007>
- Cliffe, L. J., Kieft, R., Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., & Sabatini, R. (2009). JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Research*, 37(5), 1452-1462. <https://doi.org/10.1093/nar/gkn1067>
- Cliffe, L. J., Siegel, T. N., Marshall, M., Cross, G. A. M., & Sabatini, R. (2010). Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*. *Nucleic Acids Research*, 38(12), 3923-3935. <https://doi.org/10.1093/nar/gkq146>
- Colgan, D. F., & Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes & Development*, 11(21), 2755-2766. <https://doi.org/10.1101/gad.11.21.2755>
- Connelly, S., & Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & Development*, 2(4), 440-452. <https://doi.org/10.1101/gad.2.4.440>
- Cordon-Obras, C., Gomez-Liñan, C., Torres-Rusillo, S., Vidal-Cobo, I., Lopez-Farfan, D., Barroso-del Jesus, A., Rojas-Barros, D., Carrington, M., & Navarro, M. (2022). Identification of sequence-specific promoters driving polycistronic transcription initiation by RNA polymerase II in trypanosomes. *Cell Reports*, 38(2), 110221. <https://doi.org/10.1016/j.celrep.2021.110221>
- Cross, G. A. M., Kim, H.-S., & Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Molecular and Biochemical Parasitology*, 195(1), 59-73. <https://doi.org/10.1016/j.molbiopara.2014.06.004>
- Dacks, J. B., Walker, G., & Field, M. C. (2008). Implications of the new eukaryotic systematics for parasitologists. *Parasitology International*, 57(2), 97-104. <https://doi.org/10.1016/j.parint.2007.11.004>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab008>
- Das, A., Banday, M., & Bellofatto, V. (2008). RNA Polymerase transcription machinery in trypanosomes. *Eukaryotic Cell*, 7(3), 429-434. <https://doi.org/10.1128/EC.00297-07>

BIBLIOGRAFÍA

- Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A. M., & Bellofatto, V. (2005). Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Molecular and Cellular Biology*, 25(16), 7314-7322. <https://doi.org/10.1128/MCB.25.16.7314-7322.2005>
- Davey, N. E., Cyert, M. S., & Moses, A. M. (2015). Short linear motifs – *ex nihilo* evolution of protein regulation. *Cell Communication and Signaling*, 13(1), 43. <https://doi.org/10.1186/s12964-015-0120-z>
- Davey, N. E., van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., & Gibson, T. J. (2012). Attributes of short linear motifs. *Mol. BioSyst.*, 8(1), 268-281. <https://doi.org/10.1039/C1MB05231D>
- D'avila-Levy, C. M., Boucinha, C., Kostygov, A., Lúcia, H., Santos, C., Morelli, K. A., Grybchuk-Ieremenko, A., Duval, L., Votýpka, J., Yurchenko, V., Grellier, P., Lukeš, J., & Cruz, F. O. (2015). Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem Inst Oswaldo Cruz*, 110(8), 956-965. <https://doi.org/10.1590/0074-02760150253>
- Davis, C. A., & Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 103(9), 3262-3267. <https://doi.org/10.1073/pnas.0507783103>
- de Gaudenzi, J., Frasch, A. C., & Clayton, C. (2005). RNA-binding domain proteins in kinetoplastids: a comparative analysis. *Eukaryotic Cell*, 4(12), 2106-2114. <https://doi.org/10.1128/EC.4.12.2106-2114.2005>
- de Lange, T., Liu, A. Y. C., Borst, P., Tromp, M. C., van Boom, J. H., & van der Ploeg, L. H. T. (1983). Tandem repetition of the 5' mini-exon of variant surface glycoprotein genes: A multiple promoter for VSG gene transcription? *Cell*, 34(3), 891-900. [https://doi.org/10.1016/0092-8674\(83\)90546-9](https://doi.org/10.1016/0092-8674(83)90546-9)
- de Lange, T., Michels, P. A. M., Veerman, H. J. G., Cornelissen, A. W. C. A., & Borst, P. (1984). Many trypanosome messenger RNAs share a common 5' terminal sequence. *Nucleic Acids Research*, 12(9), 3777-3790. <https://doi.org/10.1093/nar/12.9.3777>
- Dean, S., Marchetti, R., Kirk, K., & Matthews, K. R. (2009). A surface transporter family conveys the trypanosome differentiation signal. *Nature*, 459(7244), 213-217. <https://doi.org/10.1038/nature07997>
- Delan-Forino, C., Spanos, C., Rappsilber, J., & Tollervey, D. (2020). Substrate specificity of the TRAMP nuclear surveillance complexes. *Nature Communications*, 11(1), 3122. <https://doi.org/10.1038/s41467-020-16965-4>
- Delhi, P., Queiroz, R., Inchaustegui, D., Carrington, M., & Clayton, C. (2011). Is there a classical nonsense-mediated decay pathway in trypanosomes? *PLoS ONE*, 6(9). <https://doi.org/10.1371/journal.pone.0025112>

- Dillon, L. A. L., Okrah, K., Hughitt, V. K., Suresh, R., Li, Y., Fernandes, M. C., Belew, A. T., Corrada Bravo, H., Mosser, D. M., & El-Sayed, N. M. (2015). Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation. *Nucleic Acids Research*, *43*(14), 6799-6813. <https://doi.org/10.1093/nar/gkv656>
- D'Orso, I., & Frasch, A. C. C. (2001). Functionally different AU- and G-rich cis-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. *Journal of Biological Chemistry*, *276*(19), 15783-15793. <https://doi.org/10.1074/jbc.M010959200>
- Dostalova, A., Käser, S., Cristodero, M., & Schimanski, B. (2013). The nuclear mRNA export receptor Mex67-Mtr2 of *Trypanosoma brucei* contains a unique and essential zinc finger motif. *Molecular Microbiology*, *88*(4), 728-739. <https://doi.org/10.1111/mmi.12217>
- Duraisingh, M. T., & Horn, D. (2016). Epigenetic regulation of virulence gene expression in parasitic protozoa. *Cell Host & Microbe*, *19*(5), 629-640. <https://doi.org/10.1016/j.chom.2016.04.020>
- Dutrow, N., Nix, D. A., Holt, D., Milash, B., Dalley, B., Westbroek, E., Parnell, T. J., & Cairns, B. R. (2008). Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA-DNA hybrid mapping. *Nature Genetics*, *40*(8), 977-986. <https://doi.org/10.1038/ng.196>
- Ekanayake, D. K., Minning, T., Weatherly, B., Gunasekera, K., Nilsson, D., Tarleton, R., Ochsenreiter, T., & Sabatini, R. (2011). Epigenetic Regulation of transcription and virulence in *Trypanosoma cruzi* by O-linked thymine glucosylation of DNA. *Molecular and Cellular Biology*, *31*(8), 1690-1700. <https://doi.org/10.1128/MCB.01277-10>
- Elizabeth Hildreth, A., Ellison, M. A., Francette, A. M., Seraly, J. M., Lotka, L. M., & Arndt, K. M. (2020). The nucleosome DNA entry-exit site is important for transcription termination and prevention of pervasive transcription. *ELife*, *9*, 1-31. <https://doi.org/10.7554/ELIFE.57757>
- 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. M., Angiuoli, S., Anupama, A., ... Hall, N. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science*, *309*(5733), 404-409. <https://doi.org/10.1126/science.1112181>
- Estévez, A. M., Kempf, T., & Clayton, C. (2001). The exosome of *Trypanosoma brucei*. *The EMBO Journal*, *20*(14), 3831-3839. <https://doi.org/10.1093/emboj/20.14.3831>
- Estévez, A. M., Lehner, B., Sanderson, C. M., Ruppert, T., & Clayton, C. (2003). The roles of intersubunit interactions in exosome stability. *Journal of Biological Chemistry*, *278*(37), 34943-34951. <https://doi.org/10.1074/jbc.M305333200>

