



Programa de Doctorado en Bioquímica y  
Biología Molecular  
Universidad de Granada



Instituto de Parasitología y Biomedicina  
'López-Neyra' - CSIC

## **Identificación de macromoléculas como potenciales biomarcadores útiles para el control de la enfermedad de Chagas**

Directores de la Tesis Doctoral: **Dra. María del Carmen Thomas Carazo y Dr. Manuel Carlos López López.**

**Memoria presentada por la Graduada Inmaculada María Gómez García para optar al Grado de Doctora por la Universidad de Granada.**

Febrero de 2022, Granada

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Inmaculada María Gómez García  
ISBN: 978-84-1117-890-7  
URI: <https://hdl.handle.net/10481/82470>

La doctoranda Inmaculada María Gómez García y los directores de la Tesis Dra. María del Carmen Thomas Carazo y Dr. Manuel Carlos López López:

Garantizamos, al firmar esta Tesis Doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la Tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

En Granada, a 22 de febrero de 2022

Los directores

Fdo.: María del Carmen Thomas Carazo

Fdo.: Manuel Carlos López López

Doctoranda

Fdo.: Inmaculada María Gómez García

# ÍNDICE

---

<b>RESUMEN .....</b>	<b>1</b>
<b>INTRODUCCIÓN .....</b>	<b>5</b>
<b>1. Trypanosoma cruzi Y ENFERMEDAD DE CHAGAS .....</b>	<b>7</b>
1.1. Agente etiológico de la enfermedad de Chagas .....	7
1.1.1. Clasificación de las cepas de <i>T. cruzi</i> .....	8
1.2. Ciclo de vida de <i>Trypanosoma cruzi</i> .....	10
1.3. Vías de transmisión de la enfermedad de Chagas .....	12
1.4. Epidemiología de la enfermedad de Chagas .....	15
1.5. Manifestaciones clínicas de la enfermedad de Chagas .....	17
1.5.1. Fase aguda .....	17
1.5.2. Fase crónica .....	18
1.5.2.1. Fase crónica indeterminada .....	18
1.5.2.2. Fase crónica sintomática .....	19
1.6. Diagnóstico de la enfermedad de Chagas .....	21
1.7. Tratamiento de la enfermedad de Chagas .....	22
1.7.1. Eficacia del tratamiento .....	23
1.8. Biomarcadores de seguimiento de la enfermedad de Chagas .....	25
1.8.1. Biomarcadores de eficacia terapéutica .....	25
1.8.1.1. Antígenos de <i>T. cruzi</i> KMP11, HSP70, PFR2 y 3973 <sup>d</sup> .....	27
1.8.2. Biomarcadores de patología o progresión de la enfermedad .....	30
<b>2. ELEMENTOS MÓVILES DE ADN .....</b>	<b>32</b>
2.1. Clasificación de los elementos móviles de ADN .....	32
2.1.1. Retrotransposones LTR .....	33
2.1.2. Retrotransposones no-LTR .....	33
2.2. Abundancia en los genomas de los elementos móviles de ADN .....	34
2.3. Efecto de los elementos transponibles sobre el genoma del hospedador .....	34
2.4. Elementos transponibles en tripanosomátidos: Importancia de los elementos LINE y SINE .....	35
2.5. L1Tc: Elemento LINE de <i>Trypanosoma cruzi</i> .....	37
2.5.1. NARTc: Versión truncada de L1Tc en el genoma de <i>T. cruzi</i> .....	38
2.6. Transcripción del elemento L1Tc .....	39
2.7. Función ribozima de la secuencia de ARN de Pr77. Función dual de la huella Pr77: promotor (a nivel de ADN) y ribozima (a nivel de ARN) .....	40
2.8. Ribozimas de tipo HDV .....	41
<b>3. RESPUESTA INMUNE FRENTE A LA INFECCIÓN POR <i>T. cruzi</i> .....</b>	<b>44</b>
3.1. Respuesta inmunológica innata .....	44
3.2. Respuesta inmunológica adaptativa .....	45
3.2.1. Respuesta humoral .....	45
3.2.2. Respuesta celular .....	46
3.2.2.1. Respuesta celular en fase crónica de la enfermedad de Chagas .....	48
<b>OBJETIVOS .....</b>	<b>51</b>
<b>MATERIALES Y MÉTODOS .....</b>	<b>55</b>
<b>MATERIALES .....</b>	<b>57</b>
1. Material y productos químicos .....	57
2. Tampones y soluciones .....	57
3. Anticuerpos .....	58
4. Antibióticos .....	58
5. Medios de cultivo .....	58
6. Patrones de peso molecular .....	58
7. Oligonucleótidos .....	59

8. Cepas de <i>Trypanosoma cruzi</i> .....	59
9. Líneas celulares .....	60
10. Cepas bacterianas .....	60
11. Proteínas recombinantes y péptidos .....	60
12. Población de estudio y consideraciones éticas.....	60
<b>MÉTODOS</b> .....	65
1. ANÁLISIS COMPARATIVO DEL CONTENIDO GÉNICO DE LOS GENOMAS DE DIFERENTES CEPAS DE <i>T. cruzi</i> .....	65
1.1. Cultivo de formas epimastigotes y purificación de ADN genómico .....	65
1.2. Genotipado de las cepas .....	65
1.3. Preparación de librerías y secuenciación de los genomas .....	66
1.4. Control de calidad de las lecturas crudas .....	67
1.5. Ensamblaje de los genomas y visualización de los <i>contigs</i> resultantes.....	67
1.6. Evaluación de la integridad de los ensamblajes .....	68
1.7. Anotación de los genomas .....	68
1.8. Análisis de las características de los genomas ensamblados .....	69
1.9. Identificación de secuencias de retroelementos.....	69
1.10. Identificación del contexto genómico de los retroelementos .....	71
1.11. Determinación de los sitios de inserción de las secuencias Pr77Sh e identidad con Pr77 de L1Tc/NARTc .....	71
1.12. Estimación del contenido génico en los genomas de <i>T. cruzi</i> .....	72
2. DETERMINACIÓN DE LA FUNCIÓN RIBOSOMAL (TIPO-HDV) DE LAS SECUENCIAS Pr77Sh .....	72
2.1. Obtención de células competentes de <i>E. coli</i> XL1-Blue.....	72
2.2. Transformación bacteriana por choque térmico.....	73
2.3. Selección de bacterias transformadas por el gen de la $\beta$ -galactosidasa .....	73
2.4. PCR colonia .....	73
2.5. Electroforesis de ADN en geles de agarosa .....	74
2.6. Purificación de ADN de plásmidos y secuenciación .....	74
2.7. Construcción de moldes de ADN para transcripción <i>in vitro</i> .....	74
2.8. Ensayos de corte co-transcripcional de huellas Pr77 .....	75
2.9. Cinética de corte post-transcripcional de huellas Pr77 .....	76
2.10. Electroforesis en geles de poliacrilamida-urea .....	77
2.11. Elución de ARN desde gel de poliacrilamida .....	77
3. ANÁLISIS DE EXPRESIÓN GÉNICA EN LOS DIFERENTES ESTADIOS DE <i>T. cruzi</i> MEDIANTE SECUENCIACIÓN DEL TRANSCRIPTOMA DE NUEVA GENERACIÓN (ARN-Seq) .....	78
3.1. Cultivo de formas epimastigotes.....	78
3.2. Obtención de formas tripomastigotes y amastigotes .....	78
3.3. Secuenciación del transcriptoma de nueva generación (ARN-Seq) .....	79
3.4. Análisis de datos procedentes de secuenciación ARN-Seq .....	79
3.4.1. Análisis de expresión génica diferencial de datos ARN-Seq.....	80
4. DETERMINACIÓN DEL NIVEL DE ANTICUERPOS ESPECÍFICOS DE ANTÍGENOS EN EL SUERO DE PACIENTES CON ENFERMEDAD DE CHAGAS.....	81
4.1. Aislamiento de antígenos solubles de <i>T. cruzi</i> (STcA).....	81
4.2. Electroforesis de proteínas mediante SDS-PAGE .....	81
4.3. Cuantificación de proteínas.....	82
4.4. Determinación del nivel de anticuerpos antígeno-específicos mediante la técnica ELISA ....	82
5. ANÁLISIS DE LA EXPRESIÓN GÉNICA RELATIVA EN PBMC DE PACIENTES CON ENFERMEDAD DE CHAGAS.....	84
5.1. Aislamiento de células mononucleares de sangre periférica (PBMC) .....	84
5.2. Descongelación, cultivo y estimulación de PBMC .....	85
5.3. Aislamiento de ARN, cuantificación y determinación de la calidad .....	85
5.4. Transcripción inversa.....	86

5.5. Diseño del ensayo. Selección de genes .....	86
5.6. PCR cuantitativa en tiempo real de alto rendimiento (RT-qPCR) .....	88
5.7. Procesamiento de los datos de RT-qPCR.....	89
5.7.1. Limpieza de réplicas, cálculo de promedios e intercalibrado entre placas.....	89
5.7.2. Selección de genes de normalización. Normalización de datos de RT-qPCR y aplicación de cantidades relativas .....	90
5.7.2.1. Selección de genes de normalización .....	90
5.7.2.2. Normalización de datos de RT-qPCR y cantidades relativas.....	91
5.7.3. Detección de <i>outliers</i> . Criterios de exclusión .....	91
5.8. Análisis estadístico .....	92
5.9. Obtención de valores de expresión génica diferencial y representación.....	92
5.9.1. Representación en gráfico de volcán .....	93
5.9.2. Representación en gráfico de barras .....	93
5.10. Predicción de interacciones funcionales proteína-proteína .....	94
5.11. Análisis de enriquecimiento de conjuntos de genes (GSEA) .....	94

## RESULTADOS ..... 95

<b>A. BIOMARCADORES GENÉTICOS .....</b>	<b>97</b>
A.1. SECUENCIACIÓN, ENSAMBLAJE Y ANOTACIÓN DE LOS GENOMAS DE DIFERENTES CEPAS DE <i>T. cruzi</i> . ANÁLISIS COMPARATIVO DEL CONTENIDO GÉNICO Y DE RETROELEMENTOS PORTADORES DE LA HUELLA Pr77 .....	97
A.1.1. Análisis de la integridad de los ensamblajes .....	99
A.1.2. Estimación del tamaño de los genomas.....	100
A.1.3. Anotación de los genomas .....	101
A.1.4. Determinación de las características de los genomas .....	102
A.1.5. Análisis del contenido en retroelementos .....	103
A.1.5.1. Proporción de retroelementos L1Tc y NARTc.....	105
A.1.6. Contexto genómico de los retroelementos .....	106
A.1.7. Variaciones en el contenido génico en las diferentes cepas de <i>T. cruzi</i> .....	109
A.2. LOCALIZACIÓN GENÓMICA Y PREDICCIÓN DE ESTRUCTURA SECUNDARIA DE TIPO RIBOZIMA HDV EN SECUENCIAS Pr77Sh. ESTUDIO DE SU FUNCIÓN CATALÍTICA.....	111
A.2.1. Identificación de huellas Pr77 no asociadas a retroelementos en el genoma de diferentes cepas de <i>T. cruzi</i> .....	111
A.2.2. Determinación de la localización genómica de las secuencias Pr77Sh .....	111
A.2.3. Análisis de la estructura secundaria de las secuencias Pr77Sh. Compatibilidad con la estructura ribozima tipo HDV.....	114
A.2.4. Identificación de candidatos a ribozimas tipo HDV .....	115
A.2.5. Determinación de la función ribozima de las Pr77Sh candidatas seleccionadas .....	117
A.2.5.1. Actividad co-transcripcional <i>in vitro</i> .....	117
A.2.5.2. Actividad post-transcripcional <i>in vitro</i> .....	119
A.3. ESTUDIO GENÓMICO Y DEL CONTENIDO EN RETROELEMENTOS EN LA CEPA SOL DE <i>T. cruzi</i> SECUENCIADA POR TECNOLOGÍA PACBIO .....	121
A.3.1. Secuenciación, ensamblaje y anotación del genoma.....	121
A.3.2. Estimación del contenido génico .....	123
A.3.3. Análisis del contenido en retroelementos .....	125
A.3.I. ANÁLISIS DE LA EXPRESIÓN GÉNICA DIFERENCIAL EN LAS TRES FORMAS DE <i>T. cruzi</i> EPIMASTIGOTE, TRIPOMASTIGOTE Y AMASTIGOTE MEDIANTE SECUENCIACION MASIVA DE ARN (ARN-seq).....	127
A.3.I.1. Expresión génica diferencial en tripomastigote <i>versus</i> epimastigote.....	128
A.3.I.2. Expresión génica diferencial en amastigote <i>versus</i> epimastigote .....	130
A.3.I.3. Expresión génica diferencial en tripomastigote <i>versus</i> amastigote.....	132
A.3.I.4. Diferencia en la expresión de genes particulares entre los 3 estadios morfológicos de <i>T. cruzi</i> .....	134

A.3.II EXPRESIÓN DIFERENCIAL DE RETROELEMENTOS PORTADORES DE LA HUELLA Pr77 ENTRE LAS DISTINTAS FORMAS DE <i>T. cruzi</i> (CEPA SOL) .....	137
<b>B. BIOMARCADORES DE RESPUESTA INMUNOLÓGICA</b> .....	143
<b>B.1. RESPUESTA SEROLÓGICA</b> .....	143
B.1.1. ANÁLISIS DE LA SENSIBILIDAD Y ESPECIFICIDAD DE LOS ANTÍGENOS KMP11, HSP70, PFR2 Y 3973 <sup>d</sup> DE <i>T. cruzi</i> PARA LA DETECCIÓN DE INFECCIÓN EN PACIENTES CON ENFERMEDAD DE CHAGAS .....	143
B.1.2. EVALUACIÓN DEL SET DE ANTÍGENOS KMP11, HSP70, PFR2 Y 3973 <sup>d</sup> DE <i>T. cruzi</i> COMO BIOMARCADORES DE EFICACIA TERAPÉUTICA EN PACIENTES CON ENFERMEDAD DE CHAGAS. ESTABLECIMIENTO DEL ALGORITMO PARA EVALUACIÓN DE LA EFICACIA TERAPÉUTICA.....	145
<b>B.2. RESPUESTA CELULAR</b> .....	152
B.2.1. EXPRESIÓN DIFERENCIAL DE GENES DE RESPUESTA INMUNE EN PACIENTES ASINTOMÁTICOS CON ENFERMEDAD DE CHAGAS <i>VERSUS</i> SUJETOS SANOS .....	152
B.2.1.1. Análisis comparativo del perfil de expresión génica entre individuos sanos.....	153
B.2.1.2. Identificación de genes con expresión diferencial entre pacientes con Chagas crónico en fase indeterminada (IND) y sujetos sanos (DS) .....	156
B.2.1.3. Análisis de los genes expresados diferencialmente entre sujetos sanos provenientes de áreas endémicas y no endémicas de la enfermedad de Chagas.....	164
B.2.1.4. Identificación de las vías inmunológicas implicadas en el establecimiento de la infección por <i>T. cruzi</i> en pacientes con enfermedad de Chagas crónico en fase indeterminada (IND).....	166
B.2.1.5. Análisis de Enriquecimiento de Conjuntos de Genes (GSEA).....	170
B.2.2. EXPRESIÓN DIFERENCIAL DE GENES DE RESPUESTA INMUNE EN PACIENTES CON ENFERMEDAD DE CHAGAS EN FASE CARDIACA <i>VERSUS</i> PACIENTES ASINTOMÁTICOS.....	174
B.2.3. EXPRESIÓN DIFERENCIAL DE GENES DE RESPUESTA INMUNE EN PACIENTES CON ENFERMEDAD DE CHAGAS POR EFECTO DEL TRATAMIENTO CON BENZNIDAZOL .....	178
B.2.3.1. Identificación de genes con expresión diferencial entre pacientes crónicos en fase indeterminada (IND) pre y post-tratamiento.....	179
B.2.3.2. Identificación de genes con expresión diferencial entre pacientes crónicos en fase cardiaca (CCC) pre y post-tratamiento.....	182
<b>DISCUSIÓN</b> .....	187
1. VARIACIONES EN EL CONTENIDO GÉNICO Y CONTENIDO DE RETROELEMENTOS PORTADORES DE LA HUELLA Pr77. ANÁLISIS COMPARATIVO ENTRE 6 CEPAS DE <i>T. cruzi</i> DE DIFERENTE DTU.....	189
2. CONTEXTO GENÓMICO Y PREDICCIÓN DE ESTRUCTURA SECUNDARIA DE TIPO RIBOZIMA HDV EN SECUENCIAS Pr77sh. ESTUDIO DE SU FUNCIÓN CATALÍTICA.....	194
3. ESTUDIO DEL CONTENIDO GÉNICO Y EN RETROELEMENTOS DE LA CEPA SOL DE <i>T. cruzi</i> SECUENCIADA POR TECNOLOGÍA PACBIO. ANÁLISIS DE EXPRESIÓN GÉNICA DIFERENCIAL ENTRE LAS 3 FORMAS DEL PARÁSITO .....	199
4. ANÁLISIS DE LA SENSIBILIDAD Y ESPECIFICIDAD DE KMP11, HSP70, PFR2 Y 3973 <sup>d</sup> DE <i>T. cruzi</i> EN PACIENTES CON ENFERMEDAD DE CHAGAS .....	202
5. KMP11, HSP70, PFR2 Y 3973 <sup>d</sup> DE <i>T. cruzi</i> COMO BIOMARCADORES DE EFICACIA TERAPÉUTICA EN PACIENTES CON ENFERMEDAD DE CHAGAS .....	203
6. DETERMINACIÓN DE LA EXPRESIÓN DIFERENCIAL DE 106 GENES IMPLICADOS EN LA RESPUESTA INMUNE EN PACIENTES CRÓNICOS CON ENFERMEDAD DE CHAGAS .....	206
6.1. Expresión diferencial de genes implicados en la respuesta inmune en pacientes en fase indeterminada <i>versus</i> sujetos sanos.....	206
6.2. Expresión diferencial de genes implicados en la respuesta inmune en pacientes de Chagas en fase cardiaca <i>versus</i> fase indeterminada, previamente y tras el tratamiento con benznidazol.....	213



<b>CONCLUSIONES</b> .....	<b>215</b>
<b>BIBLIOGRAFÍA</b> .....	<b>219</b>
<b>ANEXOS</b> .....	<b>253</b>
ANEXO DE PUBLICACIONES.....	259

## ÍNDICE DE FIGURAS

### INTRODUCCIÓN

<b>Figura I1.</b> Distribución geográfica aproximada de las DTU de <i>T. cruzi</i> en los ciclos de transmisión domésticos y selváticos .....	<b>10</b>
<b>Figura I2.</b> Formas de <i>T. cruzi</i> : epimastigote, tripomastigote y amastigote .....	<b>10</b>
<b>Figura I3.</b> Ciclo de vida de <i>T. cruzi</i> .....	<b>12</b>
<b>Figura I4.</b> Distribución de las especies de triatomíneos de mayor relevancia epidemiológica ....	<b>13</b>
<b>Figura I5.</b> Distribución global de los casos de la enfermedad de Chagas, basado en estimaciones oficiales 2006-2015 .....	<b>17</b>
<b>Figura I6.</b> Manifestaciones clínicas de la fase aguda de la enfermedad de Chagas .....	<b>18</b>
<b>Figura I7.</b> Estructura química de los fármacos empleados para tratar la enfermedad de Chagas: Benznidazol y Nifurtimox.....	<b>23</b>
<b>Figura I8.</b> Representación esquemática del retroelemento L1Tc de <i>T. cruzi</i> .....	<b>38</b>
<b>Figura I9.</b> Representación esquemática y comparación de los retrotransposones L1Tc y NARTc de <i>T. cruzi</i> .....	<b>39</b>
<b>Figura I10.</b> Mecanismo de la reacción de transesterificación del ARN catalizado por las ribozimas HDV autocatalíticas .....	<b>42</b>
<b>Figura I11.</b> Estructura secundaria consenso y cristalina de ribozimas tipo-HDV .....	<b>42</b>

### MATERIAL Y MÉTODOS

<b>Figura M1.</b> Fases obtenidas tras la centrifugación de la muestra de sangre periférica con la solución Lymphoprep™ durante la técnica de aislamiento de células mononucleares de sangre periférica (PBMC).....	<b>84</b>
<b>Figura M2.</b> Diseño del ‘ <i>TaqMan</i> ® <i>OpenArray</i> ® <i>Real-Time PCR plate</i> ’ .....	<b>88</b>
<b>Figura M3.</b> Representación esquemática de los criterios establecidos en el algoritmo empleado para la limpieza de réplicas en <i>Visual Basic</i> .....	<b>90</b>

### RESULTADOS

<b>Figura R1.</b> Esquema que describe los procesos de secuenciación, ensamblaje y anotación llevados a cabo en los genomas de las diferentes cepas de <i>T. cruzi</i> descritas en este estudio ..	<b>98</b>
<b>Figura R2.</b> Análisis BUSCO de la integridad del ensamblaje del genoma .....	<b>100</b>

<b>Figura R3.</b> Estructura secundaria de ribozimas seleccionadas teóricamente activas según su plegamiento .....	<b>116</b>
<b>Figura R4.</b> Representación esquemática de la posición genómica de las <i>single Pr77 hallmarks</i> seleccionadas.....	<b>117</b>
<b>Figura R5.</b> Corte co-transcripcional de huellas Pr77Sh de <i>T. cruzi</i> a diferentes temperaturas.	<b>118</b>
<b>Figura R6.</b> Corte post-transcripcional de las huellas Pr77Sh IK-51 y BM-1 identificadas en el genoma de <i>T. cruzi</i> .....	<b>120</b>
<b>Figura R7.</b> Análisis BUSCO de la integridad del ensamblaje del genoma .....	<b>122</b>
<b>Figura R8.</b> Análisis de expresión génica diferencial entre los estadios epimastigote (Epi), tripomastigote (Trypo) y amastigote (Ama) de <i>T. cruzi</i> .....	<b>128</b>
<b>Figura R9.</b> Representación en gráfico de volcán de los transcritos expresados diferencialmente entre los estadios tripomastigote y epimastigote de <i>T. cruzi</i> determinado por ARN-seq .....	<b>129</b>
<b>Figura R10.</b> Representación en gráfico de volcán de los transcritos expresados diferencialmente entre los estadios amastigote y epimastigote de <i>T. cruzi</i> determinado por ARN-seq.....	<b>131</b>
<b>Figura R11.</b> Representación en gráfico de volcán de los transcritos expresados diferencialmente entre los estadios tripomastigote y amastigote de <i>T. cruzi</i> determinado por ARN-seq .....	<b>133</b>
<b>Figura R12.</b> Análisis comparativo del nivel de expresión de los genes indicados en las formas de <i>T. cruzi</i> epimastigote (Epi), tripomastigote (Tripo) y amastigote (Ama) .....	<b>136</b>
<b>Figura R13.</b> Representación en gráfico de volcán de los retroelementos expresados diferencialmente entre los estadios (A) tripomastigote <i>versus</i> epimastigote, (B) amastigote <i>versus</i> epimastigote y (C) tripomastigote <i>versus</i> amastigote determinado por ARN-seq.....	<b>139</b>
<b>Figura R14.</b> Análisis comparativo del nivel de expresión de retroelementos identificados en el genoma de SOL en las 3 formas de <i>T. cruzi</i> epimastigote (Epi), tripomastigote (Tripo) y amastigote (Ama).....	<b>141</b>
<b>Figura R15.</b> Análisis comparativo del nivel de expresión de retroelementos identificados en el genoma de SOL en el estadio epimastigote (A), tripomastigote (B) y amastigote (C).....	<b>142</b>
<b>Figura R16.</b> Diagramas de cajas que muestran la reactividad de los sueros de sujetos infectados con <i>T. cruzi</i> y sujetos sanos (valores de densidad óptica, representados en escala logarítmica) frente a cada uno de los antígenos.....	<b>144</b>
<b>Figura R17.</b> Reactividad de los sueros de pacientes asintomáticos de la enfermedad de Chagas frente a antígenos de <i>T. cruzi</i> antes y tras el tratamiento con benznidazol.....	<b>146</b>
<b>Figura R18.</b> Frecuencia de pacientes con enfermedad de Chagas que mostraron una disminución de los niveles de anticuerpos frente a los cuatro antígenos de <i>T. cruzi</i> evaluados después del tratamiento .....	<b>147</b>
<b>Figura R19.</b> Número y porcentaje de pacientes con enfermedad de Chagas cuya reactividad frente a cada una de las cuatro moléculas cumplía con el criterio STEC a los 24 y 48 meses tras el tratamiento.....	<b>148</b>
<b>Figura R20.</b> Representación en forma de gráfico circular del porcentaje de pacientes con enfermedad de Chagas que cumplieron o no el criterio de eficacia terapéutica establecido .	<b>150</b>

<b>Figura R21.</b> Gráfico de puntuación PCA para las diferencias en la reactividad (varianza) de 66 pacientes en fase indeterminada frente a cada uno de los cuatro biomarcadores (4 variables) entre el tiempo previo al tratamiento y tras 24 meses de tratamiento .....	<b>151</b>
<b>Figura R22.</b> Mapa de calor del nivel de expresión génica de 106 genes implicados en la respuesta inmunológica en sujetos sanos de áreas no endémicas (DSc) y endémicas (DSe) de la enfermedad de Chagas .....	<b>154</b>
<b>Figura R23.</b> Análisis comparativo de la expresión génica de 106 genes en donantes sanos procedentes de áreas no endémicas (DSc, n=20) y endémicas (DSe, n=10) de la enfermedad de Chagas.....	<b>155</b>
<b>Figura R24.</b> Mapa de calor de los 100 genes principales determinado por análisis GSEA en pacientes en fase indeterminada (IND) y sujetos sanos (DS).....	<b>157</b>
<b>Figura R25.</b> Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes en fase indeterminada y donantes sanos .....	<b>159</b>
<b>Figura R26.</b> Análisis de componentes principales sobre los valores de cantidades relativas normalizadas de expresión génica de 106 genes analizados en muestras correspondientes a (A) mezclas de pacientes IND y pacientes IND individuales y en (B) DSc, DSe, mezclas de pacientes IND y pacientes IND individuales .....	<b>162</b>
<b>Figura R27.</b> Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes con enfermedad de Chagas en fase crónica indeterminada (IND, n=39) y donantes sanos (DS, n=30) representado en gráfico de volcán.....	<b>163</b>
<b>Figura R28.</b> Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes indeterminados (IND) y donantes sanos (DS) .....	<b>164</b>
<b>Figura R29.</b> Análisis comparativo del nivel de expresión de los genes <i>FCER1A</i> , <i>IL12B</i> , <i>IL6</i> , <i>IL17A</i> y <i>CD27</i> medidos como cantidades relativas normalizadas (NRQ) promedio en pacientes con enfermedad de Chagas crónica en fase indeterminada (IND) y donantes sanos de áreas endémicas (DSe) y no endémicas (DSc) de la enfermedad de Chagas.....	<b>165</b>
<b>Figura R30.</b> Análisis STRING de interacción proteína-proteína (IPP) .....	<b>167</b>
<b>Figura R31.</b> Subredes STRING de interacción proteína-proteína (IPP) .....	<b>169</b>
<b>Figura R32.</b> Gráficos de enriquecimiento procedentes del análisis de enriquecimiento de conjuntos de genes (GSEA) .....	<b>171</b>
<b>Figura R33.</b> Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes en fase indeterminada y pacientes en fase cardiaca .....	<b>175</b>
<b>Figura R34.</b> Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes con enfermedad de Chagas en fase cardiaca (CCC, n=18) y en fase indeterminada (IND, n=39) representado en gráfico de volcán.....	<b>177</b>
<b>Figura R35.</b> Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes con enfermedad de Chagas en fase cardiaca (CCC) y en fase indeterminada (IND) .....	<b>178</b>
<b>Figura R36.</b> Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes en fase indeterminada pre-tratamiento y pacientes en fase indeterminada post-tratamiento.....	<b>179</b>

<b>Figura R37.</b> Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes con enfermedad de Chagas en fase indeterminada pre-tratamiento (IND-PRE, n=39) y pacientes en fase indeterminada post-tratamiento (IND-POS, n=46) representado en gráfico de volcán .....	<b>181</b>
<b>Figura R38.</b> Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes con enfermedad de Chagas en fase indeterminada pre-tratamiento (IND-PRE) y pacientes en fase indeterminada post-tratamiento (IND-POS) .....	<b>182</b>
<b>Figura R39.</b> Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes crónicos en fase cardíaca pre-tratamiento y pacientes en fase cardíaca post-tratamiento .....	<b>183</b>
<b>Figura R40.</b> Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes crónicos con enfermedad de Chagas en fase cardíaca pre-tratamiento (CCC-PRE, n=18) y pacientes en fase cardíaca post-tratamiento (CCC-POS, n=21) representado en gráfico de volcán .....	<b>184</b>
<b>Figura R41.</b> Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes crónicos con enfermedad de Chagas en fase cardíaca pre-tratamiento (CCC-PRE) y pacientes en fase cardíaca post-tratamiento (CCC-POS) .....	<b>185</b>

## DISCUSIÓN

<b>Figura D1.</b> Nivel de anticuerpos por test serológicos para la evaluación de la eficacia del tratamiento en la enfermedad de Chagas .....	<b>205</b>
--	------------

# ÍNDICE DE TABLAS

## INTRODUCCIÓN

<b>Tabla I1.</b> Clasificación según Kuschnir de la afectación cardíaca en la fase crónica de la enfermedad de Chagas .....	<b>20</b>
---	-----------

## MATERIAL Y MÉTODOS

<b>Tabla M1.</b> Cohorte de pacientes empleada para el análisis de la sensibilidad y especificidad de los antígenos KMP11, HSP70, PFR2 y 3973 <sup>d</sup> de <i>T. cruzi</i> en pacientes con enfermedad de Chagas .....	<b>61</b>
<b>Tabla M2.</b> Cohorte de pacientes empleada para el estudio de la expresión diferencial de genes de respuesta inmune en pacientes con enfermedad de Chagas mediante OpenArrays® .....	<b>63</b>
<b>Tabla M3.</b> Genes a estudio (n=112), identidad y papel funcional.....	<b>87</b>

## RESULTADOS

<b>Tabla R1.</b> DTU, origen geográfico y hospedador/vector de las cepas de <i>T. cruzi</i> analizadas en este estudio .....	<b>97</b>
--	-----------

<b>Tabla R2.</b> Resumen de métricas de secuenciación y ensamblaje de genomas de las cepas B. M. López, Dm28, Y, Ikiakarora, SOL y CL Brener de <i>T. cruzi</i> .....	<b>99</b>
<b>Tabla R3.</b> Estimación del tamaño del genoma de las cepas de <i>T. cruzi</i> .....	<b>101</b>
<b>Tabla R4.</b> Resumen del análisis de <i>Best Hit</i> .....	<b>102</b>
<b>Tabla R5.</b> Características de los genomas ensamblados y genes anotados en las cepas de <i>T. cruzi</i> a estudio .....	<b>103</b>
<b>Tabla R6.</b> Comparación del contenido en elementos transponibles.....	<b>104</b>
<b>Tabla R7.</b> Abundancia relativa de retroelementos identificados por (a) huella Pr77 más 40 nt ubicados corriente abajo de Pr77, (b) extremos 3' (55 últimos nt de L1Tc o NARTc) o (c) promedio de elementos identificados tanto por el extremo 5' como 3' .....	<b>105</b>
<b>Tabla R8.</b> Contexto genómico donde se insertan los retroelementos identificados en las diferentes cepas a estudio .....	<b>108</b>
<b>Tabla R9.</b> Comparación del contenido génico entre cepas .....	<b>110</b>
<b>Tabla R10.</b> Número y sitios de inserción de <i>single Pr77 hallmarks</i> (Pr77Sh) en el genoma de las seis cepas de <i>T. cruzi</i> a estudio .....	<b>113</b>
<b>Tabla R11.</b> Resumen de métricas de secuenciación y ensamblaje del genoma de la cepa SOL de <i>T. cruzi</i> secuenciado por tecnología PacBio .....	<b>121</b>
<b>Tabla R12.</b> Características del genoma ensamblado y genes anotados en la cepa SOL de <i>T. cruzi</i> secuenciada por tecnología PacBio.....	<b>123</b>
<b>Tabla R13.</b> Contenido génico del genoma de SOL de <i>T. cruzi</i> (PacBio).....	<b>125</b>
<b>Tabla R14.</b> Secuencias Pr77, L1Tc y NARTc identificadas en el genoma de la cepa SOL de <i>T. cruzi</i> (PacBio) mediante análisis BLASTN .....	<b>126</b>
<b>Tabla R15.</b> Secuencias DIREs identificadas en el genoma de la cepa SOL de <i>T. cruzi</i> (PacBio) mediante análisis BLASTN .....	<b>126</b>
<b>Tabla R16.</b> Número de genes expresados diferencialmente (GED) en la forma tripomastigote de <i>T. cruzi</i> versus epimastigote determinado por ARN-seq .....	<b>130</b>
<b>Tabla R17.</b> Número de genes expresados diferencialmente (GED) en la forma amastigote de <i>T. cruzi</i> versus epimastigote determinado por ARN-seq.....	<b>132</b>
<b>Tabla R18.</b> Número de genes expresados diferencialmente (GED) en la forma tripomastigote de <i>T. cruzi</i> versus amastigote determinado por ARN-seq .....	<b>134</b>
<b>Tabla R19.</b> Retroelementos portadores de Pr77 expresados diferencialmente en las tres formas del parásito identificados en el genoma de SOL.....	<b>140</b>
<b>Tabla R20.</b> Resumen de la clasificación de cada antígeno evaluado .....	<b>144</b>
<b>Tabla R21.</b> Porcentaje de pacientes con enfermedad de Chagas que presentaban una disminución en el nivel de anticuerpo frente a los antígenos de <i>T. cruzi</i> tras el tratamiento con benznidazol .....	<b>147</b>
<b>Tabla R22.</b> Pacientes con enfermedad de Chagas que cumplen con el criterio de eficacia terapéutica establecido .....	<b>149</b>

<b>Tabla R23.</b> Detección de <i>T. cruzi</i> por PCR en pacientes en fase indeterminada antes y tras el tratamiento con benznidazol .....	<b>151</b>
<b>Tabla R24.</b> Genes con carga factorial para la componente principal 1 (PC1) superior a 0,6 o inferior a -0,6 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes en fase indeterminada y donantes sanos.....	<b>160</b>
<b>Tabla R25.</b> Principales rutas biológicas en las que participan las proteínas de la red de STRING codificadas por los genes expresados diferencialmente (GED) entre pacientes en fase indeterminada y donantes sanos.....	<b>168</b>
<b>Tabla R26.</b> Lista de genes expresados diferencialmente entre pacientes en fase indeterminada (IND) y donantes sanos, enriquecidos en el fenotipo IND según el análisis GSEA, que participan en las rutas de BioCarta: activación de células B dependientes de antígeno (BIOCARTA_ASBCCELL_PATHWAY), inducción de la regulación de las HSP por el estrés (BIOCARTA_HSP27_PATHWAY), citoquinas y respuesta inflamatoria (BIOCARTA_INFLAM_PATHWAY) e IL-12 dependiente de NO2 en células NK (BIOCARTA_NO2IL12_PATHWAY).....	<b>173</b>
<b>Tabla R27.</b> Comparación de los valores de expresión génica diferencial obtenidos para los genes implicados en las vías enriquecidas predichas por GSEA entre todas las muestras IND y sólo las muestras IND de pacientes individuales frente a donantes sanos (DS).....	<b>174</b>
<b>Tabla R28.</b> Genes con carga factorial para la componente principal 2 (PC2) y componente principal 4 (PC4) superior a 0,54 o inferior a -0,54 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes en fase indeterminada y cardíaca.....	<b>176</b>
<b>Tabla R29.</b> Genes con carga factorial para la componente principal 1 (PC1) superior a 0,54 o inferior a -0,54 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes crónicos en fase indeterminada pre y post-tratamiento .....	<b>180</b>
<b>Tabla R30.</b> Genes con carga factorial para la componente principal 2 (PC2) superior a 0,54 o inferior a -0,54 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes crónicos en fase cardíaca pre y post-tratamiento .....	<b>183</b>