BIBLIOGRAFÍA

- Fadda, A., Ryten, M., Droll, D., Rojas, F., Färber, V., Haanstra, J. R., Merce, C., Bakker, B. M., Matthews, K., & 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. *Molecular Microbiology*, *94*(2), 307-326. <https://doi.org/10.1111/mmi.12764>
- Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., van der Brug, M. P., Nalls, M. A., Cookson, M. R., St-Laurent, G., & Wahlestedt, C. (2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biology*, *11*(5), R56. <https://doi.org/10.1186/gb-2010-11-5-r56>
- Faria, J., Glover, L., Hutchinson, S., Boehm, C., Field, M. C., & Horn, D. (2019). Monoallelic expression and epigenetic inheritance sustained by a *Trypanosoma brucei* variant surface glycoprotein exclusion complex. *Nature Communications*, *10*(1), 3023. <https://doi.org/10.1038/s41467-019-10823-8>
- Faria, J., Luzak, V., Müller, L. S. M., Brink, B. G., Hutchinson, S., Glover, L., Horn, D., & Siegel, T. N. (2021). Spatial integration of transcription and splicing in a dedicated compartment sustains monogenic antigen expression in African trypanosomes. *Nature Microbiology*, *6*(3), 289-300. <https://doi.org/10.1038/s41564-020-00833-4>
- Fernández-Moya, S. M., Carrington, M., & Estévez, A. M. (2014). Depletion of the RNA-binding protein RBP33 results in increased expression of silenced RNA polymerase II transcripts in *Trypanosoma brucei*. *PLoS ONE*, *9*(9). <https://doi.org/10.1371/journal.pone.0107608>
- Fernández-Moya, S. M., García-Pérez, A., Kramer, S., Carrington, M., & Estévez, A. M. (2012). Alterations in DRBD3 ribonucleoprotein complexes in response to stress in *Trypanosoma brucei*. *PLoS ONE*, *7*(11), e48870. <https://doi.org/10.1371/journal.pone.0048870>
- Figueiredo, L. M., Cross, G. A. M., & Janzen, C. J. (2009). Epigenetic regulation in African trypanosomes: a new kid on the block. *Nature Reviews Microbiology*, *7*(7), 504-513. <https://doi.org/10.1038/nrmicro2149>
- Flynn, I. W., & Bowman, I. B. R. (1973). The metabolism of carbohydrate by pleomorphic African trypanosomes. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, *45*(1), 25-42. [https://doi.org/10.1016/0305-0491\(73\)90281-2](https://doi.org/10.1016/0305-0491(73)90281-2)
- Frischmeyer, P. A., van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., & Dietz, H. C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*, *295*(5563), 2258-2261. <https://doi.org/10.1126/science.1067338>
- Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., Sengupta, A. M., & Vershon, A. K. (2011). Regulated antisense transcription controls expression of cell-type-specific genes in yeast. *Molecular and Cellular Biology*, *31*(8), 1701-1709. <https://doi.org/10.1128/MCB.01071-10>

- Ghedin, E., Bringaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J., Angiuoli, S., Wanless, D., von Arx, A., Murphy, L., Lennard, N., Salzberg, S., Adams, M. D., White, O., Hall, N., Stuart, K., ... El-Sayed, N. M. A. (2004). Gene synteny and evolution of genome architecture in trypanosomatids. *Molecular and Biochemical Parasitology*, *134*(2), 183-191. <https://doi.org/10.1016/J.molbiopara.2003.11.012>
- Gill, J. K., Maffioletti, A., García-Moliner, V., Stutz, F., & Soudet, J. (2020). Fine chromatin-driven Mechanism of transcription interference by antisense noncoding transcription. *Cell Reports*, *31*(5), 107612. <https://doi.org/10.1016/j.celrep.2020.107612>
- Glover, L., Hutchinson, S., Alford, S., & Horn, D. (2016). VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. *Proceedings of the National Academy of Sciences*, *113*(26), 7225-7230. <https://doi.org/10.1073/pnas.1600344113>
- Gómez-Liñán, C., Gómez-Díaz, E., Ceballos-Pérez, G., Fernández-Moya, S. M., & Estévez, A. M. (2022). The RNA-binding protein RBP33 dampens non-productive transcription in trypanosomes. *Nucleic Acids Research*, *50*(21), 12251-12265. <https://doi.org/10.1093/nar/gkac1123>
- Gommers-Ampt, J., Lutgerink, J., & Borst, P. (1991). A novel DNA nucleotide in *Trypanosoma brucei* only present in the mammalian phase of the life-cycle. *Nucleic Acids Research*, *19*(8), 1745-1751. <https://doi.org/10.1093/nar/19.8.1745>
- Goos, C., Dejung, M., Janzen, C. J., Butter, F., & Kramer, S. (2017). The nuclear proteome of *Trypanosoma brucei*. *PLOS ONE*, *12*(7), e0181884. <https://doi.org/10.1371/journal.pone.0181884>
- Goos, C., Dejung, M., Wehman, A. M., M-Natus, E., Schmidt, J., Sunter, J., Engstler, M., Butter, F., & Kramer, S. (2019). Trypanosomes can initiate nuclear export co-transcriptionally. *Nucleic Acids Research*, *47*(1), 266-282. <https://doi.org/10.1093/nar/gky1136>
- Graham, S. v., & Barry, J. D. (1995). Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of *Trypanosoma brucei*. *Molecular and Cellular Biology*, *15*(11), 5945-5956. <https://doi.org/10.1128/MCB.15.11.5945>
- Grosso, A. R., Leite, A. P., Carvalho, S., Matos, M. R., Martins, F. B., Vítor, A. C., Desterro, J. M., Carmo-Fonseca, M., & de Almeida, S. F. (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *ELife*, *4*. <https://doi.org/10.7554/eLife.09214>
- Gull K. (1999). The cytoskeleton of trypanosomatid parasites. *Annual review of microbiology*, *53*, 629–655. <https://doi.org/10.1146/annurev.micro.53.1.629>