## ABREVIATURAS

ADN	Ácido desoxirribonucleico
ADNc	ADN complementario
ADNk	ADN del kinetoplasto
AP	<i>Apurínico / Apirimidínico</i>
ARN	Ácido ribonucleico
ARNm	ARN mensajero
ARNr	ARN ribosómico
ARNt	ARN transferente
BLAST	<i>Basic local alignment search tool</i>
BMK	Biomarcador
CCC	Forma crónica cardiaca de la enfermedad de Chagas
Cq	<i>Quantification cycle</i>
Ct	Ciclo umbral de la PCR cuantitativa
DEPC	Dietilpirocarbonato
DIRE	<i>Degenerate ingi/L1Tc-related elements</i>
DMSO	Dimetil sulfóxido
dNTP	Desoxirribonucleótidos trifosfato
DPE	<i>Downstream promoter elements motif</i>
DS	Donantes sanos
DSc	Donantes sanos procedentes de zonas no endémicas
DSe	Donantes sanos procedentes de zonas endémicas
DTT	Ditiotreitol
DTU	<i>Discrete typing unit</i>
D.O.	Densidad óptica
ECG	Electrocardiograma
EDTA	Ácido etilendiaminotetraacético
FC	<i>Fold change</i>
g, mg, µg	Gramos, miligramos, microgramos
GSEA	Análisis de enriquecimiento de conjuntos de genes
h, min, seg	Horas, minutos, segundos

HDV	Virus de la hepatitis delta
Ig	Inmunoglobulina
IL	Interleuquina
IND	Forma crónica indeterminada de la enfermedad de Chagas
IPP	Interacción proteína-proteína
kb	Kilobase
kDa	Kilodalton
kHz	Kilohercios
L, ml, µl, nl	Litro, mililitro, microlitro, nanolitro
LINE	<i>Long Interspersed Nuclear Element</i>
LTR	Repeticiones largas terminales ( <i>Long Terminal Repeats</i> )
Log	Logaritmo
L1Tc	<i>Long Interspersed Nucleotide Element from T. cruzi</i>
M, mM	Molar, milimolar
mg, ng, µg	Miligramo, nanogramo, microgramo
NARTc	<i>Non-Autonomous Retrotransposon in T. cruzi</i>
NK	<i>Natural killer</i>
NO	Óxido nítrico
NO <sub>2</sub>	Dióxido de nitrógeno
NP-40	Nonidet P-40
NRQ	Cantidad relativa normalizada
nt	Nucleótido
ODP	O-fenilendiamina
OMS	Organización Mundial de la Salud
ORF	Marco abierto de lectura ( <i>Open Reading Frame</i> ).
PacBio	<i>Pacific Biosciences</i>
PAGE	Electroforesis en gel de poliacrilamida
pb	Pares de bases
PBMC	Células mononucleares de sangre periférica
PBS	Tampón fosfato salino
PCA	Análisis de componentes principales
PCR	Reacción en cadena de la polimerasa



PMSF	Fluoruro de fenilmetilsulfonilo
Q1, Q2, Q3	Cuartil 1, Cuartil 2, Cuartil 3
qPCR	PCR cuantitativa
RI	Rango intercuartílico
RNasa	Ribonucleasa
RPKM	Lecturas por kilobase por millón de lecturas mapeadas
rpm	Revoluciones por minuto
RT	Transcriptasa inversa
RTEC	<i>Restrictive therapeutic efficacy criteria</i>
Rz	Ribozima
SFBI	Suero fetal bovino inactivado
SDS	Dodecil sulfato sódico
Se	Sensibilidad
SINE	<i>Short Interspersed Nuclear Elements</i>
Sp	Especificidad
SSD	Dodecilsulfato sódico
STcA	Antígenos solubles de <i>Trypanosoma cruzi</i>
STEC	<i>Standard therapeutic efficacy criteria</i>
TEMED	N-N-N'-N'-tetrametil etilén diamina
Th1	Linfocitos T cooperadores de tipo 1 ( <i>T helper cells Type 1</i> )
Th2	Linfocitos T cooperadores de tipo 2 ( <i>T helper cells Type 2</i> )
Th17	Linfocitos T cooperadores de tipo 17 ( <i>T helper cells Type 17</i> )
Tris	Trihidroximetil aminometano
U	Unidad enzimática
UTR	Región no traducida ( <i>Untranslated region</i> )
°C	Grados Celsius



## RESUMEN

La infección por el parásito *Trypanosoma cruzi* causa la enfermedad de Chagas, considerada una de las enfermedades tropicales desatendidas más prevalentes, afectando aproximadamente a 6-7 millones de personas en el mundo, estando presente en zonas no endémicas debido a los flujos migratorios. La enfermedad, tras la fase aguda, evoluciona hacia una fase crónica asintomática, clínicamente silente, llamada fase indeterminada que puede durar décadas y en la que existe un frágil equilibrio entre la replicación del parásito y la respuesta inmunitaria del huésped. Así, el 30-40% de los pacientes crónicos desarrollan una fase sintomática que se caracteriza, principalmente, por alteraciones cardíacas y digestivas. El tratamiento se basa en la eliminación del parásito en sangre, siendo los fármacos de elección el benznidazol y nifurtimox. Existe controversia sobre su eficacia en la fase crónica y si su efectividad disminuye en la fase sintomática. Para el diagnóstico de la enfermedad de Chagas, se aplican técnicas serológicas convencionales. Sin embargo, éstas no son útiles para evaluar el status clínico del paciente ni el impacto del tratamiento. El diagnóstico por PCR se ve dificultado por la escasa presencia de parásitos en sangre y fluctuaciones en la parasitemia por lo que se genera un alto porcentaje de falsos negativos.

La infección por *T. cruzi* desencadena múltiples mecanismos inmunitarios de respuesta innata y adaptativa en el huésped para combatir el patógeno. Dado que el parásito es intracelular y se replica dentro de las células, la respuesta celular de la inmunidad adaptativa del huésped desempeña un papel fundamental. Esta función está orquestada principalmente por los linfocitos T, que reconocen los antígenos del parásito y promueven funciones específicas para controlar la infección. La alta variabilidad existente entre las cepas del parásito *T. cruzi*, que afecta a su biología y, consecuentemente, a su virulencia, tropismo, etc., dificulta aún más su control. Los genomas de tripanosomátidos están colonizados por retrotransposones sin LTR que presentan una secuencia de 77 nt altamente conservada en sus extremos 5', conocida como huella Pr77. La amplia distribución de la huella Pr77 podría estar relacionada con los procesos de regulación génica en estos organismos, ya que esta presenta actividad promotora y ribozima, similar a las del virus de la hepatitis delta (HDV), a nivel de ADN y ARN, respectivamente.

Sobre estas bases, en la presente Tesis se marcó como un primer objetivo la identificación de los retrotransposones portadores de la huella Pr77 y de material genético de interés contenido en los genomas de seis cepas de *T. cruzi* que pertenecen a distintos DTU y provienen de diferentes orígenes geográficos y hospedadores/vectores. Los resultados mostraron diferencias entre los seis genomas en el número de copias de genes implicados en procesos de virulencia, en la abundancia de retrotransposones que portan la huella Pr77, así como la presencia de huellas Pr77 no asociados a retroelementos (Pr77Sh). También, se observaron

asociaciones frecuentes de retrotransposones portadores de Pr77 y de Pr77Sh con genes que codifican para proteínas relevantes desde el punto de vista de la biología del parásito, tales como trans-sialidasas, RHS, MASP, mucinas, etc., mostrando una proporción variable según el tipo de retroelemento, la clase de gen y la cepa del parásito.

Dada la función dual promotor/ribozima descrita para Pr77 de L1Tc, otro de los objetivos de esta Tesis fue la caracterización de las mencionadas huellas Pr77 no asociadas a retroelementos. Los resultados obtenidos mostraron que en la mayoría de las cepas a estudio existían huellas Pr77Sh que mantienen conservado el motivo DPE (*Downstream Promoter Element*) en términos de composición de la secuencia y localización respecto al sitio de inicio de la transcripción. Interesantemente, ensayos *in vitro* de corte co-transcripcional y post-transcripcional mostraron la existencia de actividad ribozima en algunas de las huellas Pr77Sh estudiadas. En cuanto a su sitio de inserción, se observó que se encontraban localizadas en diferentes puntos del genoma y frecuentemente asociadas a genes relevantes para la biología del parásito. Todo ello sugiere un importante papel funcional de estas secuencias en la regulación de la expresión génica de este parásito. Por otro lado, el genoma de la cepa SOL de *T. cruzi*, fue secuenciado por tecnología PacBio, formando parte de esta Tesis la determinación del contenido de retroelementos portadores de Pr77 y genes de interés identificados en dicha cepa, así como la expresión diferencial existente entre las formas epimastigotas, amastigotas y tripomastigotas del parásito.

Por otra parte, y sobre la base de los antecedentes expuestos, la identificación de marcadores de progresión de la enfermedad, así como de eficacia terapéutica resulta esencial para el adecuado diagnóstico y seguimiento de los pacientes de Chagas en su fase crónica. En este contexto, se propuso como un segundo objetivo de la presente Tesis evaluar varias moléculas como biomarcadores serológicos de diagnóstico de la enfermedad y eficacia del tratamiento farmacológico. Los resultados obtenidos mostraron una alta sensibilidad y especificidad para las moléculas recombinantes KMP11, PFR2 y para el péptido sintético 3973<sup>d</sup>. Asimismo, se determinó que el set formado por los antígenos de *T. cruzi* KMP11, HSP70, PFR2 y 3973<sup>d</sup> era una herramienta útil como marcador temprano de eficacia del tratamiento con benznidazol en los pacientes crónicos de Chagas.

Además, se planteó como un tercer objetivo de la presente Tesis el estudio de los patrones de expresión de genes implicados en la respuesta del sistema inmunitario humano a la infección por *T. cruzi*, mediante tecnología OpenArray®, en pacientes crónicos con enfermedad de Chagas en sus distintas fases clínicas y tras el tratamiento con benznidazol. Los resultados obtenidos revelaron la existencia de un perfil de expresión diferencial, específica de antígeno, entre los pacientes en fase indeterminada y sujetos sanos, identificándose tanto la activación como la depleción significativa de numerosos genes en dichos pacientes *versus* sujetos sanos, así como la activación de varias relevantes rutas de respuesta inmunitaria. Del mismo modo, se observaron patrones diferenciales de expresión entre los pacientes crónicos en

fase indeterminada y aquellos en fase cardiaca en genes implicados en la respuesta inmune celular antígeno-específica. Interesantemente, también hemos puesto de manifiesto que el tratamiento con benznidazol modula el patrón de expresión de dichos genes, tanto en pacientes asintomáticos como con sintomatología cardiaca. Estos perfiles de expresión génica diferencial detectados en los pacientes crónicos de Chagas, así como las rutas inmunológicas que parecen estar activadas o inhibidas, podrían dar lugar a la identificación de potenciales biomarcadores de progresión de la enfermedad y/o de eficacia del tratamiento, así como actuar como posibles dianas terapéuticas útiles para el control inmuno-terapéutico de la enfermedad crónica de Chagas.



# **INTRODUCCIÓN**

---





## 1. *Trypanosoma cruzi* Y ENFERMEDAD DE CHAGAS

La enfermedad de Chagas, también llamada tripanosomiasis americana, fue descubierta por Carlos Ribeiro Justiniano Chagas en 1909. Se clasifica dentro de las enfermedades olvidadas o relacionadas con la pobreza, siendo, además, dentro de estas una de las más desatendidas. Constituye una de las enfermedades parasitarias más importantes del continente americano, y a día de hoy, debido a los movimientos migratorios, un grave problema de salud mundial (Rassi et al., 2012).

### 1.1. Agente etiológico de la enfermedad de Chagas

El agente causal de la enfermedad de Chagas es *Trypanosoma cruzi*, un parásito protozoo flagelado intracelular perteneciente al dominio Eukaryota, reino Protista, filo *Sarcomastigophora*, subfilo *Mastigophora*, clase *Zoomastigophora*, orden *Kinetoplastida* y familia *Trypanosomatidae* (Rassi et al., 2012).

El orden *Kinetoplastida* se caracteriza por presentar una única mitocondria de gran tamaño llamada kinetoplasto y cuyo ADN (denominado ADNk) consiste en miles de copias de minicírculos y algunas copias de maxicírculos interconectados topológicamente constituyendo una red (Simpson, 1987). Los minicírculos (0,5-2,5 kb) codifican pequeños ácidos ribonucleicos (ARNs) guías que modifican los transcritos de los maxicírculos por edición del ARN y de los que pueden hallarse hasta 30.000 copias por parásito. Los maxicírculos (20-40 kb) son menos abundantes y análogos estructurales y funcionalmente al ADN mitocondrial de eucariotas superiores que codifica los ARN ribosómicos y las subunidades de los complejos respiratorios (de Souza, 2009). El ADNk constituye del 10% al 30% del ADN total, en función de la especie, y su replicación se inicia casi simultáneamente a la replicación del ADN nuclear lo cual sugiere que las fases de síntesis de ADN (S) en el núcleo y en la mitocondria están reguladas coordinadamente (Pasion et al., 1994). Todos los kinetoplastidos examinados comparten otras dos características únicas: compartimentalización de la glucólisis dentro del glicosoma (Michels and Hannaert, 1994) y *trans-splicing* de una pequeña secuencia de ARN conservada entre todos los ARN mensajeros, denominada *spliced leader* (SL o mini-exón) que se añade al extremo 5' de todos los mensajeros para su maduración y traducción (Campbell et al., 1984).

### 1.1.1. Clasificación de las cepas de *T. cruzi*

Debido a la alta heterogeneidad genotípica y fenotípica presente en la población de *T. cruzi*, las diferentes cepas se han subdividido en seis DTUs (*Discrete Typing Units*) definidas como TcI a TcVI, a la que posteriormente se añadió Tcbat (Lima et al., 2015; Zingales et al., 2012). Una DTU se define como un conjunto de cepas que comparten características en su genotipo y que pueden ser identificadas por marcadores genéticos, moleculares e inmunológicos comunes. Los parásitos de las diferentes DTUs presentan ciclos selváticos o domésticos mostrando diferencias en su distribución entre las especies de insecto vector y hospedador mamífero, así como en su localización geográfica (Brenière et al., 2016; Zingales et al., 2012).

Independientemente del DTU al que pertenezca la cepa, todos los parásitos de *T. cruzi* pueden causar la enfermedad de Chagas. Los parásitos de distintas DTU muestran diferencias notables a nivel celular y molecular, como en el tropismo de tejidos y órganos, antigenicidad, grado de virulencia, susceptibilidad a fármacos, capacidad para infectar vectores, número de cromosomas y contenido génico (Macedo et al., 2004; Macedo and Pena, 1998; Zingales et al., 2012). Respecto a la última característica, se ha encontrado que diferentes DTU poseen un contenido génico diferente en términos del número de copias de ciertas secuencias, como el ADN satélite, el cual se utiliza para el diagnóstico de la enfermedad de Chagas por PCR, comprometiendo la sensibilidad de la técnica de diagnóstico, lo que hace que parásitos de ciertas DTU se detecten con mayor facilidad en sangre que otros (Duffy et al., 2009).

TcI es el linaje de *T. cruzi* más abundante y ampliamente distribuido pudiendo relacionarse con ciclos domésticos y selváticos. La infección humana por TcI se concentra en el norte de Sudamérica, Centroamérica y México y se asocia con miocardiopatía chagásica (Zingales, 2018; Zingales et al., 2012). Esta DTU presenta la mayor heterogeneidad genética, en la que se han descrito cinco genotipos intra-DTU TcI (Ia, Ib, Ic, Id, Ie) basados en polimorfismos de secuencia en la región intergénica del mini-exón (Cura et al., 2010; Falla et al., 2009). TcII es tan antiguo como TcI, pero tiene una distribución geográfica más limitada, predominando en las regiones sur y central de América del Sur. Se asocia con sintomatología cardíaca y puede darse megaesófago y megacolon concomitante. Presenta una diversidad

genética mucho menor que TcI, no teniendo la misma capacidad de expansión y siendo aislada, mayoritariamente, en ciclos de transmisión domésticos. Los huéspedes y vectores naturales de TcII son imprecisos y la mayoría de los aislamientos registrados se han hecho de primates en regiones de la selva atlántica de Brasil y, esporádicamente, de otras especies de mamíferos (Fernandes et al., 1999; Lisboa et al., 2007; Zingales et al., 1999). TcIII se asocia, principalmente, con el ciclo selvático y el nicho terrestre en Brasil y países adyacentes (Marcili et al., 2009). La infección humana es rara, pero se han documentado algunos casos en Colombia (Hernández et al., 2016; Ramírez et al., 2010) y en la amazonia brasileña (Monteiro et al., 2010). TcIV también se asocia predominantemente con el ciclo selvático, y se encuentra en América del Norte y del Sur. En América del Norte, el mapache es el principal reservorio, mientras que, en América del Sur el hospedador más importante es el coatí (*Nasua nasua*) (Zingales, 2018). TcIV es la segunda cepa responsable de la enfermedad de Chagas en Venezuela (Miles et al., 1981) y la causa de brotes orales en la amazonia brasileña (Monteiro et al., 2012). TcV y TcVI (derivado de la hibridación de los DTUs TcII y TcIII), son las dos DTU cuya asociación con la enfermedad de Chagas es más evidente en los países del sur de América del Sur. Ambas DTU están asociadas con ciclos domésticos (Zingales et al., 2012) (**Figura 11**). Por último, Tcbat, se describió posteriormente en murciélagos como DTU independiente, respaldado por análisis filogenéticos y filogeográficos junto con múltiples marcadores moleculares específicos de Tcbat (Lima et al., 2015). La técnica empleada para el genotipado de los aislados de *T. cruzi* se basa en la amplificación mediante PCR de secuencias de genes codificantes del ARN ribosómico 24S $\alpha$ , 18S y mini-exón (Araújo et al., 2011; Brisse et al., 2001).