BIBLIOGRAFÍA

- Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., Lee, P. T., & Lee, M. G. S. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryotic Cell*, 2(3), 542-551. <https://doi.org/10.1128/EC.2.3.542-551.2003>
- Günzl, A., Kirkham, J. K., Nguyen, T. N., Badjatia, N., & Park, S. H. (2015). Mono-allelic VSG expression by RNA polymerase I in *Trypanosoma brucei*: expression site control from both ends? *Gene*, 556(1), 68-73. <https://doi.org/10.1016/j.gene.2014.09.047>
- Günzl, A., Srivastava, A., & Gosavi, U. (2020). Transcription Factor Analysis in Trypanosomatids. En *Methods in Molecular Biology* (Vol. 2116, pp. 265-284). Humana Press Inc. https://doi.org/10.1007/978-1-0716-0294-2_16
- Gupta, S. K., Carmi, S., Ben-Asher, H. W., Tkacz, I. D., Naboishchikov, I., & Michaeli, S. (2013). Basal splicing factors regulate the stability of mature mRNAs in trypanosomes. *Journal of Biological Chemistry*, 288(7), 4991-5006. <https://doi.org/10.1074/jbc.M112.416578>
- Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, 482(7385), 339-346. <https://doi.org/10.1038/nature10887>
- Haanstra, J. R., Stewart, M., Luu, V.-D., van Tuijl, A., Westerhoff, H. v., Clayton, C., & Bakker, B. M. (2008). Control and regulation of gene expression. *Journal of Biological Chemistry*, 283(5), 2495-2507. <https://doi.org/10.1074/jbc.M705782200>
- Haile, S., Estévez, A. M., & Clayton, C. (2003). A role for the exosome in the in vivo degradation of unstable mRNAs. *RNA*, 9(12), 1491-1501. <https://doi.org/10.1261/rna.5940703>
- Haile, S., & Papadopoulou, B. (2007). Developmental regulation of gene expression in trypanosomatid parasitic protozoa. En *Current Opinion in Microbiology* (Vol. 10, Issue 6, pp. 569-577). <https://doi.org/10.1016/j.mib.2007.10.001>
- Han, S. M., Lee, T. H., Mun, J. Y., Kim, M. J., Kritikou, E. A., Lee, S.-J., Han, S. S., Hengartner, M. O., & Koo, H.-S. (2006). Deleted in cancer 1 (DICE1) is an essential protein controlling the topology of the inner mitochondrial membrane in *C. elegans*. *Development*, 133(18), 3597-3606. <https://doi.org/10.1242/dev.02534>
- Han, Z., Jasnovidova, O., Haidara, N., Tudek, A., Kubicek, K., Libri, D., Stefl, R., & Porrua, O. (2020). Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1. *The EMBO Journal*, 39(7). <https://doi.org/10.15252/embj.2019101548>
- Hata, T., & Nakayama, M. (2007). Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. *Biochimica et Biophysica Acta*, 1773(7), 1039-1051. <https://doi.org/10.1016/j.bbamcr.2007.04.010>

- Häusler, T., & Clayton, C. (1996). Post-transcriptional control of hsp70 mRNA in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 76(1-2), 57-71. [https://doi.org/10.1016/0166-6851\(95\)02538-3](https://doi.org/10.1016/0166-6851(95)02538-3)
- He, S., Valkov, E., Cheloufi, S., & Murn, J. (2022). The nexus between RNA-binding proteins and their effectors. *Nature Reviews Genetics*. <https://doi.org/10.1038/s41576-022-00550-0>
- Hegedúsová, E., Kulkarni, S., Burgman, B., Alfonzo, J. D., & Paris, Z. (2019). The general mRNA exporters Mex67 and Mtr2 play distinct roles in nuclear export of tRNAs in *Trypanosoma brucei*. *Nucleic Acids Research*, 47(16), 8620-8631. <https://doi.org/10.1093/nar/gkz671>
- Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., ... Berriman, M. (2008). Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS ONE*, 3(10), e3527. <https://doi.org/10.1371/journal.pone.0003527>
- Hirumi, H., & Hirumi, K. (1989). Continuous Cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *The Journal of Parasitology* (Vol. 75, Issue 6).
- Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., & Keene, J. D. (1991). RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proceedings of the National Academy of Sciences*, 88(6), 2495-2499. <https://doi.org/10.1073/pnas.88.6.2495>
- Holmes, R. K., Tuck, A. C., Zhu, C., Dunn-Davies, H. R., Kudla, G., Clauder-Munster, S., Granneman, S., Steinmetz, L. M., Guthrie, C., & Tollervey, D. (2015). Loss of the yeast SR protein Npl3 alters gene expression due to transcription readthrough. *PLoS Genetics*, 11(12). <https://doi.org/10.1371/journal.pgen.1005735>
- Horn, D. (2014). Antigenic variation in African trypanosomes. *Molecular and Biochemical Parasitology*, 195(2), 123-129. <https://doi.org/10.1016/j.molbiopara.2014.05.001>
- Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., & Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast*, 23(6), 439-454. <https://doi.org/10.1002/yea.1369>
- Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., & Vogelauer, M. (2008). A ncRNA Modulates histone modification and mRNA induction in the yeast *GAL* gene cluster. *Molecular Cell*, 32(5), 685-695. <https://doi.org/10.1016/j.molcel.2008.09.027>
- Hu, W., Alvarez-Dominguez, J. R., & Lodish, H. F. (2012). Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Reports*, 13(11), 971-983. <https://doi.org/10.1038/embor.2012.145>