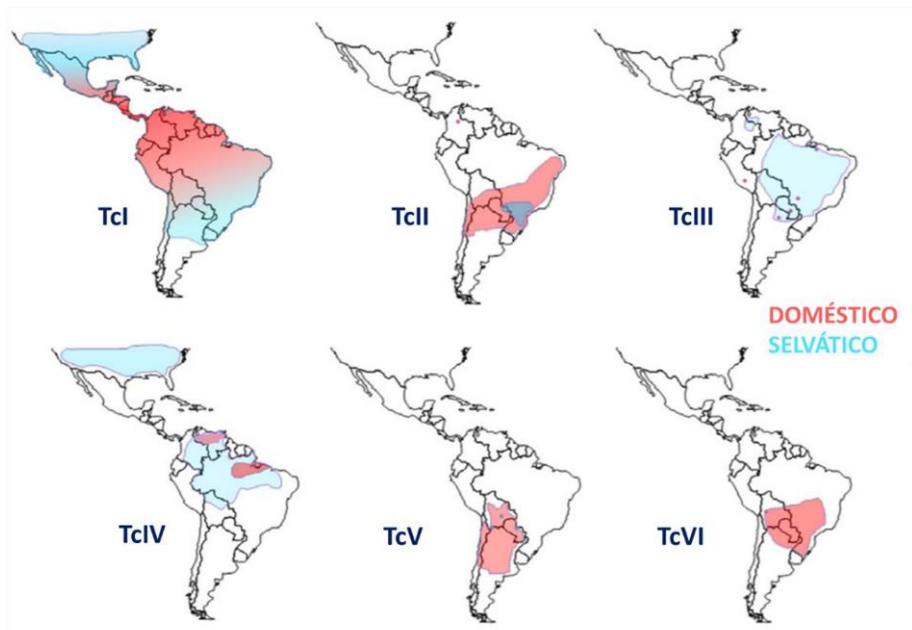


Figura I1. Distribución geográfica aproximada de las DTU de *T. cruzi* en los ciclos de transmisión domésticos y selváticos. Adaptada de Zingales et al., 2012.

### 1.2. Ciclo de vida de *Trypanosoma cruzi*

El ciclo biológico natural de *T. cruzi* es complejo y digénico, con fases de desarrollo en el insecto vector (perteneciente a la familia *Reduviidae*) y en el hospedador mamífero, que pueden ser humanos, o animales domésticos o salvajes. El parásito presenta diferentes estadios morfológica y molecularmente diferentes: epimastigote, tripomastigote y amastigote (**Figura I2**). El epimastigote corresponde con formas alargadas, flageladas, extracelulares y replicativas (**Figura I2A**). Los amastigotes son intracelulares, redondeados y sin flagelo visible (**Figura I2C**). La forma tripomastigote es también alargada y flagelada, pero, a diferencia de las anteriores, esta forma no es replicativa (**Figura I2B**).

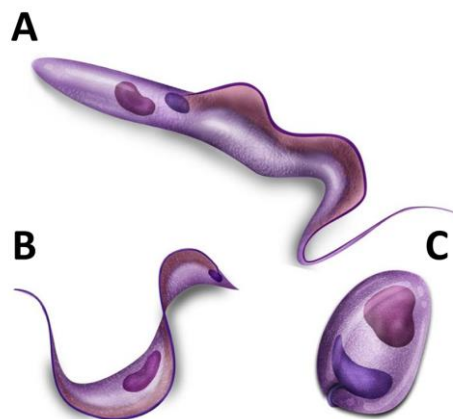
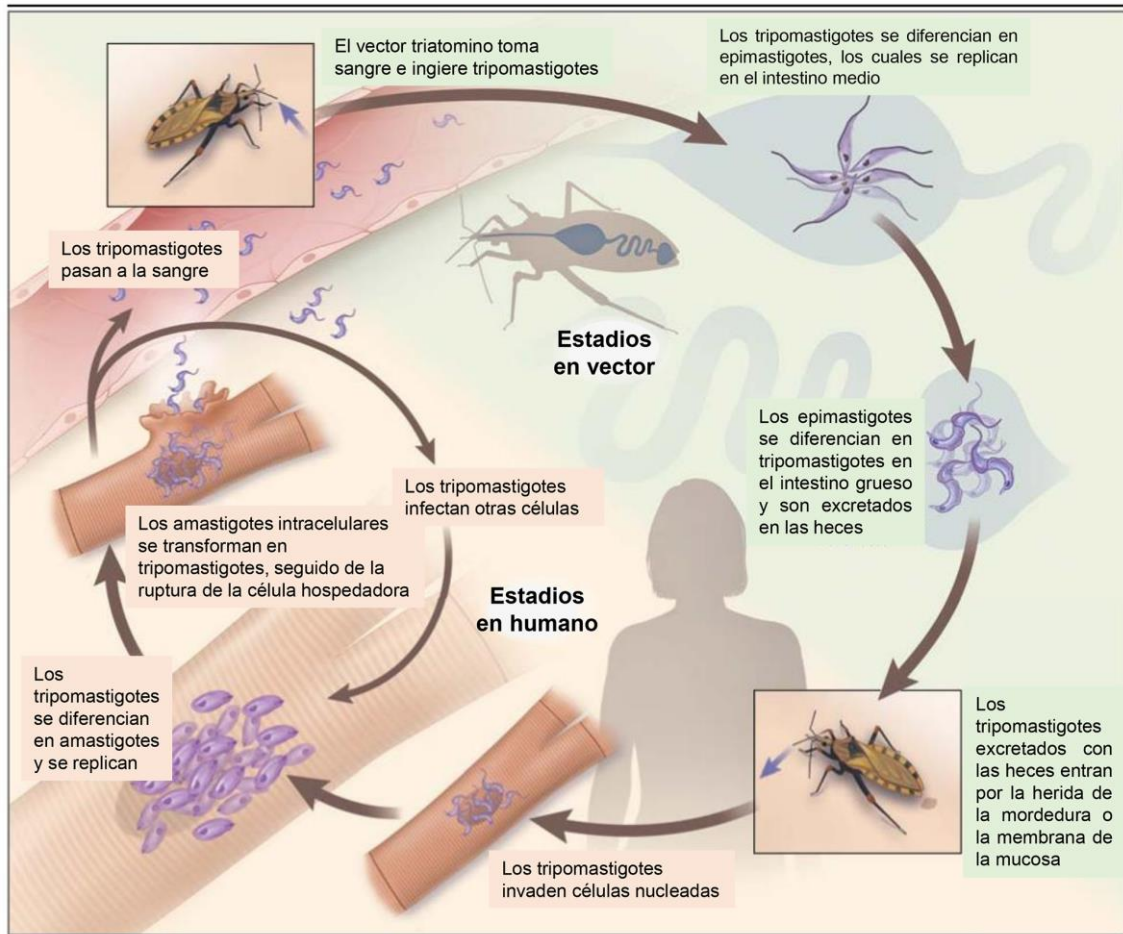


Figura I2. Formas de *T. cruzi*: epimastigote (A), tripomastigote (B) y amastigote (C). Figura obtenida de: <https://www.pinterest.es/pin/440086194812061375/?lp=true>

Los tripomastigotes no replicativos del torrente sanguíneo y los amastigotes intracelulares replicativos son las formas del parásito que se identifican en los huéspedes mamíferos, mientras que los epimastigotes replicativos y los tripomastigotes metacíclicos infecciosos son las identificadas con mayor frecuencia en el vector triatomino (Rassi et al., 2010).

El insecto vector pica a un mamífero infectado e ingiere los tripomastigotes presentes en su torrente sanguíneo, estos se diferencian en epimastigotes en la región media del intestino del vector donde se multiplican por fisión binaria; posteriormente, en la región posterior del intestino (intestino grueso o recto), tiene lugar el proceso de metaciclogénesis en el que los parásitos se diferencian en el estadio infectivo no replicativo, tripomastigote metacíclico. Los tripomastigotes excretados en las heces del insecto vector, ingresan en el hospedador mamífero por mucosas o lesiones previas o aquellas producidas por la picadura del insecto. Una vez en el hospedador mamífero, los tripomastigotes invaden células nucleadas, adyacentes al sitio de infección, como macrófagos, fibroblastos y células del sistema nervioso y muscular (Burleigh and Andrews, 1995). En el interior de las células, los tripomastigotes se diferencian en la forma replicativa amastigote, la cual se multiplica dentro de diferentes tipos celulares, preferentemente los de origen mesenquimatoso. Finalmente, los amastigotes se diferencian en tripomastigotes que producen la lisis de las células hospedadoras, siendo liberados al torrente sanguíneo. Los tripomastigotes infectan las células de los tejidos de alrededor y se mueven a través de los vasos linfáticos y el torrente sanguíneo hasta lugares lejanos. El ciclo de transmisión se completa en el momento que el insecto vector ingiere formas tripomastigotes de la sangre de mamíferos con la infección (Pérez-Molina and Molina, 2018; Rassi et al., 2012) (**Figura I3**).



**Figura I3. Ciclo de vida de *T. cruzi*.** Se muestran las diferentes formas del parásito presentes en el vector invertebrado y en el hospedador mamífero. *Adaptada de Bern, 2015.*

### 1.3. Vías de transmisión de la enfermedad de Chagas

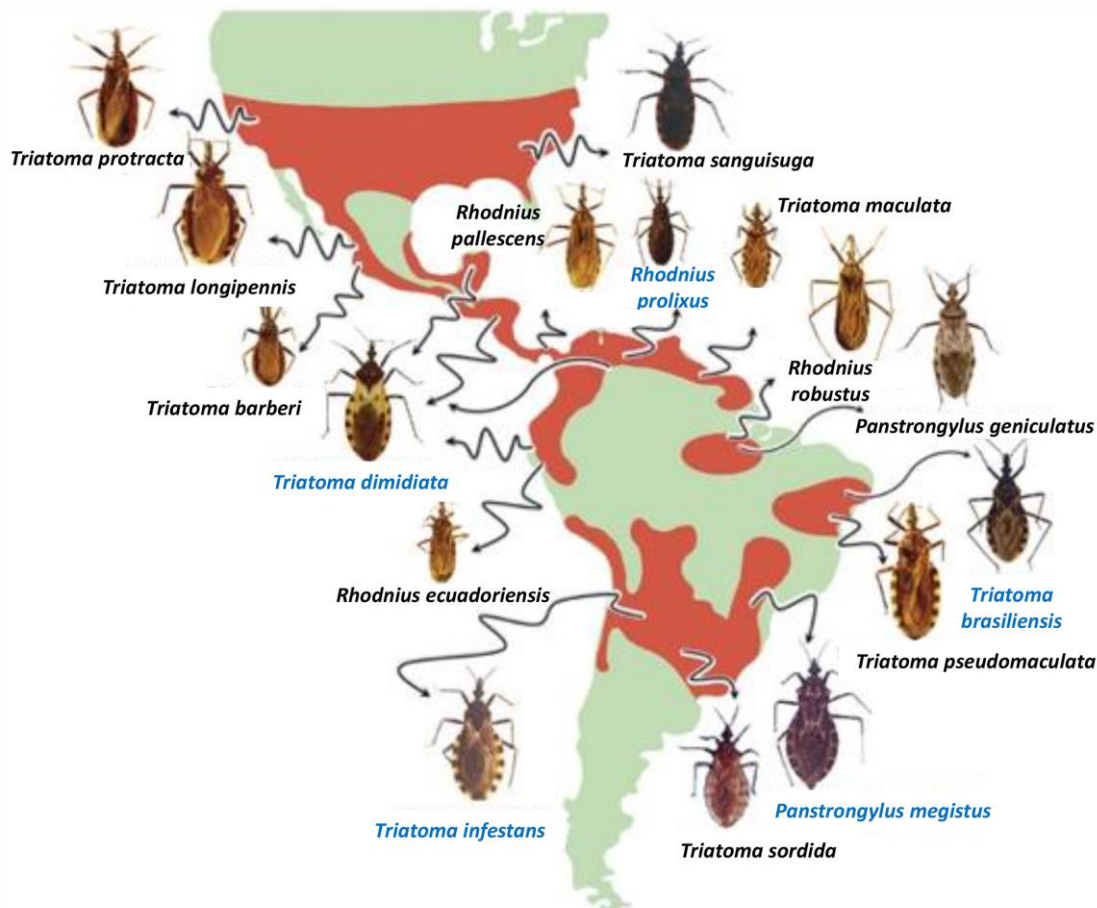
La transmisión de *T. cruzi* en zona endémica ocurre principalmente por transmisión vectorial, pero existen otras vías de transmisión que se dan tanto en zonas endémicas como no endémicas de la enfermedad. Estas son la transmisión congénita, por transfusiones sanguíneas o trasplante de órganos, por vía oral o por accidentes de laboratorio.

#### - Trasmisión vectorial (a través de las heces del insecto)

La más común de todas las formas de transmisión de la enfermedad de Chagas, tanto en humanos como en más de 150 especies de animales domésticos, peridomésticos o salvajes, es la que se produce a través de insectos hematófagos (machos y hembras) del orden *Hemiptera*, familia *Reduviidae*, subfamilia *Triatominae* (Rassi et al., 2012).



Aunque se han identificado alrededor de 150 especies de triatominos, sólo algunas son vectores competentes de *T. cruzi*. Los principales vectores son *Triatoma infestans*, *Triatoma brasiliensis* y *Panstrongylus megistus* en los países del Cono Sur, *Rhodnius prolixus* y *Triatoma dimidiata* en la región andina y partes de América Central, y *Triatoma dimidiata* y *Triatoma barberi* en México (Gorla and Noireau, 2017) (Figura I4).



**Figura I4. Distribución de las especies de triatominos de mayor relevancia epidemiológica.** Las áreas rojas señalan la distribución geográfica aproximada de las distintas especies. Los vectores del parásito que se consideran más relevantes se resaltan en azul. *Adaptada de Gourbière et al., 2012.*

Se distinguen 3 ciclos de transmisión vectorial de *T. cruzi*: doméstico, peridoméstico y selvático. El ciclo doméstico se da a través de triatominos que transmiten el parásito de animales domésticos a humanos, así como entre humanos, y el ciclo selvático se conserva entre animales salvajes y triatominos. El ciclo peridoméstico sirve de nexo entre ambos ciclos. Este se originó a partir del ciclo silvestre y mantiene la infección entre los animales domésticos en las zonas que rodean las viviendas humanas y, ocasionalmente, por la ingesta de animales salvajes infectados por parte de los animales domésticos o por la invasión de viviendas con

animales salvajes. Por ello, los tres ciclos de transmisión pueden interrelacionarse entre sí (Coura and Dias, 2009).

### - Trasmisión vertical o congénita (de madre a hijo, durante el embarazo)

La transmisión de madre a hijo es una forma de transmisión frecuente, también en zonas no endémicas de la enfermedad. Puede ocurrir en cualquier estadio de la enfermedad, en cualquier momento del embarazo y producirse en diferentes embarazos (Schenone et al., 2001).

Se cree que la transmisión materno fetal es multifactorial (Torricco et al., 2007) donde influyen la disminución de la respuesta inmune de la madre, su origen geográfico (que puede condicionar la o las cepas de parásitos implicados y la densidad vectorial intradomiciliaria), su estado nutricional, edad, bases genéticas, la fase evolutiva de la enfermedad materna y la reactivación de la infección o las reinfecciones (de origen vectorial) a lo largo del embarazo (Bittencourt, 1992; de Andrade et al., 1994; Moretti et al., 2005).

La transmisión congénita se da en, al menos, el 5% de las mujeres embarazadas con infección crónica de ciertas regiones de Bolivia, Chile y Paraguay, y en el 1-2% o menos en la mayoría de los demás países endémicos (Russomando et al., 1998; Torricco et al., 2004; WHO, 2002) pudiendo alcanzar hasta un 12% (Torricco et al., 2007). En un estudio realizado en 1.350 mujeres embarazadas con origen latinoamericano residentes en Barcelona se han obtenido cifras de seroprevalencia frente a *T. cruzi* del 3,4% y de transmisión a la prole del 7,3% (Muñoz et al., 2009). El tratamiento de mujeres en edad fértil ha mostrado interrumpir la transmisión congénita (Murcia et al., 2013, 2017; Sosa-Estani et al., 2009).

### - Trasmisión por transfusiones sanguíneas y trasplantes de órganos

La infección también se puede transmitir vía transfusión de sangre (o sus derivados) o tras trasplante de órganos de una persona infectada.

El riesgo de contraer la infección por *T. cruzi* después de la transfusión de 1 unidad de sangre de un donante infectado oscila de un 10% a 20% (Rassi et al., 2010). El riesgo de transmisión es superior para la transfusión de plaquetas que para otras células sanguíneas, por la predominancia de tripomastigotes en esta fracción tras la centrifugación (Rassi et al., 2010). La tasa de infección tras el



trasplante de órganos sólidos de un sujeto infectado parece ser menor para los individuos receptores de riñón (0-19%) (Cicora et al., 2014; Huprikar et al., 2013; Riarte et al., 1999) que para receptores de hígado (0-29%)(Chin-Hong et al., 2011; Huprikar et al., 2013) y receptores de corazón (75-100%) (Huprikar et al., 2013; Kun et al., 2009).

En la actualidad, siguiendo las recomendaciones de la OMS, se realiza cribado de los donantes que proceden o han viajado a Sudamérica por lo que esta vía de transmisión está controlada.

### - Trasmisión oral

Aunque este tipo de trasmisión es menos frecuente, también es posible contraer la enfermedad Chagas al ingerir comida o bebida contaminada con triatomos infectados o con sus heces, presentes en zumos naturales de açaí, caña de azúcar o guayaba y carne cruda. El cuadro clínico es agudo y a esta vía se asocia una alta mortalidad (Pereira et al., 2009). En los últimos años, se han identificado varios brotes de trasmisión oral, entre ellos uno que afectó a más de 100 personas en un colegio de Venezuela por consumo de zumo contaminado (de Noya et al., 2010, 2015).