BIBLIOGRAFÍA

- Huang, J., & van der Ploeg, L. H. (1991). Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. *The EMBO Journal*, *10*(12), 3877-3885. <https://doi.org/10.1002/j.1460-2075.1991.tb04957.x>
- Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, L. M., & Knop, M. (2016). Protein abundance control by non-coding antisense transcription. *Cell Reports*, *15*(12), 2625-2636. <https://doi.org/10.1016/j.celrep.2016.05.043>
- Hug, M., Hotz, H. R., Hartmann, C., & Clayton, C. (1994). Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. *Molecular and Cellular Biology*, *14*(11), 7428-7435. <https://doi.org/10.1128/mcb.14.11.7428-7435.1994>
- Hummel, H. S., Gillespie, R. D., & Swindle, J. (2000). Mutational analysis of 3' splice site selection during trans-splicing. *Journal of Biological Chemistry*, *275*(45), 35522-35531. <https://doi.org/10.1074/jbc.M002424200>
- Hutchinson, S., Foulon, S., Crouzols, A., Menafrá, R., Rotureau, B., Griffiths, A. D., & Bastin, P. (2021). The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of *Trypanosoma brucei* in the tsetse fly salivary glands. *PLoS Pathogens*, *17*(9). <https://doi.org/10.1371/journal.ppat.1009904>
- Inoue, H., Nojima, H., & Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, *96*(1), 23-28. [https://doi.org/10.1016/0378-1119\(90\)90336-P](https://doi.org/10.1016/0378-1119(90)90336-P)
- Irmer, H. (2001). Degradation of the unstable EP1 mRNA in *Trypanosoma brucei* involves initial destruction of the 3'-untranslated region. *Nucleic Acids Research*, *29*(22), 4707-4715. <https://doi.org/10.1093/nar/29.22.4707>
- Jackson, A. P., Quail, M. A., & Berriman, M. (2008). Insights into the genome sequence of a free-living Kinetoplastid: *Bodo saltans* (Kinetoplastida: Euglenozoa). *BMC Genomics*, *9*(1), 594. <https://doi.org/10.1186/1471-2164-9-594>
- Jacobs, R. T., Nare, B., Wring, S. A., Orr, M. D., Chen, D., Sligar, J. M., Jenks, M. X., Noe, R. A., Bowling, T. S., Mercer, L. T., Rewerts, C., Gaukel, E., Owens, J., Parham, R., Randolph, R., Beaudet, B., Bacchi, C. J., Yarlett, N., Plattner, J. J., ... Don, R. (2011). SCYX-7158, an Orally-Active Benzoxaborole for the Treatment of Stage 2 Human African Trypanosomiasis. *PLoS Neglected Tropical Diseases*, *5*(6), e1151. <https://doi.org/10.1371/journal.pntd.0001151>
- Jacobs, R. T., Plattner, J. J., Nare, B., Wring, S. A., Chen, D., Freund, Y., Gaukel, E. G., Orr, M. D., Perales, J. B., Jenks, M., Noe, R. A., Sligar, J. M., Zhang, Y.-K., Bacchi, C. J., Yarlett, N., & Don, R. (2011). Benzoxaboroles: a new class of potential drugs for human African trypanosomiasis. *Future Medicinal Chemistry*, *3*(10), 1259-1278. <https://doi.org/10.4155/fmc.11.80>

- Jensen, T. H., Jacquier, A., & Libri, D. (2013). Dealing with pervasive transcription. *Molecular Cell* (Vol. 52, Issue 4, pp. 473-484). <https://doi.org/10.1016/j.molcel.2013.10.032>
- Kaplan, C. D., Laprade, L., & Winston, F. (2003). Transcription elongation factors Repress transcription initiation from cryptic sites. *Science*, *301*(5636), 1096-1099. <https://doi.org/10.1126/science.1087374>
- Kapp, L. D., Abrams, E. W., Marlow, F. L., & Mullins, M. C. (2013). The integrator complex subunit 6 (Ints6) confines the dorsal organizer in vertebrate embryogenesis. *PLoS Genetics*, *9*(10), e1003822. <https://doi.org/10.1371/journal.pgen.1003822>
- Kassem, A., Pays, E., & Vanhamme, L. (2014). Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences*, *111*(24), 8943-8948. <https://doi.org/10.1073/pnas.1404873111>
- Kawano, M., Aravind, L., & Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Molecular Microbiology*, *64*(3), 738-754. <https://doi.org/10.1111/j.1365-2958.2007.05688.x>
- Keeling, P. J., & Burki, F. (2019). Progress towards the Tree of Eukaryotes. *Current biology : CB*, *29*(16), R808–R817. <https://doi.org/10.1016/j.cub.2019.07.031>
- Keene, J. D. (2010). The global dynamics of RNA stability orchestrates responses to cellular activation. *BMC Biology*, *8*(1), 95. <https://doi.org/10.1186/1741-7007-8-95>
- Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B., & Carrington, M. (2007). Functional genomics in *Trypanosoma brucei*: A collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Molecular and Biochemical Parasitology*, *154*(1), 103-109. <https://doi.org/10.1016/j.molbiopara.2007.03.012>
- Kennedy, P. G. E. (2008). The continuing problem of human African trypanosomiasis (sleeping sickness). *Annals of Neurology*, *64*(2), 116-126. <https://doi.org/10.1002/ana.21429>
- Kim, H. S. (2021). Genetic interaction between site-specific epigenetic marks and roles of H4v in transcription termination in *Trypanosoma brucei*. *Frontiers in Cell and Developmental Biology*, *9*. <https://doi.org/10.3389/fcell.2021.744878>
- Koch, H., Raabe, M., Urlaub, H., Bindereif, A., & Preußner, C. (2016). The polyadenylation complex of *Trypanosoma brucei*: Characterization of the functional poly(A) polymerase. *RNA Biology*, *13*(2), 221-231. <https://doi.org/10.1080/15476286.2015.1130208>
- Kolev, N. G., Franklin, J. B., Carmi, S., Shi, H., Michaeli, S., & Tschudi, C. (2010). The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathogens*, *6*(9), e1001090. <https://doi.org/10.1371/journal.ppat.1001090>

BIBLIOGRAFÍA

- Kolev, N. G., Ramey-Butler, K., Cross, G. A. M., Ullu, E., & Tschudi, C. (2012). Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science*, 338(6112), 1352-1353. <https://doi.org/10.1126/science.1229641>
- Kooter, J. M., & Borst, P. (1984). α -amanitin-insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Research*, 12. <https://academic.oup.com/nar/article/12/24/9457/1051958>
- Kramer, S. (2017). The ApaH-like phosphatase TbALPH1 is the major mRNA decapping enzyme of trypanosomes. *PLOS Pathogens*, 13(6), e1006456. <https://doi.org/10.1371/journal.ppat.1006456>
- Kramer, S., Piper, S., Estévez, A., & Carrington, M. (2016). Polycistronic trypanosome mRNAs are a target for the exosome. *Molecular and Biochemical Parasitology*, 205(1-2), 1-5. <https://doi.org/10.1016/j.molbiopara.2016.02.009>
- Krchňáková, Z., Krajčovič, J., & Vesteg, M. (2017). On the possibility of an early evolutionary origin for the spliced leader trans-splicing. *Journal of Molecular Evolution*, 85(1-2), 37-45. <https://doi.org/10.1007/s00239-017-9803-y>
- Kuehner, J. N., & Brow, D. A. (2008). Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Molecular Cell*, 31(2), 201-211. <https://doi.org/10.1016/j.molcel.2008.05.018>
- Kuehner, J. N., Pearson, E. L., & Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*, 12(5), 283-294. <https://doi.org/10.1038/nrm3098>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. <https://doi.org/10.1038/nmeth.1923>
- Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I. R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penadés, J. R., Valle, J., Solano, C., & Gingeras, T. R. (2011). Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences*, 108(50), 20172-20177. <https://doi.org/10.1073/pnas.1113521108>
- LeBowitz, J. H., Coburn, C. M., McMahon-Pratt, D., & Beverley, S. M. (1990). Development of a stable *Leishmania* expression vector and application to the study of parasite surface antigen genes. *Proceedings of the National Academy of Sciences*, 87(24), 9736-9740. <https://doi.org/10.1073/pnas.87.24.9736>
- Lecordier, L., Devaux, S., Uzureau, P., Dierick, J. F., Walgraffe, D., Poelvoorde, P., Pays, E., & Vanhamme, L. (2007). Characterization of a TFIIF homologue from *Trypanosoma brucei*. *Molecular Microbiology*, 64(5), 1164-1181. <https://doi.org/10.1111/j.1365-2958.2007.05725.x>

- Lee, J. H., Cai, G., Panigrahi, A. K., Dunham-Ems, S., Nguyen, T. N., Radolf, J. D., Asturias, F. J., & Günzl, A. (2010). A TFIIH-associated mediator head is a basal factor of small nuclear spliced leader RNA gene transcription in early-diverged trypanosomes. *Molecular and Cellular Biology*, *30*(23), 5502-5513. <https://doi.org/10.1128/MCB.00966-10>
- Lejeune, F., Li, X., & Maquat, L. E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylation, and exonucleolytic activities. *Molecular Cell*, *12*(3), 675-687. [https://doi.org/10.1016/S1097-2765\(03\)00349-6](https://doi.org/10.1016/S1097-2765(03)00349-6)
- Li, C., Irmer, H., Gudjonsdottir-Planck, D., Freese, S., Salm, H., Haile, S., Estévez, A. M., & Clayton, C. (2006). Roles of a *Trypanosoma brucei* 5'→3' exoribonuclease homolog in mRNA degradation. *RNA*, *12*(12), 2171-2186. <https://doi.org/10.1261/rna.291506>
- Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108-e108. <https://doi.org/10.1093/nar/gkt214>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930. <https://doi.org/10.1093/bioinformatics/btt656>
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*, *25*(4), 402-408. <https://doi.org/10.1006/meth.2001.1262>
- Logan, J., Falck-Pedersen, E., Darnell, J. E., & Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences*, *84*(23), 8306-8310. <https://doi.org/10.1073/pnas.84.23.8306>
- Lorković, Z. J. (2002). Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Research*, *30*(3), 623-635. <https://doi.org/10.1093/nar/30.3.623>
- Lukeš, J., Butenko, A., Hashimi, H., Maslov, D. A., Votýpka, J., & Yurchenko, V. (2018). Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. *Trends in parasitology*, *34*(6), 466-480. <https://doi.org/10.1016/j.pt.2018.03.002>
- Lukeš, J., Skalický, T., Týč, J., Votýpka, J., & Yurchenko, V. (2014). Evolution of parasitism in kinetoplastid flagellates. *Molecular and Biochemical Parasitology*, *195*(2), 115-122. <https://doi.org/10.1016/j.molbiopara.2014.05.007>
- Luo, W., Johnson, A. W., & Bentley, D. L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes & Development*, *20*(8), 954-965. <https://doi.org/10.1101/gad.1409106>

BIBLIOGRAFÍA

- Lykke-Andersen, S., Žumer, K., Molska, E. Š., Rouvière, J. O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., & Jensen, T. H. (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Molecular Cell*, 81(3), 514-529.e6. <https://doi.org/10.1016/j.molcel.2020.12.014>
- Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., & Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *ELife*, 4. <https://doi.org/10.7554/eLife.06722>
- Manful, T., Fadda, A., & Clayton, C. (2011). The role of the 5'–3' exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA*, 17(11), 2039-2047. <https://doi.org/10.1261/rna.2837311>
- Manley, J. L., Sharp, P. A., & Gefter, M. L. (1982). RNA synthesis in isolated nuclei. *Journal of Molecular Biology*, 159(4), 581-599. [https://doi.org/10.1016/0022-2836\(82\)90102-4](https://doi.org/10.1016/0022-2836(82)90102-4)
- Marchetti, M. (1998). Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersions of RNA polymerase II- and III-transcribed genes. *Nucleic Acids Research*, 26(15), 3591-3598. <https://doi.org/10.1093/nar/26.15.3591>
- Maree, J. P., & Patterson, H.-G. (2014). The epigenome of *Trypanosoma brucei*: A regulatory interface to an unconventional transcriptional machine. *Biochimica et Biophysica Acta*, 1839(9), 743-750. <https://doi.org/10.1016/j.bbagr.2014.05.028>
- Maree, J. P., Povelones, M. L., Clark, D. J., Rudenko, G., & Patterson, H. G. (2017). Well-positioned nucleosomes punctuate polycistronic pol II transcription units and flank silent VSG gene arrays in *Trypanosoma brucei*. *Epigenetics and Chromatin*, 10(1). <https://doi.org/10.1186/s13072-017-0121-9>
- Martens, J. A., Laprade, L., & Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature*, 429(6991), 571-574. <https://doi.org/10.1038/nature02538>
- Martínez-Calvillo, S., Vizuet-de-Rueda, J. C., Florencio-Martínez, L. E., Manning-Cela, R. G., & Figueroa-Angulo, E. E. (2010). Gene expression in trypanosomatid parasites. *Journal of Biomedicine and Biotechnology*, 2010, 1-15. <https://doi.org/10.1155/2010/525241>
- Martínez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., & Myler, P. J. (2003). Transcription of *Leishmania major* friedlin chromosome 1 initiates in both directions within a single region. *Molecular Cell*, 11(5), 1291-1299. [https://doi.org/10.1016/S1097-2765\(03\)00143-6](https://doi.org/10.1016/S1097-2765(03)00143-6)
- Matthews, K. R. (2005). The developmental cell biology of *Trypanosoma brucei*. *Journal of Cell Science*, 118(2), 283-290. <https://doi.org/10.1242/jcs.01649>

- Matthews, K. R., Tschudi, C., & Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes & Development*, *8*(4), 491-501. <https://doi.org/10.1101/gad.8.4.491>
- Mattick, J. S. (2007). A new paradigm for developmental biology. *Journal of Experimental Biology*, *210*(9), 1526-1547. <https://doi.org/10.1242/jeb.005017>
- McAndrew, M., Graham, S., Hartmann, C., & Clayton, C. (1998). Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Experimental Parasitology*, *90*(1), 65-76. <https://doi.org/10.1006/expr.1998.4317>
- Medina, D. A., Jordán-Pla, A., Millán-Zambrano, G., Chávez, S., Choder, M., & Pérez-Ortín, J. E. (2014). Cytoplasmic 5'-3' exonuclease Xrn1p is also a genome-wide transcription factor in yeast. *Frontiers in Genetics*, *5*. <https://doi.org/10.3389/fgene.2014.00001>
- Mesu, V. K. B. K., Kalonji, W. M., Bardonneau, C., Mordt, O. V., Blesson, S., Simon, F., Delhomme, S., Bernhard, S., Kuziena, W., Lubaki, J.-P. F., Vuvu, S. L., Ngima, P. N., Mbembo, H. M., Ilunga, M., Bonama, A. K., Heradi, J. A., Solomo, J. L. L., Mandula, G., Badibabi, L. K., ... Tarral, A. (2018). Oral fexinidazole for late-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a pivotal multicentre, randomised, non-inferiority trial. *The Lancet*, *391*(10116), 144-154. [https://doi.org/10.1016/S0140-6736\(17\)32758-7](https://doi.org/10.1016/S0140-6736(17)32758-7)
- Michaeli, S. (2011). Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiology*, *6*(4), 459-474. <https://doi.org/10.2217/fmb.11.20>
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., & Tollervey, D. (1997). The Exosome: A Conserved Eukaryotic RNA Processing Complex Containing Multiple 3'→5' Exoribonucleases. *Cell*, *91*(4), 457-466. [https://doi.org/10.1016/S0092-8674\(00\)80432-8](https://doi.org/10.1016/S0092-8674(00)80432-8)
- Morris, K. v., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, *15*(6), 423-437. <https://doi.org/10.1038/nrg3722>
- Müller, L. S. M., Cosentino, R. O., Förstner, K. U., Guizzetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A. E., Sebra, R. P., & Siegel, T. N. (2018). Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature*, *563*(7729), 121-125. <https://doi.org/10.1038/s41586-018-0619-8>
- Müller-McNicoll, M., & Neugebauer, K. M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nature Reviews Genetics*, *14*(4), 275-287. <https://doi.org/10.1038/nrg3434>
- Myler, P. J., Allison, J., Agabian, N., & Stuart, K. (1984). Antigenic variation in African trypanosomes by gene replacement or activation of alternate telomeres. *Cell*, *39*(1), 203-211. [https://doi.org/10.1016/0092-8674\(84\)90206-X](https://doi.org/10.1016/0092-8674(84)90206-X)