### - Trasmisión por accidentes de laboratorio

Esta forma de transmisión es muy poco frecuente. Puede darse en trabajadores que manipulen el insecto vector o directamente el parásito por inoculación accidental a causa de pinchazos o exposición a mucosas. Las personas expuestas deben recibir instrucciones exhaustivas sobre las precauciones de seguridad antes de comenzar a trabajar y mediante programas de capacitación continua, así como proporcionarse protocolos para la manipulación de muestras, el uso de ropa y equipo de protección, el tratamiento de derrames de organismos infecciosos y la respuesta a accidentes (Herwaldt, 2001).

## **1.4. Epidemiología de la enfermedad de Chagas**

La enfermedad de Chagas se considera una de las enfermedades tropicales desatendidas más prevalentes, la cual causa entre 10.000 y 14.000 muertes al año (PAHO and WHO, 2021; Rassi et al., 2012). La Organización Mundial de la Salud

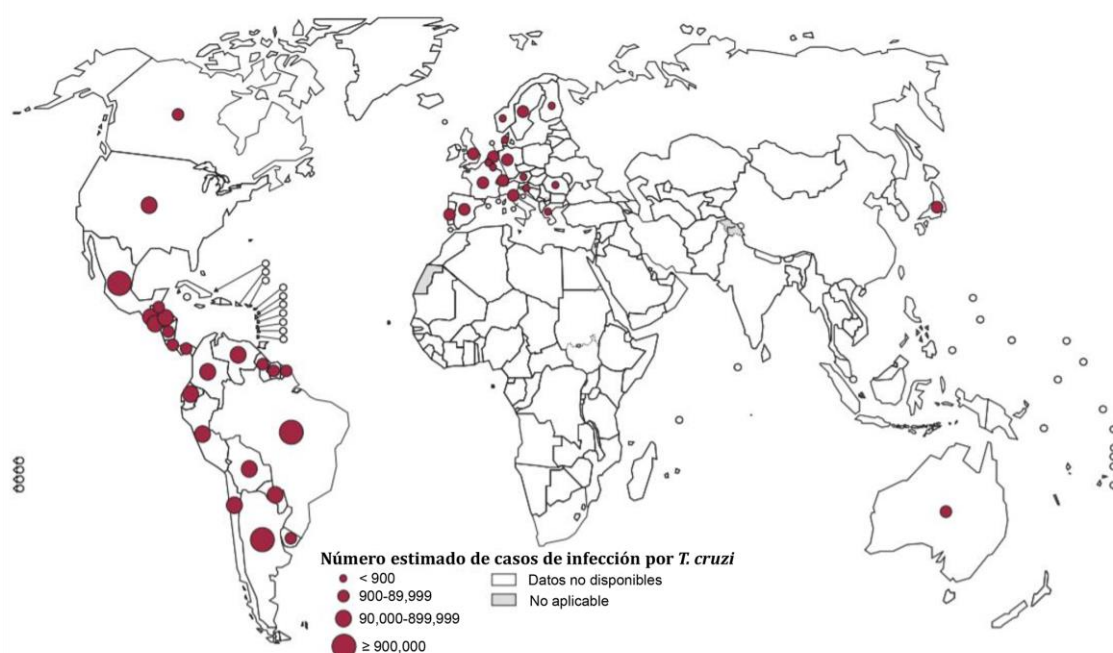
estima que aproximadamente 6-7 millones de personas en el mundo están infectadas por *T. cruzi*, principalmente en áreas endémicas de 21 países de América Latina: Argentina, Belice, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guayana Francesa, Guatemala, Guyana, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, Surinam, Uruguay y Venezuela (WHO, 2017).

Los países en los que se han identificado mayores prevalencias de la enfermedad de Chagas son Bolivia (6,8%), Argentina (4,1%), El Salvador (3,4%), Honduras (3,1%) y Paraguay (2,5%). Sin embargo, Brasil y México (con prevalencias de alrededor el 1%), junto con Argentina, son los países donde se encuentran casi el 60% de todas las personas infectadas por *T. cruzi* en América Latina (Rassi et al., 2012).

En la mayoría de los países endémicos, se han desarrollado iniciativas con el objetivo de erradicar la enfermedad de Chagas (Moncayo, 2003). Desde principios de los años 90, las medidas más efectivas para controlar la enfermedad en América Latina han sido programas para la contención del vector y el cribado obligatorio en bancos de sangre de donantes (WHO, 2002; Yamagata and Nakagawa, 2006). Como consecuencia, la prevalencia ha disminuido sustancialmente interrumpiéndose la transmisión vectorial en Uruguay en 1997, en Chile en 1999, en Brasil en 2006 y, años posteriores, en otras regiones de América Latina (Salvatella et al., 2014; WHO, 2002).

El patrón epidemiológico de la enfermedad ha cambiado pasando de ser una afección relacionada con la pobreza de las áreas rurales latinoamericanas a estar presente mayoritariamente en áreas urbanizadas, debido a la movilidad de la población (WHO, 2017) (**Figura 15**). La migración desde zonas endémicas a países desarrollados ha hecho que la enfermedad también esté presente en áreas externas a las tradicionales fronteras geográficas, donde el vector no está presente, como Estados Unidos (Manne-Goehler et al., 2016), Canadá, Australia, así como en muchos países europeos y algunos de África, el Mediterráneo Oriental y el Pacífico Occidental (Antinori et al., 2017; Schmunis, 2007; Schmunis and Yadon, 2010; WHO, 2021). España es el país fuera del continente americano donde vive el mayor número de personas infectadas, aproximadamente 52.000 personas, pero sólo un 10% de ellos están diagnosticados (Romay-Barja et al., 2019).

El elevado número de personas que siguen sin ser diagnosticadas o tratadas y que tienen una infección en curso combinada con la transmisión vectorial y oral activa ponen a millones de personas en riesgo de adquirir la infección. En 2020, se incorporaron a la hoja de ruta para las enfermedades tropicales desatendidas 2021-2030 de la OMS, cinco objetivos relacionados con la enfermedad de Chagas: verificación de la interrupción de la transmisión vectorial domiciliaria, por transfusiones y por trasplante de órganos, eliminación de la enfermedad de Chagas congénita y 75% de cobertura del tratamiento antiparasitario de la población en riesgo (WHO, 2021).



**Figura I5.** Distribución global de los casos de la enfermedad de Chagas, basado en estimaciones oficiales 2006-2015. *Adaptada de WHO, 2017.*

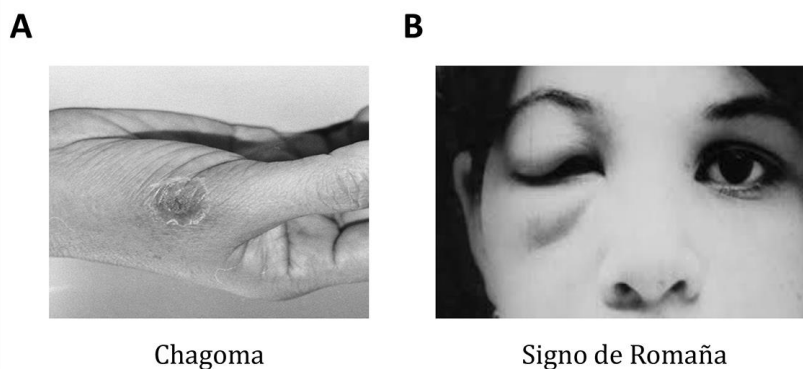
## 1.5. Manifestaciones clínicas de la enfermedad de Chagas

La enfermedad de Chagas se caracteriza por desarrollarse en dos etapas o fases clínicas: fase aguda y fase crónica.

### 1.5.1. Fase aguda

Esta fase se caracteriza por una alta parasitemia detectable en sangre periférica y suele tener una duración de entre 4 a 8 semanas (Bern et al., 2011; Pérez-Molina and Molina, 2018). Los pacientes en esta fase suelen ser asintomáticos. Los síntomas clínicos suelen aparecer entre 1 y 2 semanas tras la infección o a los 2

meses tras la transfusión con sangre infectada. Son leves e inespecíficos que incluyen fiebre, linfadenopatía, hepatoesplenomegalia e inflamación del lugar de entrada como el chagoma (**Figura I6A**), zona eritematosa que aparece donde ha tenido lugar la picadura del triatomino 1-2 semanas después, o el signo de Romaña (**Figura I6B**), edema periorbitario, unilateral e indoloro que aparece cuando la conjuntiva es el lugar de entrada (Pérez-Molina and Molina, 2018; Rassi et al., 2010).



**Figura I6. Manifestaciones clínicas de la fase aguda de la enfermedad de Chagas.** A) Chagoma. B) Signo de Romaña. *Imágenes tomadas de:* Kinoshita-Yanaga et al., 2009; Pérez Yanez et al., 2017.

La mortalidad en esta fase es baja (afectando a menos de un 5-10% de los casos sintomáticos) y suele darse como resultado de una miocarditis severa, meningoencefalitis o ambas (Prata, 2001; Rassi et al., 2010).

### 1.5.2. Fase crónica

Si el paciente en fase aguda no se trata, la enfermedad se cronifica (Pérez-Molina and Molina, 2018; Rassi et al., 2010). Esto se debe a que el parásito pasa de la sangre al tejido, dónde permanece oculto y, a su vez, menos accesible a la respuesta inmune generada frente a él (Tarleton, 2001). Esta fase se caracteriza por un largo periodo de latencia asintomático que desemboca para algunos pacientes en la presentación de manifestaciones clínicas que comprometen al sistema cardiaco, digestivo y/o neurológico (Prata, 2001).

#### 1.5.2.1. Fase crónica indeterminada

Durante la fase crónica, la mayoría de los pacientes son asintomáticos, no presentando ningún síntoma ni signo clínico. Esto es posible debido a la existencia de un frágil balance entre la respuesta inmune del hospedador y la replicación del parásito, lo que hace que los pacientes puedan llegar a permanecer silentes

clínicamente de 10 a 25 años (dos Santos Virgilio et al., 2014). Esta fase asintomática es conocida como fase indeterminada de la enfermedad de Chagas, con una buena prognosis, caracterizada por seropositividad para *T. cruzi*, ausencia de síntomas cardiacos o digestivos, con electrocardiograma normal, así como radiografía de tórax, esófago y colon (Pinto Dias, 1989; WHO, 2021).

### **1.5.2.2. Fase crónica sintomática**

El desbalance entre la respuesta del sistema inmune y la replicación del parásito es crucial para la progresión de la enfermedad. Así, un 30-40% de los pacientes crónicos desarrollan una fase sintomática (Rassi et al., 2010). Alrededor del 30% de ellos presenta alteraciones cardíacas y hasta un 15% presentan megasíndromes digestivos y/o trastornos neurológicos (Rassi et al., 2010).

#### **- Afectación cardíaca**

Es la sintomatología más importante a tener en cuenta, debido a la frecuencia de aparición, así como, a las consecuencias de su progresión. Las alteraciones que pueden ocurrir incluyen agrandamiento del corazón o miocardiopatía, frecuencia o ritmo cardíaco alterado, insuficiencia cardíaca, paro cardíaco, tromboembolismo y muerte súbita. Los casos más graves de alteraciones cardíacas conducen al desarrollo de una miocardiopatía chagásica crónica, asociada a altas tasas de mortalidad en la enfermedad de Chagas (Morris et al., 1990; Rocha et al., 2003).

En 1985, Kuschnir propuso una clasificación de la cardiopatía chagásica en 4 estadios atendiendo al grado de afectación cardíaca, y, por tanto, su gravedad (Kuschnir et al., 1985) (**Tabla I1**):

- Grupo 0: pacientes asintomáticos con serología positiva frente a *T. cruzi*, ECG y radiografía de tórax normal.
- Grupo I: pacientes con serología positiva, ECG patológico (portadores de alteraciones de conducción intraventricular y/o extrasístole ventricular) y radiografía de tórax normal.
- Grupo II: pacientes que, además de la serología y de las alteraciones en el ECG, muestran cardiomegalia.
- Grupo III: pacientes con serología positiva, alteraciones electrocardiográficas severas, cardiomegalia y síntomas y signos clínicos de insuficiencia cardíaca.

## INTRODUCCIÓN

---

Tabla I1. Clasificación según Kuschnir de la afectación cardíaca en la fase crónica de la enfermedad de Chagas. ECG: Electrocardiograma.

	Clasificación de Kuschnir			
	0	I	II	III
Serología	+	+	+	+
ECG	Normal	Patológico	Patológico	Patológico
Rx Tórax	Normal	Normal	Cardiomegalia	Cardiomegalia
Sintomatología	Asintomático	Asintomático	Asintomático	Insuficiencia cardíaca

### - Afectación digestiva

La sintomatología digestiva aparece con menor frecuencia y gravedad y aunque no suele suponer un riesgo vital, puede tener un gran impacto negativo en la calidad de vida de los pacientes. Afecta aproximadamente entre el 15-20% de los sujetos infectados, los cuales pueden desarrollar alteraciones de la motilidad, secreción y absorción en el tracto digestivo, principalmente en esófago y colon. En primer lugar, las modificaciones en la motilidad surgen con un tránsito lento y estreñimiento, seguido de la dilatación del órgano afectado dando lugar a megaesófago y megacolon (Prata, 2001). La afectación esofágica es progresiva en algunos pacientes y evoluciona de forma independiente a la afectación cardíaca. Sin embargo, aproximadamente la mitad de pacientes con esofagopatía presentan en su electrocardiograma alteraciones compatibles con afectación cardíaca, existiendo, por tanto, formas de afectaciones cardiodigestivas de la enfermedad de Chagas (Prata, 2001).

### - Afectación del sistema nervioso

Es la sintomatología menos frecuente y se presenta, principalmente, con polineuritis sensitivomotora caracterizada por parestesias e hipoestesia en miembros y descenso o abolición de los reflejos osteotendinosos (Prata, 2001).

La reactivación de la enfermedad de Chagas puede darse en pacientes en fase crónica bajo situaciones que provoquen inmunosupresión con recuento de células CD4<sup>+</sup> menor a 200 células/mm<sup>3</sup>. Este suele ser el caso de pacientes coinfectados con VIH o que tomen inmunosupresores, como es el caso de personas sometidas a trasplante de órganos. Con la reactivación, los parásitos proliferan pudiendo ocasionar lesiones necróticas o tumorales en el cerebro (75%) e intensificación de la miocarditis (44%) (Prata, 2001).

## 1.6. Diagnóstico de la enfermedad de Chagas

Para el diagnóstico de la enfermedad de Chagas pueden aplicarse varios métodos directos o indirectos. Los métodos directos detectan el parásito en sangre periférica o tejidos y la mayoría están basados en la microscopía mientras que los métodos indirectos detectan el ADN del parásito (principalmente por PCR) o la secreción de anticuerpos específicos frente a *T. cruzi*.

En la fase aguda o durante la reactivación de la enfermedad, la alta densidad de parásitos en sangre permite su detección en muestras de sangre periférica, aplicando técnicas como la visualización al microscopio óptico, por métodos de concentración como Scrot y microhematocrito (generalmente usado en recién nacidos), así como otras técnicas más usadas en áreas endémicas como el xenodiagnóstico o hemocultivo (Abramo-Orrego et al., 1980; Freilij and Altchek, 1995; Strout, 1962).

En fase crónica, el nivel de parasitemia disminuye, por lo que se requieren técnicas moleculares más sensibles (PCR) y serológicas. Las técnicas serológicas para la detección de anticuerpos frente a antígenos de *T. cruzi* incluyen inmunofluorescencia indirecta (IFI), hemaglutinación indirecta (HAI) o enzimoimmunoanálisis de adsorción (ELISA) (Camargo, 1966; Camargo et al., 1971; Engvall and Perlmann, 1971; Fife and Muschel, 1959; Voller et al., 1975). Los antígenos empleados en estas técnicas pueden ser extractos solubles de proteínas del parásito, antígenos purificados, proteínas recombinantes o péptidos sintéticos. La OMS (WHO, 2002) recomienda el uso de dos técnicas serológicas diferentes para la confirmación de resultados positivos y, en caso de discordancia, la aplicación de una tercera técnica. En los últimos años, se han desarrollado varios test comerciales basados en antígenos recombinantes de *T. cruzi* como *BioELISA Chagas* (Biokit, Lliçà d'Amunt, España), *Chagatest ELISA recombinante* (Wiener Lab) y *Chagas Antibody EIA* (Abbott Laboratories, Wiesbaden, Alemania), entre otros.

Además de las técnicas de laboratorio, se recomienda la exploración física de los pacientes aplicando técnicas de diagnóstico como el electrocardiograma, radiografía de tórax o ecocardiografía para la detección de afectación cardiaca. Para establecer la afectación gastrointestinal en los pacientes asintomáticos, los

procedimientos diagnósticos más comunes son la deglución o enema de bario y la manometría anorrectal (Salvador et al., 2015).

Si bien el diagnóstico de la enfermedad de Chagas no supone un problema, sí lo es el acceso de toda la población a los servicios de salud, especialmente en zonas rurales.