BIBLIOGRAFÍA

- Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L. M., de Nadal, E., & Posas, F. (2014). Control of Cdc28 CDK1 by a stress-induced lncRNA. *Molecular Cell*, 53(4), 549-561. <https://doi.org/10.1016/j.molcel.2014.01.006>
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J., & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, 348(6301), 515-520. <https://doi.org/10.1038/348515a0>
- Nagarajavel, V., Iben, J. R., Howard, B. H., Maraia, R. J., & Clark, D. J. (2013). Global 'bootprinting' reveals the elastic architecture of the yeast TFIIB-TFIIC transcription complex in vivo. *Nucleic Acids Research*, 41(17), 8135-8143. <https://doi.org/10.1093/nar/gkt611>
- Navarro, M. (1999). *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. *The EMBO Journal*, 18(8), 2265-2272. <https://doi.org/10.1093/emboj/18.8.2265>
- Navarro, M., & Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature*, 414(6865), 759-763. <https://doi.org/10.1038/414759a>
- Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., & Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 457(7232), 1038-1042. <https://doi.org/10.1038/nature07747>
- Ouna, B. A., Nyambega, B., Manful, T., Helbig, C., Males, M., Fadda, A., & Clayton, C. (2012). Depletion of trypanosome CTR9 Leads to gene expression defects. *PLoS ONE*, 7(4), e34256. <https://doi.org/10.1371/journal.pone.0034256>
- Palenchar, J. B., & Bellofatto, V. (2006). Gene transcription in trypanosomes. *Molecular and Biochemical Parasitology*, 146(2), 135-141. <https://doi.org/10.1016/j.molbiopara.2005.12.008>
- Pandya, U. M., Sandhu, R., & Li, B. (2013). Silencing subtelomeric VSGs by *Trypanosoma brucei* RAP1 at the insect stage involves chromatin structure changes. *Nucleic Acids Research*, 41(16), 7673-7682. <https://doi.org/10.1093/nar/gkt562>
- Parker, R., & Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. *Nature Structural & Molecular Biology*, 11(2), 121-127. <https://doi.org/10.1038/nsmb724>
- Patraquim, P., Magny, E. G., Pueyo, J. I., Platero, A. I., & Couso, J. P. (2022). Translation and natural selection of micropeptides from long non-canonical RNAs. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-34094-y>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikit-learn: machine learning in {P}ython. *J. Mach. Learn. Res.*, 12, 2825-2830.

- Pelechano, V., Wei, W., & Steinmetz, L. M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, *497*(7447), 127-131. <https://doi.org/10.1038/nature12121>
- Perry, K. L., Watkins, K. P., & Agabian, N. (1987). Trypanosome mRNAs have unusual «cap 4» structures acquired by addition of a spliced leader. *Proceedings of the National Academy of Sciences*, *84*(23), 8190-8194. <https://doi.org/10.1073/pnas.84.23.8190>
- Pinskaya, M., Gourvenec, S., & Morillon, A. (2009). H3 lysine 4 di- and trimethylation deposited by cryptic transcription attenuates promoter activation. *The EMBO Journal*, *28*(12), 1697-1707. <https://doi.org/10.1038/emboj.2009.108>
- Podlipaev S. (1990). Catalogue of world fauna of trypanosomatidae (Protozoa). (Vol. 217). *Proceedings of the Zoological Institute*, *217*, 1-177.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H., & Jensen, T. H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, *322*(5909), 1851-1854. <https://doi.org/10.1126/science.1164096>
- Priest, J. W., & Hajduk, S. L. (1994). Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*. *Journal of Bioenergetics and Biomembranes*, *26*(2), 179-191. <https://doi.org/10.1007/BF00763067>
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., & Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, *24*(3), 218-229. <https://doi.org/10.1006/meth.2001.1183>
- Rajan, K. S., Doniger, T., Cohen-Chalamish, S., Rengaraj, P., Galili, B., Aryal, S., Unger, R., Tschudi, C., & Michaeli, S. (2020). Developmentally regulated novel non-coding anti-sense regulators of mRNA translation in *Trypanosoma brucei*. *IScience*, *23*(12), 101780. <https://doi.org/10.1016/j.isci.2020.101780>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160-W165. <https://doi.org/10.1093/nar/gkw257>
- Rastrojo, A., Carrasco-Ramiro, F., Martín, D., Crespillo, A., Reguera, R. M., Aguado, B., & 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), 223. <https://doi.org/10.1186/1471-2164-14-223>
- Rettig, J., Wang, Y., Schneider, A., & Ochsenreiter, T. (2012). Dual targeting of isoleucyl-tRNA synthetase in *Trypanosoma brucei* is mediated through alternative trans-splicing. *Nucleic Acids Research*, *40*(3), 1299-1306. <https://doi.org/10.1093/nar/gkr794>