### 1.7. Tratamiento de la enfermedad de Chagas

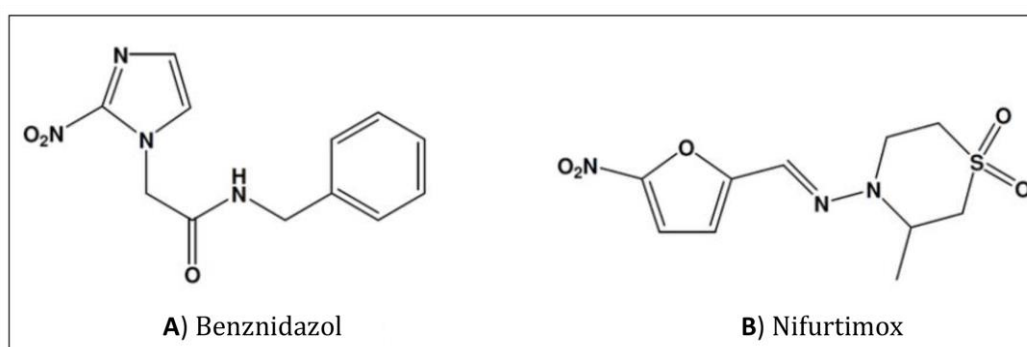
El tratamiento de la enfermedad de Chagas se centra en la eliminación del parásito. A pesar de que se han investigado muchos compuestos candidatos, en la actualidad sólo se encuentran disponibles dos fármacos que han demostrado eficacia frente a *T. cruzi*: Benznidazol y nifurtimox. Ambos son compuestos nitroheterocíclicos y su mecanismo de acción no está claro. El benznidazol (Radanil®, Roche (1972)) (**Figura I7A**) es un nitroimidazol cuya actividad tripanocida se atribuye al enlace covalente de los metabolitos reducidos del grupo nitro a lípidos, ADN y proteínas del parásito (Díaz de Toranzo et al., 1988). También, se ha descrito que el benznidazol puede mejorar la fagocitosis e incrementar la muerte tripanosomal a través de la inducción de IFN- $\gamma$  (Murta et al., 1999). En adultos se administran 5 mg/kg/día en dos dosis, durante 60 días.

El nifurtimox (Lampit®, Bayer (1967)) (**Figura I7B**) es un nitrofurano cuyo mecanismo de acción se basa en el estrés oxidativo, debido a la transformación del fármaco en un radical nitroanión, y en la generación de especies reactivas de oxígeno (ROS) (Docampo and Stoppani, 1979). La ausencia o baja actividad en el parásito de enzimas que eliminan los radicales libres lo hace más sensible que las células humanas (Maya et al., 2007). Su administración es de 10 mg/kg/día, en tres dosis, durante 60-90 días.

El benznidazol se considera el fármaco de primera elección debido a su mejor tolerabilidad y eficacia (Murcia et al., 2012; Viotti et al., 2009). No obstante, cabe destacar que la administración de ambos fármacos produce frecuentes y severos efectos secundarios. Los más comunes tras la administración de nifurtimox son pérdida de peso, anorexia, alteraciones psíquicas, somnolencia, excitabilidad, manifestaciones digestivas como náuseas o vómitos, y, ocasionalmente, diarrea y



cólicos intestinales. En el caso del benznidazol, son más notorias las alteraciones cutáneas (hipersensibilidad, dermatitis con erupciones cutáneas, edema generalizado), fiebre, linfadenopatía, dolor articular y muscular, así como manifestaciones más graves como la depresión de la médula ósea, púrpura trombocitopénica y agranulocitosis (Castro et al., 2006). Estos efectos secundarios están directamente relacionados con la toxicidad de estos fármacos lo cual es una de las principales limitaciones del tratamiento provocando interrupciones y abandonos en los pacientes bajo tratamiento. En un ensayo aleatorizado reciente en pacientes crónicos con cardiomiopatía chagásica, un 23,9% de ellos interrumpieron el tratamiento (Morillo et al., 2015).



**Figura 17. Estructura química de los fármacos empleados para tratar la enfermedad de Chagas: (A) Benznidazol y (B) Nifurtimox.** Adaptada de Salomão and de Castro, 2017.

### 1.7.1. Eficacia del tratamiento

La eficacia de estos fármacos varía en función del tiempo que el paciente lleve infectado, es decir, de la fase en la que se encuentre en el momento de la administración, disminuyendo con la evolución de la enfermedad (Altcheh et al., 2005; Meymandi et al., 2018).

#### - Fase aguda

En la fase aguda, el tratamiento debe comenzarse lo antes posible, en infecciones congénitas, en niños menores de 14 años, accidentes de laboratorio y en los casos de reactivación de la infección (Pinazo et al., 2013; WHO, 2021). En estas situaciones, se ha demostrado que el tratamiento es eficaz al eliminar el parásito de sangre en un 60% de los casos (Coura and de Castro, 2002) llegando a alcanzar tasas del 76% (Cançado, 2002) y del 96% en infecciones congénitas (Alonso-Vega et al., 2013) y hasta el 100% de curación cuando el tratamiento se administra en

transmisión congénita a niños en los 8 primeros meses de vida (Altcheh et al., 2005). Adicionalmente, el tratamiento de mujeres infectadas en edad fértil, antes del embarazo, previene la transmisión congénita de la enfermedad (Meymandi et al., 2018; Murcia et al., 2013, 2017; Sosa-Estani et al., 2009).

### - Fase crónica

En la fase crónica hay controversias acerca de la efectividad del tratamiento al no existir parámetros que lo puedan medir. Se estima que la tasa de curación disminuye significativamente en fase crónica siendo de entre un 8-20% (Sales et al., 2017). No obstante, varios estudios llevados a cabo en pacientes crónicos han demostrado que el tratamiento antiparasitario reduce la detección del parásito en sangre (Morillo et al., 2015; Vallejo et al., 2016), previene el desarrollo de alteraciones en el electrocardiograma en pacientes crónicos con la forma indeterminada (Bustamante et al., 2007; Fragata-Filho et al., 2016; Garcia et al., 2005; Viotti et al., 2014) y conduce hacia una mejora en la respuesta inmune frente a *T. cruzi* (Albareda and Laucella, 2015; Vallejo et al., 2016). Debido a estos datos, la OMS recomienda la administración del tratamiento en pacientes crónicos. En 2014, miembros de la red NHEPACHA (Nuevas Herramientas para el diagnóstico y la Evaluación del Paciente con enfermedad de CHAgas), basándose en la evidencia inmunológica y clínica, también recomendaron la administración obligatoria del tratamiento antiparasitario a todos los pacientes adultos con enfermedad de Chagas en fase crónica (Viotti et al., 2014). Cabe destacar que en el caso de pacientes crónicos con enfermedad cardíaca avanzada el tratamiento no ha demostrado eficacia dado que no es posible la reversión del daño miocárdico (Morillo et al., 2015).

Según la OMS, el fármaco ideal para el tratamiento de la enfermedad de Chagas debe cumplir las siguientes características: eliminación del parásito durante la fase aguda y crónica, eficacia en una o pocas dosis, bajo coste, ausencia de efectos secundarios o teratogénicos y no inducir resistencia. Hasta la fecha, ningún fármaco cumple estos criterios, por lo que son necesarios nuevos compuestos más efectivos y mejor tolerados.

En busca de la mejora de la eficacia y reducción de los efectos secundarios de los fármacos existentes, en la actualidad, nuevas moléculas anti-*T. cruzi* están siendo

investigadas. Algunas de las principales aproximaciones experimentales y preclínicas se basan en inhibidores de la biosíntesis del ergosterol, metabolismo de tripanotomas, inhibidores de cisteína proteasas, metabolismo de pirofosfatos, síntesis de proteínas y purinas, análogos de lisofosfolípidos e inhibidores del proteosoma del kinetoplasto o en estrategias de microencapsulación del benznidazol que mejoren su farmacocinética con objeto de poder reducir la dosis administrada y, por consecuencia, los efectos adversos (Sales et al., 2017).

### **1.8. Biomarcadores de seguimiento de la enfermedad de Chagas**

#### **1.8.1. Biomarcadores de eficacia terapéutica**

Además de que la eficacia se encuentra reducida en la fase crónica de la enfermedad de Chagas, una de las limitaciones adicionales es la falta de marcadores que midan la eficacia del tratamiento (Pinazo et al., 2014). Hasta ahora, el criterio admitido para aceptar la eficacia del tratamiento es la seronegativización medida por los test serológicos convencionales, sin embargo, esta suele tardar décadas en establecerse (Fabbro et al., 2007; Viotti et al., 2006). Un estudio de seguimiento serológico de 430 pacientes con enfermedad crónica de Chagas mostró que la seronegativización completa se alcanzaba en una mediana de 11,7 años (Viotti et al., 2006).

La técnica de PCR se ha propuesto como marcador temprano de fracaso terapéutico, de tal forma que si tras el tratamiento se obtiene un valor positivo indicaría que el parásito sigue estando presente y que, por tanto, el tratamiento no ha eliminado al parásito. Sin embargo, debido a las fluctuaciones de la parasitemia y a variaciones en el resultado de la PCR a lo largo del tiempo post-tratamiento durante la fase crónica, un resultado negativo no indicaría cura sino únicamente que no se ha detectado el parásito en la muestra (Rassi et al., 2012).

Esta situación pone de manifiesto la necesidad de desarrollar biomarcadores tempranos que permitan medir la eficacia del tratamiento en la enfermedad de Chagas. El Programa de Seguridad Química dirigido por la OMS (WHO and IPCS, 1993) define biomarcador como "cualquier sustancia, estructura o proceso que pueda medirse en el cuerpo, o sus productos, e influye o predice la incidencia o resultado de la enfermedad" (WHO and IPCS, 2001). En los últimos años, varios

grupos de biomarcadores han demostrado ser útiles para evaluar la respuesta a tratamientos específicos frente a *T. cruzi* (Pinazo et al., 2015). Estos se clasifican en biomarcadores de respuesta/daño del hospedador o biomarcadores del parásito. Entre los biomarcadores de respuesta/daño del hospedador se encuentran marcadores bioquímicos, como la apolipoproteína (APO) y fragmentos de fibronectina, marcadores de hipercoagulabilidad, así como la respuesta de células T anti-*T. cruzi*, expresión génica diferencial de citoquinas y moléculas de superficie, entre otros (Pinazo et al., 2011, 2015; Santamaria et al., 2014). Los biomarcadores del parásito incluyen proteínas o glicoproteínas aisladas del parásito, proteínas recombinantes o grupos de proteínas. Algunas de las proteínas que han sido evaluadas son el extracto antigénico F105, las fracciones F-III y F-IV de *T. cruzi*, exoantígenos de ratón infectado con formas tripomastigotes (EXO), la proteína rTc24, la proteína ribosomal P2 $\beta$ , set de 16 proteínas en multiplex, la proteína flagelar F29, tGPI-mucinas (antígenos F2, F2/3 y AT) y los antígenos recombinantes 1, 2, 13, 30, 36 y SAPA, así como otras 14 proteínas recombinantes de *T. cruzi* (Cooley et al., 2008; de Andrade et al., 1996; Fabbro et al., 2011, 2013; Krautz et al., 1995; Moretti et al., 1998; Negrette et al., 2008; Sosa-Estani et al., 1998; Viotti et al., 2011).

En estudios realizados en nuestro laboratorio se ha demostrado que las proteínas recombinantes KMP11, PFR2, HSP70 y el péptido 3973 podrían ser moléculas útiles para evaluar la eficacia del tratamiento en pacientes con enfermedad de Chagas. De esta forma, en el suero de pacientes se observó una disminución de la reactividad frente a KMP11 a los 6 meses tras el tratamiento con benznidazol y frente a PFR2, HSP70 y 3973 a los 9 meses post-tratamiento (Fernández-Villegas et al., 2011, 2016). Además, otro estudio mostró que el péptido 3973 era reconocido con una alta sensibilidad (> 90%) y especificidad (> 98%) por el suero de pacientes con enfermedad crónica de Chagas (Thomas et al., 2012). Las proteínas KMP11, PFR2, HSP70 junto con Tgp63, se emplearon en un estudio ciego post-tratamiento de dos gemelos infectados congénitamente (Fernández-Villegas et al., 2014) observándose una disminución en la reactividad frente a estas proteínas lo cual se correlacionó con la eficacia del tratamiento en uno de los hermanos. Sin embargo, en el suero del otro hermano no se observó modificación de la reactividad frente a estas moléculas tras el tratamiento. Posteriormente, tras entrevista a la familia se informó que el hermano 2 tan solo había tomado 1 de las 4 dosis de

benznidazol que le habían sido prescritas. Cuarenta y cinco días después de la administración de un segundo tratamiento con benznidazol al hermano, se observó un descenso en la reactividad frente a las moléculas analizadas.

El descenso en el nivel de anticuerpos específicos frente a los antígenos específicos del parásito se ha asociado con una reducción importante de la carga parasitaria y con una mejora del estadio clínico de los pacientes tratados con enfermedad de Chagas (Fragata-Filho et al., 2016; Viotti et al., 2011). Por tanto, la evaluación de la reactividad frente a estos antígenos podría ser un sistema útil de marcadores serológicos para el seguimiento de la eficacia del tratamiento en pacientes con enfermedad de Chagas que permitiría la detección precoz del potencial fracaso terapéutico.

#### **1.8.1.1. Antígenos de *T. cruzi* KMP11, HSP70, PFR2 y 3973<sup>d</sup>**

- **KMP11** (*Kinetoplastid Membrane Protein 11 kDa*)

KMP11 es una glicoproteína de la membrana de kinetoplastidos de 11 kDa. Se expresa abundantemente en todos los estadios de vida de kinetoplastidos con una mayor frecuencia en las formas que interactúan con el insecto vector. Es específica de tripanosomátidos conservando un elevado grado de homología entre ellos. Se localiza asociada al citoesqueleto y muestra una gran capacidad inmunogénica (Carvalho et al., 2005; Goel et al., 1999; Thomas et al., 2000, 2001).

KMP11 es codificada por genes repetidos organizados en tándem de los que se han identificado 3 copias en las especies *Leishmania* (Berberich et al., 1997; Jardim et al., 1995; Ramírez et al., 1998) y 4 en tripanosomas (Bridge et al., 1998; Thomas et al., 2000). Esta proteína estimula a células B específicas, induciendo proliferación celular, respuesta citotóxica e inmunoprotección frente a la infección en modelos animales (Lasso et al., 2010; Marañón et al., 2001; Planelles et al., 2002; Thomas et al., 2001). Del mismo modo, se ha descrito que los pacientes con Chagas presentan una notable respuesta humoral frente a KMP11, con anticuerpos, principalmente, de isotipo IgG1 (Flechas et al., 2009). Es reconocida tanto por el suero de pacientes infectados por *T. cruzi* como infectados por *Leishmania* (Thomas et al., 2001). No obstante, los sueros de pacientes con Chagas reconocen epítopes lineales contenidos en KMP11 mientras que los sueros de pacientes con

leishmaniasis reconocen predominantemente epítopes conformacionales contenidos en la proteína (Thomas et al., 2001).

- **HSP70** (*Heat Shock Protein 70 kDa*)

HSP70 es una proteína de choque térmico de 70 kDa, perteneciente a la familia de las chaperonas presentes en todo tipo de organismos. Entre las funciones biológicas que desempeñan se encuentran el plegamiento y traslocación de proteínas, unión a proteínas evitando la generación de agregados, así como el control de la actividad de proteínas reguladoras (Mayer and Bukau, 2005). Se ha descrito que la expresión de HSP70 aumenta hasta 4 veces tras un choque térmico a 37°C (Requena et al., 1992).

HSP70 tiene una alta capacidad adyuvante/inmunomoduladora, asociada a su capacidad para interactuar con células dendríticas y macrófagos (Singh-Jasuja et al., 2001). Así, se ha descrito que la inmunización de ratones con un vector vacuna que porta el gen KMP11 de *T. cruzi* (Thomas et al., 2000) fusionado al codificante para la proteína HSP70 provoca un descenso en la parasitemia y otorga protección frente a la infección por *T. cruzi*, generándose anticuerpos IgG2a, específicos para KMP11 y activación de linfocitos T citotóxicos (CTL) frente a péptidos contenidos en el antígeno (Planelles et al., 2001). Además, se ha publicado que la HSP70 de *T. cruzi*, tanto de forma individual como unida a KMP11, así como un fragmento de la HSP70, puede madurar células dendríticas humanas y murinas (Cuellar et al., 2008; Planelles et al., 2002). Adicionalmente, en la secuencia de HSP70 de *T. cruzi* se han identificado 4 epítopes inmunodominantes siendo, además, 2 de ellos reconocidos por CTLs de pacientes con enfermedad de Chagas (Marañón et al., 2011).

- **PFR2** (*Paraflagellar Rod protein 68 kDa*)

PFR2 pertenece al grupo de proteínas paraflagelares, las cuales se localizan a lo largo del flagelo y en el bolsillo flagelar de los kinetoplastos, euglenoides y dinoflagelados presentando un alto grado de conservación entre especies de tripanosomátidos.

Las proteínas paraflagelares se han descrito como altamente inmunogénicas, desempeñando un papel esencial en la inducción de inmunidad protectora (Luhrs et al., 2003; Miller et al., 1997; Morell et al., 2006). Así, la inmunización de ratones con las proteínas PFR2 y PFR3, induce a una respuesta protectora (Wrightsmann and

Manning, 2000). Además, pudo observarse que los ratones infectados poseían células T CD8<sup>+</sup> ante los epítopes presentes en esta proteína y que la inmunización con vacunas que portan el gen que codifica PFR2 es protectora (Morell et al., 2006).

En los últimos años, se han descrito epítopes contenidos en las proteínas PFR2 y PFR3 que son reconocidos por células T CD8<sup>+</sup> de pacientes con enfermedad crónica de Chagas en fase indeterminada (IND) y con sintomatología cardíaca (CCC), así como actividad citotóxica por parte de las células T CD8<sup>+</sup> epítipo-específicas en pacientes de Chagas IND (Egui et al., 2012). Recientemente, un estudio realizado en pacientes de Chagas crónico en fase indeterminada ha mostrado que todos poseían linfocitos T CD8<sup>+</sup> antígeno-específicos para epítopes específicos contenidos en las proteínas PRF2, así como en KMP11, y que la proporción de estas células disminuía ligeramente tras el tratamiento con benznidazol (Egui et al., 2020).

- **3973<sup>d</sup>**

3973 es un péptido contenido en la proteína de membrana de *T. cruzi* TcCA-2. La proteína TcCA-2 posee un tamaño de 85 kDa y una alta capacidad inmunogénica, expresándose en los estadios epimastigote y tripomastigote. Esta contiene el epítipo TcMe (específico de *T. cruzi*) implicado en la internalización del parásito a la célula huésped (Martín et al., 1997).