BIBLIOGRAFÍA

- Reynolds, D., Cliffe, L., Förstner, K. U., Hon, C. C., Siegel, T. N., & Sabatini, R. (2014). Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Research*, 42(15), 9717-9729. <https://doi.org/10.1093/nar/gku714>
- Reynolds, D., Hofmeister, B. T., Cliffe, L., Alabady, M., Siegel, T. N., Schmitz, R. J., & Sabatini, R. (2016). Histone H3 variant regulates RNA polymerase II transcription termination and dual strand transcription of siRNA loci in *Trypanosoma brucei*. *PLoS Genetics*, 12(1). <https://doi.org/10.1371/journal.pgen.1005758>
- Richard, P., & Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes & Development*, 23(11), 1247-1269. <https://doi.org/10.1101/gad.1792809>
- Rico-Jiménez, M., Ceballos-Pérez, G., Gómez-Liñán, C., & Estévez, A. M. (2021). An RNA-binding protein complex regulates the purine-dependent expression of a nucleobase transporter in trypanosomes. *Nucleic Acids Research*, 49(7), 3814-3825. <https://doi.org/10.1093/nar/gkab181>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roditi, I., & Liniger, M. (2002). Dressed for success: the surface coats of insect-borne protozoan parasites. *Trends in Microbiology*, 10(3), 128-134. [https://doi.org/10.1016/S0966-842X\(02\)02309-0](https://doi.org/10.1016/S0966-842X(02)02309-0)
- Ruan, J., Arhin, G. K., Ullu, E., & Tschudi, C. (2004). Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related Factor points to a universal regulator of transcription in trypanosomes. *Molecular and Cellular Biology*, 24(21), 9610-9618. <https://doi.org/10.1128/MCB.24.21.9610-9618.2004>
- Rudenko, G., Bishop, D., Gottesdiener, K., & van der Ploeg, L. H. (1989). Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form *Trypanosoma brucei*. *The EMBO Journal*, 8(13), 4259-4263. <https://doi.org/10.1002/j.1460-2075.1989.tb08611.x>
- Rutkowski, R. J., & Warren, W. D. (2009). Phenotypic analysis of deflated/Ints7 function in *Drosophila* development. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, 238(5), 1131-1139. <https://doi.org/10.1002/dvdy.21922>
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press.
- Sambrook, J., & Russell, D. W. (2006). SDS-polyacrylamide gel electrophoresis of proteins. *Cold Spring Harbor Protocols*, 4. <https://doi.org/10.1101/pdb.prot4540>
- Schilders, G. (2005). MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation. *Nucleic Acids Research*, 33(21), 6795-6804. <https://doi.org/10.1093/nar/gki982>

- Schimanski, B. (2006). A TFIIB-like protein is indispensable for spliced leader RNA gene transcription in *Trypanosoma brucei*. *Nucleic Acids Research*, *34*(6), 1676-1684. <https://doi.org/10.1093/nar/gkl090>
- Schimanski, B., Nguyen, T. N., & Günzl, A. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in *Trypanosoma brucei*. *Molecular and Cellular Biology*, *25*(16), 7303-7313. <https://doi.org/10.1128/MCB.25.16.7303-7313.2005>
- Schmid, M., & Jensen, T. H. (2018). Controlling nuclear RNA levels. *Nature Reviews Genetics*, *19*(8), 518-529. <https://doi.org/10.1038/s41576-018-0013-2>
- Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., & Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*, *155*(5), 1075. <https://doi.org/10.1016/j.cell.2013.10.024>
- Schulz, D., Zaringhalam, M., Papavasiliou, F. N., & Kim, H. S. (2016). Base J and H3.V regulate transcriptional termination in *Trypanosoma brucei*. *PLoS Genetics*, *12*(1). <https://doi.org/10.1371/journal.pgen.1005762>
- Schwede, A., Ellis, L., Luther, J., Carrington, M., Stoecklin, G., & Clayton, C. (2008). A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. *Nucleic Acids Research*, *36*(10), 3374-3388. <https://doi.org/10.1093/nar/gkn108>
- Schwede, A., Manful, T., Jha, B. A., Helbig, C., Bercovich, N., Stewart, M., & Clayton, C. (2009). The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. *Nucleic Acids Research*, *37*(16), 5511-5528. <https://doi.org/10.1093/nar/gkp571>
- Shearwin, K., Callen, B., & Egan, J. (2005). Transcriptional interference – a crash course. *Trends in Genetics*, *21*(6), 339-345. <https://doi.org/10.1016/j.tig.2005.04.009>
- Shi, H., Butler, K., & Tschudi, C. (2018). A single-point mutation in the RNA-binding protein 6 generates *Trypanosoma brucei* metacyclics that are able to progress to bloodstream forms in vitro. *Molecular and Biochemical Parasitology*, *224*, 50-56. <https://doi.org/10.1016/j.molbiopara.2018.07.011>
- Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., Wang, X., Dewell, S., & Cross, G. A. M. (2009). Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes & Development*, *23*(9), 1063-1076. <https://doi.org/10.1101/gad.1790409>
- Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S., & Cross, G. A. M. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Research*, *38*(15), 4946-4957. <https://doi.org/10.1093/nar/gkq237>

BIBLIOGRAFÍA

- Siegel, T. N., Tan, K. S. W., & Cross, G. A. M. (2005). Systematic Study of Sequence Motifs for RNA trans Splicing in *Trypanosoma brucei*. *Molecular and Cellular Biology*, 25(21), 9586-9594. <https://doi.org/10.1128/MCB.25.21.9586-9594.2005>
- Smith, M., Blanchette, M., & Papadopoulou, B. (2008). Improving the prediction of mRNA extremities in the parasitic protozoan *Leishmania*. *BMC Bioinformatics*, 9(1), 158. <https://doi.org/10.1186/1471-2105-9-158>
- Srivastava, A., Badjatia, N., Lee, J. H., Hao, B., & Günzl, A. (2018). An RNA polymerase II-associated TFIIIF-like complex is indispensable for SL RNA gene transcription in *Trypanosoma brucei*. *Nucleic Acids Research*, 46(4), 1695-1709. <https://doi.org/10.1093/nar/gkx1198>
- Staneva, D. P., Bresson, S., Auchynnikava, T., Spanos, C., Rappsilber, J., Jeyaprakash, A. A., Tollervey, D., Matthews, K. R., & Allshire, R. C. (2022). The SPARC complex defines RNAPII promoters in *Trypanosoma brucei*. *ELife*, 11. <https://doi.org/10.7554/eLife.83135>
- Staneva, D. P., Carloni, R., Auchynnikava, T., Tong, P., Rappsilber, J., Arockia Jeyaprakash, A., Matthews, K. R., & Allshire, R. C. (2021). A systematic analysis of *Trypanosoma brucei* chromatin factors identifies novel protein interaction networks associated with sites of transcription initiation and termination. *Genome Research*, 31(11), 2138-2154. <https://doi.org/10.1101/gr.275368.121>
- Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, 22(2), 96-118. <https://doi.org/10.1038/s41580-020-00315-9>
- Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M. F., Rifkin, S. A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P. E., Bussemaker, H. J., & White, K. P. (2004). A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science*, 306(5696), 655-660. <https://doi.org/10.1126/science.1101312>
- Tatomer, D. C., Elrod, N. D., Liang, D., Xiao, M.-S., Jiang, J. Z., Jonathan, M., Huang, K.-L., Wagner, E. J., Cherry, S., & Wilusz, J. E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes & Development*, 33(21-22), 1525-1538. <https://doi.org/10.1101/gad.330167.119>
- Taylor, J., & Rudenko, G. (2006). Switching trypanosome coats: what's in the wardrobe? *Trends in Genetics*, 22(11), 614-620. <https://doi.org/10.1016/j.tig.2006.08.003>
- Tetley, L., Turner, C. M., Barry, J. D., Crowe, J. S., & Vickerman, K. (1987). Onset of expression of the variant surface glycoproteins of *Trypanosoma brucei* in the tsetse fly studied using immunoelectron microscopy. *Journal of Cell Science*, 87(2), 363-372. <https://doi.org/10.1242/jcs.87.2.363>

- Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., & Libri, D. (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Molecular Cell*, *55*(3), 467-481. <https://doi.org/10.1016/j.molcel.2014.05.031>
- Uhler, J. P., Hertel, C., & Svejstrup, J. Q. (2007). A role for noncoding transcription in activation of the yeast *PHO5* gene. *Proceedings of the National Academy of Sciences*, *104*(19), 8011-8016. <https://doi.org/10.1073/pnas.0702431104>
- Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell*, *154*(1), 26-46. <https://doi.org/10.1016/j.cell.2013.06.020>
- Ullu, E., Matthews, K. R., & Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Molecular and Cellular Biology*, *13*(1), 720-725. <https://doi.org/10.1128/mcb.13.1.720-725.1993>
- van Dijk, E. L., Chen, C. L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., Nicolas, A., Thermes, C., & Morillon, A. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, *475*(7354), 114-117. <https://doi.org/10.1038/nature10118>
- van Hoof, A., Frischmeyer, P. A., Dietz, H. C., & Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science*, *295*(5563), 2262-2264. <https://doi.org/10.1126/science.1067272>
- van Luenen, H. G. A. M., Farris, C., Jan, S., Genest, P. A., Tripathi, P., Velds, A., Kerkhoven, R. M., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., van Steensel, B., Myler, P. J., & Borst, P. (2012). Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell*, *150*(5), 909-921. <https://doi.org/10.1016/j.cell.2012.07.030>
- van Roey, K., Uyar, B., Weatheritt, R. J., Dinkel, H., Seiler, M., Budd, A., Gibson, T. J., & Davey, N. E. (2014). Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chemical Reviews*, *114*(13), 6733-6778. <https://doi.org/10.1021/cr400585q>
- Vanhamme, L., & Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiological reviews*, *59*(2), 223-240. <https://doi.org/10.1128/mr.59.2.223-240.1995>
- van Werven, F. J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., & Amon, A. (2012). Transcription of two Long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell*, *150*(6), 1170-1181. <https://doi.org/10.1016/j.cell.2012.06.049>
- Vazquez, M. P., Muallem, D., Bercovich, N., Stern, M. Z., Nyambega, B., Barda, O., Masiga, D., Gupta, S. K., Michaeli, S., & Levin, M. J. (2009). Functional characterization and protein-protein interactions of trypanosome splicing factors U2AF35, U2AF65 and SF1. *Molecular and Biochemical Parasitology*, *164*(2), 137-146. <https://doi.org/10.1016/j.molbiopara.2008.12.009>

BIBLIOGRAFÍA

- Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. *Journal of Cell Science*, 5(1), 163-193. <https://doi.org/10.1242/jcs.5.1.163>
- Vigneron, A., O'Neill, M. B., Weiss, B. L., Savage, A. F., Campbell, O. C., Kamhawi, S., Valenzuela, J. G., & Aksoy, S. (2020). Single-cell RNA sequencing of *Trypanosoma brucei* from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens. *Proceedings of the National Academy of Sciences of the United States of America*, 117(5), 2613–2621. <https://doi.org/10.1073/pnas.1914423117>
- Villa, T., & Porrua, O. (2022). Pervasive transcription: a controlled risk. *FEBS Journal*. John Wiley and Sons Inc. <https://doi.org/10.1111/febs.16530>
- Wade, J. T., & Grainger, D. C. (2014). Pervasive transcription: Illuminating the dark matter of bacterial transcriptomes. *Nature Reviews Microbiology* (Vol. 12, Issue 9, pp. 647-653). Nature Publishing Group. <https://doi.org/10.1038/nrmicro3316>
- Wall, R. J., Rico, E., Lukac, I., Zuccotto, F., Elg, S., Gilbert, I. H., Freund, Y., Alley, M. R. K., Field, M. C., Wyllie, S., & Horn, D. (2018). Clinical and veterinary trypanocidal benzoxaboroles target CPSF3. *Proceedings of the National Academy of Sciences*, 115(38), 9616-9621. <https://doi.org/10.1073/pnas.1807915115>
- Wallace, F. G. (1966). The Trypanosomatid Parasites of Insects and Arachnids. *Reviews Experimental Parasitology*, (Vol. 18). [https://doi:10.1016/0014-4894\(66\)90015-4](https://doi:10.1016/0014-4894(66)90015-4)
- Wang, K. C., & Chang, H. Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell*, 43(6), 904-914. <https://doi.org/10.1016/j.molcel.2011.08.018>
- Wedel, C., Förstner, K. U., Derr, R., & Siegel, T. N. (2017). GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *The EMBO Journal*, 36(17), 2581-2594. <https://doi.org/10.15252/embj.201695323>
- Wickstead, B., Ersfeld, K., & Gull, K. (2004). The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes. *Genome Research*, 14(6), 1014-1024. <https://doi.org/10.1101/gr.2227704>
- Wippel, H. H., Malgarin, J. S., Martins, S. de T., Vidal, N. M., Marcon, B. H., Miot, H. T., Marchini, F. K., Goldenberg, S., & Alves, L. R. (2019). The nuclear RNA-binding protein RBSR1 interactome in *Trypanosoma cruzi*. *Journal of Eukaryotic Microbiology*, 66(2), 244-253. <https://doi.org/10.1111/jeu.12666>
- Wirtz, E., Leal, S., Ochatt, C., & Cross, George A. M. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 99(1), 89-101. [https://doi.org/10.1016/S0166-6851\(99\)00002-X](https://doi.org/10.1016/S0166-6851(99)00002-X)
- Wright, J. R., Siegel, T. N., & Cross, G. A. M. (2010). Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 172(2), 141-144. <https://doi.org/10.1016/j.molbiopara.2010.03.013>

- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnauld, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., & Jacquier, A. (2005). Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, *121*(5), 725-737. <https://doi.org/10.1016/j.cell.2005.04.030>
- Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W., & Steinmetz, L. M. (2011). Antisense expression increases gene expression variability and locus interdependency. *Molecular Systems Biology*, *7*(1), 468. <https://doi.org/10.1038/msb.2011.1>
- Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., & Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, *457*(7232), 1033-1037. <https://doi.org/10.1038/nature07728>
- Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*, *21*(1), 22. <https://doi.org/10.1186/s13059-020-1929-3>
- Yang, X., Figueiredo, L. M., Espinal, A., Okubo, E., & Li, B. (2009). RAP1 Is essential for silencing telomeric Variant Surface Glycoprotein genes in *Trypanosoma brucei*. *Cell*, *137*(1), 99-109. <https://doi.org/10.1016/j.cell.2009.01.037>
- Zaheed, O., Kiniry, S. J., Baranov, P. v., & Dean, K. (2021). Exploring evidence of non-coding RNA translation with Trips-Viz and GWIPS-Viz browsers. *Frontiers in Cell and Developmental Biology*, *9*. <https://doi.org/10.3389/fcell.2021.70337>