El péptido 3973 (FGQAAAGDKPSL) es reconocido con una elevada sensibilidad (> 90%) por el suero de pacientes con Chagas crónico, mientras que no es reconocido por el suero de pacientes en la fase aguda de la enfermedad (Thomas et al., 2012). El péptido 3973 presenta una alta especificidad (> 98%), pues no es reconocido por sujetos con procesos autoinmunes e inflamatorios o por individuos con una cardiomiopatía no chagásica. Del mismo modo, no se dan reacciones cruzadas con sueros de pacientes infectados con *Leishmania* u otras enfermedades infecciosas similares (Thomas et al., 2012). El péptido 3973<sup>d</sup>, hace referencia a dos secuencias del péptido 3973 unidas por dos glicinas (FGQAAAGDKPSLGGFGQAAAGDKPSL).

### 1.8.2. Biomarcadores de patología o progresión de la enfermedad

Hasta la fecha se han descrito diferentes biomarcadores que podrían resultar útiles para determinar la patología o progresión de la enfermedad de Chagas. Entre ellos se encuentran los péptidos natriuréticos, como el NTproBNP el cual se ha correlacionado con la parasitemia encontrándose elevado en pacientes CCC (Keating et al., 2015). Del mismo modo, las proteínas CRA (antígenos del citoplasma de las formas epimastigote y amastigote) y FRA (antígenos del flagelo de las formas tripomastigote y epimastigote) también se han propuesto como biomarcadores de las distintas fases de la enfermedad. Así, la frecuencia de anticuerpos de isotipo IgG2 frente a FRA detectado es diferencial entre los pacientes IND y CCC (Verçosa et al., 2007). Por otra parte, el nivel de anticuerpos de isotipo IgA frente a ambas proteínas permitió diferenciar pacientes asintomáticos de aquellos con sintomatología cardíaca y digestiva (Vasconcelos et al., 2010).

El péptido 3973<sup>d</sup>, descrito en el apartado anterior, también se ha propuesto como indicador del estadio patológico de la enfermedad, pues la reactividad frente al mismo detectada en los sueros de pacientes crónicos asintomáticos con enfermedad de Chagas, fue en promedio un 40% menor que la detectada en los sueros de pacientes que presentaban alteraciones cardíacas o digestivas (Thomas et al., 2012).

Por otro lado, los anticuerpos monoclonales 5F2 y 5A9B11 frente a la fracción microsomal de *T. cruzi* (Mc) y su reactividad cruzada con tejidos de mamíferos se utilizaron para estudiar las diferencias en la respuesta inmunitaria frente los epítopes del parásito de pacientes con enfermedad de Chagas crónica con diferentes grados de disfunción cardíaca. La reactividad entre 5F2 o 5A9B11 y sus antígenos diana fue inhibida significativamente por los sueros de individuos con la enfermedad de Chagas en comparación con los sueros de donantes sanos. Además, el anticuerpo 5F2 permitió distinguir los sueros de pacientes en la fase inicial de la forma sintomática de la de los pacientes con enfermedad grave ( $p < 0,05$ ), sugiriendo el potencial uso de anticuerpos anti-Mc como marcador de pronóstico (Laucella et al., 1996).

El estudio del perfil de secreción de citoquinas, así como los estudios fenotípicos de las células T de los pacientes en las diferentes formas clínicas de la



enfermedad, también ha puesto de manifiesto su potencial como biomarcadores de progresión. Un ejemplo serían las citoquinas IFN- $\gamma$ , IL4 y TNF- $\alpha$  propuestas como marcadores inmunológicos de la forma clínica cardíaca de la enfermedad de Chagas. La estimulación de células de pacientes IND y CCC con las proteínas CRA y FRA, reveló que el nivel de IFN- $\gamma$  y TNF- $\alpha$  frente a la proteína CRA era superior en pacientes CCC (Lorena et al., 2010). Por otro lado, el estudio del fenotipo de las células T CD8<sup>+</sup> circulantes reveló que en el caso de células T CD8<sup>+</sup> específicas de los epítopes PFR y TcCA-2, se observó un mayor nivel de IFN- $\gamma$  en pacientes de Chagas asintomáticos frente a los que presentan sintomatología cardíaca (Egui et al., 2012, 2015). Además, sólo se detectó actividad citotóxica por parte de las células T CD8<sup>+</sup> epítotope-específicas de los pacientes asintomáticos (Egui et al., 2012). Curiosamente, se observó un perfil fenotípico diferencial de las células T CD8<sup>+</sup> específicas de TcCA-2 de pacientes con síntomas cardíacos, las cuales eran principalmente células efectoras de memoria (T<sub>EM</sub> y T<sub>EMRA</sub>), mientras que las presentes en los pacientes IND eran predominantemente células *naive* (T<sub>NAIVE</sub>). Asimismo, en los pacientes CCC, el porcentaje de células T CD8<sup>+</sup> con características de senescencia (CD8<sup>+</sup>/CD44<sup>+</sup>/CD57<sup>+</sup>) fue significativamente mayor que el detectado en los pacientes en la fase asintomática de la enfermedad (Egui et al., 2015).

Al igual que con los biomarcadores de eficacia del tratamiento, en la actualidad, el desarrollo de biomarcadores de patología o progresión de la enfermedad sigue siendo una necesidad y supondría un importante logro para mejorar el tratamiento clínico de los pacientes con enfermedad de Chagas.

## 2. ELEMENTOS MÓVILES DE ADN

Los elementos móviles o elementos transponibles son secuencias de ADN con capacidad de saltar de un sitio a otro del genoma, generando frecuentemente nuevas copias de sí mismos, lo que les permite expandirse a través de los genomas. Barbara McClintock los descubrió mientras estudiaba la herencia en la pigmentación de los granos de maíz en los años 50, hallazgo que le otorgó el Nobel de Medicina en 1983. Estos han sido catalogados como elementos dinámicos que remodelan el genoma del hospedador mediante la generación de reordenamientos, con el potencial de crear o alterar genes, barajar genes existentes, así como modular sus patrones de expresión (Thomas et al., 2010).

### 2.1. Clasificación de los elementos móviles de ADN

La primera clasificación de los elementos transponibles fue propuesta por Finnegan en 1989. Se basó en el mecanismo de transposición, así se dividieron en dos clases: elementos de Clase I, que se transponen mediante la retrotranscripción de un intermediario de ARN, y los de Clase II, que se transponen directamente de ADN a ADN. Dentro de la Clase I, se definieron a su vez dos subclases: elementos LTR (*Long Terminal Repeats*), con largas secuencias terminales repetidas (semejantes a los retrovirus) y elementos no-LTR que carecían de estas repeticiones, pero poseían una secuencia rica en adenina en la posición 3' (no semejantes a los retrovirus). Posteriormente, se propusieron otras clasificaciones más detalladas, pero fue en 2006 cuando finalmente se constituyó, durante la primera conferencia internacional sobre elementos transponibles, un Comité Internacional para la clasificación de los elementos transponibles (Kapitonov and Jurka, 2008) cuya misión era la de establecer una clasificación actualizada conservando la configuración principal de la propuesta por Finnegan. Acorde a esta clasificación, los elementos transponibles pueden clasificarse en dos tipos: transposones de ADN y retrotransposones, dentro de este último tipo distinguiéndose retrotransposones LTR y retrotransposones no-LTR.

### 2.1.1. Retrotransposones LTR

Los elementos transponibles del tipo retrotransposones LTR son semejantes en estructura a los retrovirus y poseen regiones solapadas de 3 genes diferentes: *gag*, proteasa *prt* y polimerasa *pol*; a veces presente también un dominio *env*. El gen *pol* presenta dominios para una retrotranscriptasa (RT), una RNasa H, y una endonucleasa (EN).

### 2.1.2. Retrotransposones no-LTR

Los retrotransposones que carecen de LTR presentan una estructura constituida por una región 5' no traducida (UTR) que desempeña función promotora, uno o dos marcos abiertos de lectura (ORF1 y ORF2; *Open Reading Frames*) con una región espaciadora, y una región UTR en 3' conteniendo una cola de poli(A). Dentro de la clase pueden diferenciarse *clades*, agrupados filogenéticamente según la secuencia de su retrotranscriptasa (Malik et al., 1999) y que, además, comparten las mismas características estructurales como el tipo de endonucleasa, la presencia de ribonucleasa H o motivos de reconocimiento de ARN o dedos de zinc.

Los elementos transponibles se clasifican como retroelementos autónomos si portan secuencias con capacidad teórica para codificar las proteínas necesarias para su propia transposición (denominados LINE, *Long Interspersed Nuclear Element*) y elementos no autónomos que carecen de dicha capacidad codificante (denominados SINE, *Short Interspersed Nuclear Elements*). Los elementos LINE están presentes en un alto número de copias dispersas por el genoma de diferentes organismos. Presentan un tamaño de 4 a 6 kb, aunque también se han descrito copias truncadas en los extremos 5' y 3'. Estos elementos están flanqueados por secuencias de duplicación directa (TSD). Es el caso de los elementos SINE, en los que se ha mostrado que sus transcritos pueden ser reconocidos por la maquinaria de retrotransposición de elementos LINE localizados en el mismo hospedador, pudiendo ser, por tanto, movilizados por las proteínas codificadas por los elementos LINE (Dewannieux et al., 2003; Kajikawa and Okada, 2002; Weiner, 2002). Estos dos elementos, LINE y SINE, están presentes en casi todos los genomas de eucariotas y juntos representan aproximadamente un 34% del genoma humano (Weiner, 2002).

### 2.2. Abundancia en los genomas de los elementos móviles de ADN

Todos los organismos conocidos presentan elementos móviles en su genoma, con la excepción de *Bacillus subtilis* (Kunst et al., 1997) y *Plasmodium falciparum* (Gardner et al., 2002). Entre especies, se observa una gran variación en la abundancia de estos elementos, existiendo una correlación entre el tamaño del genoma y la abundancia de elementos transponibles (Kidwell, 2002). En cuanto a la abundancia relativa de cada clase de elementos, son los retrotransposones LTR y no-LTR, los que se encuentran en mayor proporción, siendo los principales responsables del aumento o disminución del tamaño de los genomas (Boulesteix et al., 2006; Sen et al., 2006; Vitte and Bennetzen, 2006). Alrededor del 45% del genoma humano corresponde a elementos móviles y secuencias repetidas, gran parte de este porcentaje correspondiente a retroelementos no-LTR, siendo LINE y SINE los más representativos por el elevado número de copias presentes (Lander et al., 2001).

### 2.3. Efecto de los elementos transponibles sobre el genoma del hospedador

A lo largo del tiempo, se ha reconocido el gran potencial de los elementos transponibles para generar diversidad genética, pudiendo ocasionar reordenaciones cromosómicas, mutaciones espontáneas, recombinación y disgénesis híbrida (Engels, 1979; Mathiopoulos et al., 1998).

Uno de los efectos provocados por los elementos transponibles es la movilización de otras secuencias génicas y la formación de pseudogenes. En 2000, Esnault y colaboradores mostraron que los LINE humanos son capaces de movilizar transcritos de ADN ajenos al propio elemento, pudiendo representar esta formación de pseudogenes procesados mediante la retrotranscripción de LINE un importante mecanismo para la duplicación génica (Esnault et al., 2000). Además de su implicación en procesos de mutagénesis o reordenaciones cromosómicas, algunos elementos se utilizan como vehículos para movilizar regiones flanqueantes del genoma. De esta forma, estas secuencias génicas, exones o promotores, pueden ser transducidas o insertadas en un nuevo gen proporcionando un mecanismo de evolución génica.

#### 2.4. Elementos transponibles en tripanosomátidos: Importancia de los elementos LINE y SINE

La mayoría de los genomas de tripanosomátidos secuenciados hasta la fecha contienen un gran número de retrotransposones que representan hasta un 5% del genoma nuclear (Bringaud et al., 2015; El-Sayed, Myler, Bartholomeu, et al., 2005). Los retrotransposones más abundantes en tripanosomátidos son los no-LTR, los cuales son capaces de movilizarse por un mecanismo conocido como TPRT (*Target-Primed Reverse Transcription*) (Eickbush and Jamburuthugoda, 2008; Luan et al., 1993) generando reordenamientos con el potencial de interrumpir o crear genes, así como reorganizar genes existentes y/o modificar sus patrones de expresión (Thomas et al., 2010). Además, se han identificado versiones degeneradas largas y cortas (DIREs y SIDERs) en la mayoría de los genomas de tripanosomátidos (Bringaud, Ghedin, et al., 2006).

El estudio comparativo de los genomas de *T. brucei*, *T. cruzi* y *L. major*, mostró que estas especies han mantenido un alto nivel de conservación en la localización de los genes (sintenia), lo cual sugiere que la presión selectiva ha conservado el orden génico cientos de millones de años de evolución (El-Sayed, Myler, Blandin, et al., 2005). En los pocos lugares donde hay reordenamientos genómicos entre estas especies, pueden encontrarse retrotransposones o retrotransposones muy degenerados, lo que sugiere el importante papel que estos elementos transponibles pueden haber desempeñado en la evolución y generación de la actual estructura genómica de los tripanosomátidos (El-Sayed, Myler, Blandin, et al., 2005).

Los genomas de tripanosomátidos están organizados en grandes grupos de genes policistrónicos direccionales que se transcriben por la ARN polimerasa II y están separados por regiones no codificantes de cambio de cadena, denominados puntos de inflexión transcripcional o SSR (*Strand Switch Region*). Los lugares donde se pierde la sintenia están relacionados, mayoritariamente, con estos puntos de inflexión transcripcional o con regiones subteloméricas. Las regiones de inflexión transcripcional que han sido analizadas presentan retrotransposones no-LTR, DIREs y/o la familia de pseudogenes RHS (*Retrotransposon Hot Spot*) caracterizada por poseer puntos calientes para la inserción de estos elementos, lo que sugiere que los retrotransposones pueden haber desempeñado un papel importante en la

configuración de la organización del genoma de los tripanosomátidos (Ghedin et al., 2004). Estas regiones se localizan entre los grupos de genes codificados por cadenas opuestas y se postula que es donde se inicia la transcripción específica en ambas direcciones (Clayton, 2002; Martínez-Calvillo et al., 2004). Por otro lado, las regiones subteloméricas son zonas no sinténicas con un elevado grado de variabilidad a través de la pérdida de nucleótidos y de recombinaciones (Wickstead et al., 2003). En las regiones subteloméricas de *T. brucei* y *T. cruzi* se encuentran repeticiones en tándem de genes que codifican para importantes familias de proteínas de superficie (El-Sayed, Myler, Blandin, et al., 2005). Se ha sugerido que dicha localización facilita modificaciones en las proteínas expresadas en la superficie del parásito generando una gran variabilidad, lo que favorece la evasión de la respuesta inmunitaria del hospedador (Cano, 2001; Niang et al., 2011). Las regiones subteloméricas de *T. cruzi* se caracterizan por un ensamblaje polimórfico de la superfamilia de las trans-sialidasas (Frasch, 2000), de DGF-1 (*Dispersed Gene Family-1*) (Wincker et al., 1992), así como de retrotransposones; VIPER/SIRE, L1Tc y/o retroelementos degenerados (DIREs) y de genes o pseudogenes de la familia RHS (Bringaud, Biteau, et al., 2002). Otra peculiaridad del genoma de *T. cruzi*, es el contenido en grandes “islas” no sinténicas (de hasta 600 kb) de genes codificantes de proteínas de superficie como trans-sialidasas, peptidasas GP63, mucinas y proteínas de superficie asociadas a mucinas (MASP), conteniendo también un alto número de retrotransposones y genes RHS (El-Sayed, Myler, Blandin, et al., 2005).

Los elementos LINE en tripanosomátidos también podrían participar en mecanismos de reparación cromosómica. De hecho, en *T. cruzi* la sobreexpresión de la endonucleasa AP codificada por el elemento L1Tc (*Long Interspersed Nucleotide Element from T. cruzi*) le proporciona resistencia al daño en el ADN ocasionado por daunorubicina y por radiación  $\gamma$ , suponiendo un decremento de hasta el 60% en las roturas del ADN generadas por el tratamiento, aumentando la supervivencia del parásito (Olivares et al., 2003).

Por todo ello, los elementos transponibles generan una gran variabilidad en el genoma hospedador, haciendo que este sea más flexible a diferentes condiciones externas. Esto podría explicar el elevado número de copias de retrotransposones, mayoritariamente SINE y LINE, presentes en los genomas de tripanosomátidos

como *T. cruzi*, cuyo ciclo de vida digénico requiere su adaptación a diferentes entornos.

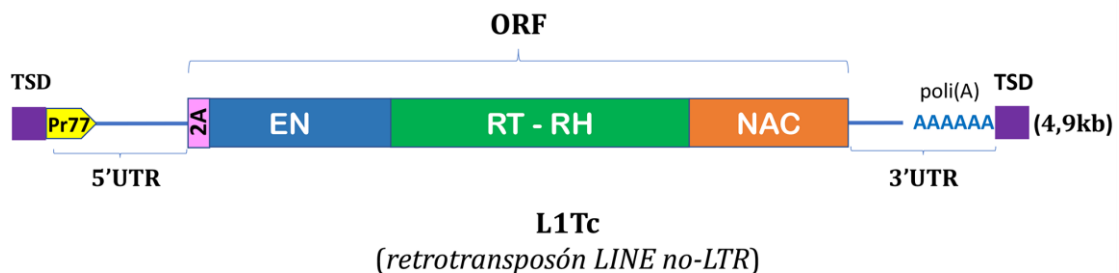
### **2.5. L1Tc: Elemento LINE de *Trypanosoma cruzi***

L1Tc es el retrotransposón no-LTR (LINE) de *T. cruzi* mejor caracterizado en tripanosomátidos (Martin et al., 1995; Thomas et al., 2010) y el más representado en el genoma de *T. cruzi*, estando presente en la mayoría, si no en todos, los cromosomas de diferentes cepas del parásito (Martin et al., 1995; Olivares et al., 2000). L1Tc se transcribe en los tres estadios del parásito y posee una longitud de aproximadamente 5 kb. Presenta características propias de este grupo como estar flanqueado por secuencias de duplicación directa y contener una cola poli(A) en su extremo 3' (Bringaud, Bartholomeu, et al., 2006; Martin et al., 1995; Olivares et al., 2000).

El estudio de la distribución genómica y organización de L1Tc de *T. cruzi* ha revelado que existe un elevado número de copias de este retroelemento en todas las cepas del parásito, existiendo copias únicas y agrupadas en tándem con orientación cabeza-cola (Olivares et al., 2000) y que pueden encontrarse asociados a elementos NARTc y a familias de genes codificantes para proteínas de superficie (El-Sayed, Myler, Blandin, et al., 2005; Olivares et al., 2000). Como se ha indicado anteriormente, se ubica fundamentalmente en puntos de inflexión transcripcional y regiones subteloméricas, frecuentemente insertado en la familia de pseudogenes RHS (Bringaud, Biteau, et al., 2002; Ghedin et al., 2004).

Las dos primeras copias de L1Tc identificadas correspondían a una secuencia de ADNc que contenía tres ORFs, así como otra copia que poseía un único ORF, portando ambas, secuencias que codificaban proteínas con dominios conservados de la AP endonucleasa, transcriptasa inversa y chaperona de ácidos nucleicos similar a los factores de transcripción de tipo TFIIIA (Martin et al., 1995). Mediante el clonaje de dichas secuencias en vectores de expresión y la producción de las diferentes proteínas recombinantes, se demostró experimentalmente que estas secuencias codificaban para proteínas con función AP endonucleasa (Olivares et al., 1997), reverso transcriptasa (García-Pérez et al., 2003), RNasa H en aquellas copias que contenían 1 ORF (Olivares et al., 2002) y proteína chaperona de ácidos nucleicos

(Heras et al., 2005). Corriente arriba de la secuencia codificante de la AP endonucleasa se encuentra una secuencia que codifica el dominio consenso 2A (-DxExNPGP-) de las secuencias 2A autocatalíticas (Heras et al., 2006), la cual se mantiene conservada en virus de pequeño tamaño (Donnelly et al., 2001). Esta secuencia autocatalítica, llamada L1Tc2A, mostró ser activa empleando construcciones bicistrónicas tanto en ensayos *in vitro* como *in vivo* al digerir entre la G y la P de la secuencia 2A. De igual manera, se demostró que la actividad de esta secuencia influía en la abundancia relativa de los productos de traducción que se encontraban corriente arriba y abajo de ella (Heras et al., 2006) (**Figura 18**). Se ha postulado que todas estas actividades constituyen la maquinaria enzimática para la movilización por sí misma de L1Tc (Heras et al., 2009; Macías et al., 2018).

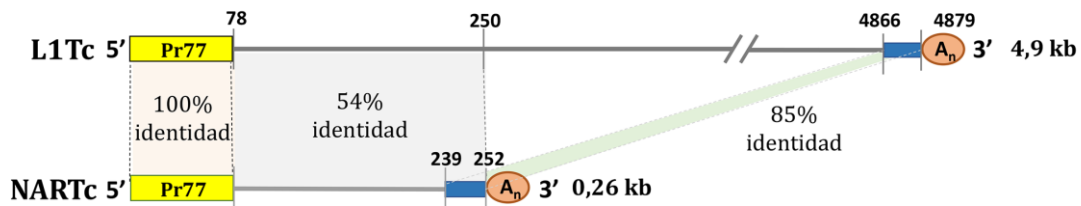


**Figura 18. Representación esquemática del retroelemento L1Tc de *T. cruzi*.** El elemento representado contiene un solo ORF que incluye los dominios endonucleasa AP (EN, azul), transcriptasa inversa y RNAsa (RT - RH, verde) y chaperona de ácidos nucleicos (NAC, naranja). Los cuadrados morados flanqueando el elemento representan las duplicaciones directas de 9 a 12 nucleótidos (TSD). Se indican las regiones 5' UTR y 3' UTR. En amarillo, se indica la secuencia Pr77 con función dual promotor-ribozima HDV. El rectángulo rosa representa la secuencia autocatalítica 2A.

### 2.5.1. NARTc: Versión truncada de L1Tc en el genoma de *T. cruzi*

La versión truncada de L1Tc en *T. cruzi* se denomina NARTc (*Non-Autonomous Retrotransposon in T. cruzi*). Estos retroelementos poseen un tamaño de 0,26 kb y no tienen capacidad codificante por lo que se estima que utilizan *in trans* la maquinaria enzimática de L1Tc para su transposición. La comparación del extremo 5' de NARTc y L1Tc mostró que los primeros 77 nt de ambos retroelementos son 100% idénticos, manteniéndose un 54% de identidad en el resto de la secuencia. Los 13 nucleótidos localizados corriente arriba de la cola poli(A) mantienen una identidad del 85% (Bringaud, García-Pérez, et al., 2002) (**Figura 19**).





**Figura 19. Representación esquemática y comparación de los retrotransposones L1Tc y NARTc de *T. cruzi*.** Los primeros 77 nucleótidos (cajas amarillas) son idénticos entre ambos elementos, mientras que los siguientes 175 nucleótidos de NARTc presentan una identidad del 54% con la secuencia de L1Tc. En la región 3' precedente a las colas poli(A) (círculos naranjas), existe una región de 13 nt con una identidad del 85% entre ambos elementos (cajas azules). *Adaptada de* Bringaud, García-Pérez, et al., 2002.

## 2.6. Transcripción del elemento L1Tc

Puesto que la transcripción constituye el primer paso necesario para la movilización de estos elementos, se postuló que, para mantener la autonomía desde el punto de vista de la movilidad, los elementos LINE debieran portar un promotor interno que transcribiera la copia del elemento que fuera a retrotransponerse e insertarse en un nuevo lugar del genoma. El alto grado de identidad de la secuencia de 77 nucleótidos localizados en el extremo 5' de las diferentes copias L1Tc y NARTc, sugería que el promotor interno estuviera localizado en esta secuencia. Mediante el clonaje de esta secuencia corriente arriba de un gen reportero en un vector de transformación para *T. cruzi* y la generación de transformantes estables, se puso de manifiesto que la secuencia Pr77 tenía actividad promotora y generaba abundantes transcritos que eran traducibles y carecían de SL por lo que se denominó promotor Pr77 (Heras et al., 2007). Los transcritos generados por Pr77 se inician en el nucleótido +1 del retroelemento y presentan una vida media de aproximadamente 1 hora (Heras et al., 2007).

Pr77 porta una secuencia de 4 nucleótidos 'CGTG', denominada motivo DPE (*Downstream Promoter Elements motif*) (Macías et al., 2016) encontrándose en Pr77 en las posiciones +25 a +28 del nucleótido +1 en el ARNm de L1Tc. Este DPE se conserva en términos de composición de la secuencia y localización respecto al sitio de inicio de la transcripción (TSS, *Transcription Start Site*) en el 99% de secuencias Pr77 de L1Tc con alto grado de identidad. Se ha descrito que el motivo DPE de Pr77 participa en la transcripción de Pr77 (Macías et al., 2016), así como los primeros nucleótidos de Pr77 y otros nucleótidos localizados a lo largo de la secuencia de

Pr77 (Macías et al., 2018), estando implicados en la unión de los factores de transcripción TBP (TATA-box binding protein) y SNAP50 (*Small Nuclear RNA-Activating Protein-50 kDa*) y otras proteínas nucleares (Macías et al., 2021).

El alto grado de identidad de la secuencia de 77 nucleótidos colocados al inicio de las diferentes copias L1Tc y NARTc, mostró estar conservada en cuanto a posición y composición de secuencia en los elementos *ingi*, así como en aquellos elementos degenerados de *T. cruzi* y *T. brucei* los cuales carecían de capacidad codificante al haber acumulado gran cantidad de mutaciones que generaban codones de parada a lo largo de las diferentes secuencias. Es por ello que a esta secuencia de 77 nt se la denominó huella Pr77 o firma Pr77 (Bringaud, Ghedin, et al., 2006).

En *T. brucei* existen copias homólogas a L1Tc en el elemento denominado *ingi* constituyendo ambos elementos el *clade ingi/L1Tc*. Se han identificado elementos homólogos a L1Tc y a *ingi*, así como versiones truncadas de estos retrotransposones en genomas de tripanosomátidos como *T. cruzi* (NARTc), *T. brucei* (RIME), *Trypanosoma vivax* y *Trypanosoma congolense* (Tvingi, L1Tco, y Tcoingi; , TvRIME y TcoRIME) (Bringaud et al., 2009; Bringaud, García-Pérez, et al., 2002).

### **2.7. Función ribozima de la secuencia de ARN de Pr77. Función dual de la huella Pr77: promotor (a nivel de ADN) y ribozima (a nivel de ARN)**

Análisis de la estructura secundaria de la secuencia de ARN codificada por la huella Pr77 puso de manifiesto que esta adoptaba una estructura secundaria compatible con la de las ribozimas de tipo HDV (*Hepatitis Delta Virus*). Mediante ensayos *in vitro* co-transcripcionales y post-transcripcionales se ha puesto de manifiesto que esta secuencia posee actividad ribozima de tipo HDV por lo que se la ha calificado L1TcRz (Sánchez-Luque et al., 2012). L1TcRz actúa postranscripcionalmente produciendo un corte en el extremo 5' del ARN en el nucleótido anterior al que forma parte de la secuencia que adopta la estructura, justo por encima del nucleótido +1 de L1Tc y NARTc (Sánchez-Luque et al., 2011). De esta manera, permite que los ARNm transcritos de estos retroelementos conserven intacta su secuencia promotora y puedan transcribirse de forma autónoma tras ser transcritos de forma inversa e integrados en una nueva posición

genómica. Por tanto, a la secuencia Pr77 se la ha calificado como un sistema dual (Sánchez-Luque et al., 2012) al actuar como promotor, a nivel de ADN, y como ribozima HDV a nivel de ARN. La implicación de estas actividades en la regulación de la expresión génica de los parásitos que las hospedan podría garantizar la escisión y consecuente liberación de una copia activa del elemento que se haya cotranscrito como parte de un ARN policistrónico, a la vez que contribuiría al mantenimiento de su autonomía desde un punto de vista transcripcional, la transcripción de las copias presentes en orientación antisentido con respecto a los *clusters* policistrónicos y las que se encuentran fuera de ellos, así como podría estar impidiendo un decaimiento del nivel de transcripción de las regiones distales, asegurando la correcta expresión de los últimos genes de cada *cluster* (Sánchez-Luque et al., 2012).

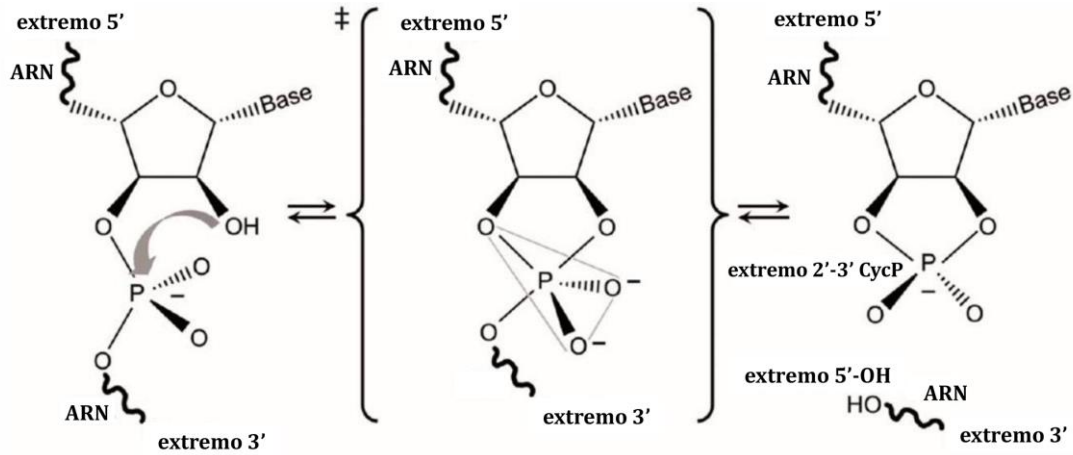
La secuencia Pr77 se encuentra altamente conservada en el extremo 5' de otros retrotransposones que residen en el genoma de *T. cruzi* y de otros tripanosomátidos como *Trypanosoma vivax*, *Trypanosoma congolense* y *Trypanosoma brucei*, así como en retrotransposones degenerados, tanto de estas especies como de la especie *Leishmania* (Bringaud et al., 2007, 2009). Por ello, se ha sugerido que los retroelementos L1Tc y aquellos portadores de la huella Pr77, son los responsables de expandir la secuencia dual Pr77 en los genomas de los tripanosomátidos (Sánchez-Luque et al., 2014).

## 2.8. Ribozimas de tipo HDV

Las ribozimas HDV se describieron por primera vez en el ARN circular del virus de la hepatitis delta (Kuo et al., 1988). Se definen como moléculas pequeñas de ARN autocatalíticas. El plegamiento de su ARN da lugar a la catálisis de una reacción de transesterificación, en la que tiene lugar la escisión de su propia cadena principal de fosfodiéster en *cis*, generándose como consecuencia un producto con 2', 3' fosfato cíclico y otro producto con un 5' hidroxilo en su extremo (Sharmeen et al., 1988) (**Figura I10**). La reacción *in vitro* no requiere de proteínas ni de factores celulares pero la presencia de un ion de metal divalente estimula las ratios de escisión (Wu et al., 1989). De esta forma, se ha descrito que los iones de metales divalentes, además de estabilizar la estructura general, parecen tener un papel catalítico directo en las

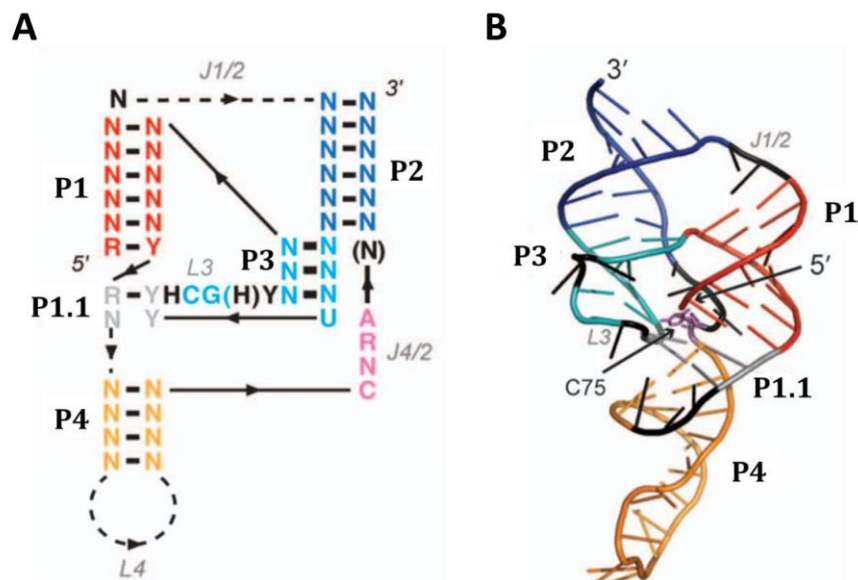
## INTRODUCCIÓN

reacciones de escisión de las ribozimas HDV (Been, 2006). Así, existe evidencia significativa de que un ion  $Mg^{2+}$  participa en la reacción de escisión de la ribozima HDV en condiciones de tampón relevantes biológicamente (Chen et al., 2013).



**Figura I10. Mecanismo de la reacción de transesterificación del ARN catalizado por las ribozimas HDV autocatalíticas.** Los productos de escisión son un fosfato 2'-3'-cíclico en el producto de ARN 5' y un 5'-hidroxilo en el producto de ARN 3'. *Adaptado de De la Peña et al., 2017.*

La estructura secundaria de estas ribozimas consiste en un doble pseudonudo anidado que incluye las regiones helicoidales pareadas P1.1, P1, P2, P3 y P4 (Shih and Been, 2002; Webb and Lupták, 2011) (**Figura I11**).



**Figura I11. Estructura secundaria consenso (A) y cristalina (B) de ribozimas tipo-HDV.** Las hélices P1 y P4 representan regiones de covariación de pares de bases; J1/2, J4/2 y L3 son regiones de cadena simple. N: indica cualquier nucleótido, R: purina, Y: pirimidina, H: adenina, C: citosina o uracilo. Las líneas discontinuas representan inserciones de secuencia de longitud variable. Las líneas continuas representan conexiones directas sin inserciones de nucleótidos. *Adaptada de Webb and Lupták, 2011.*

El sitio de escisión se asigna al primer nucleótido de la hélice P1, típicamente en un residuo de guanosina. El núcleo catalítico está formado por regiones monocatenarias L3 y J4/2, y los nucleótidos de la base de las hélices P2-P3. Es por ello que la orientación correcta de los segmentos helicoidales P1 y P2-P3 es crucial para la actividad catalítica (Webb et al., 2016). Solo unas pocas posiciones en la secuencia son invariantes entre las que se encuentra el par de bases guanosina-pirimidina en la base de la hélice P1. Las hélices P2 y P4 tienden a mostrar más variaciones de longitud, desapareamientos e inserciones (Webb and Lupták, 2011). Además, la hélice P4 no es absolutamente necesaria para la actividad, aunque parece estabilizar la ribozima activa (Been et al., 1992; Thill et al., 1993).

Durante muchos años, se ha creído que las ribozimas autocatalíticas de tipo HDV eran raras en la naturaleza y que podían encontrarse solo en ARNs del virus HDV (Chen et al., 2013). Sin embargo, las ribozimas de tipo HDV, y probablemente las ribozimas en general, se distribuyen ampliamente en la naturaleza y pueden desempeñar variadas funciones biológicas (Webb et al., 2009). Es por ello que no están limitadas a este organismo, siendo también identificadas en plantas, insectos, protozoos, bacterias, hongos, e incluso en mamíferos como es el caso de la identificada en el intrón del gen *CPEB3* el cual forma parte de la familia de genes implicados en la regulación de la poliadenilación del ARNm (Salehi-Ashtiani et al., 2006). Estas ribozimas, además, se han localizado en retrotransposones. Así, en 2010, se localizó una ribozima activa de tipo HDV en el elemento R2, presente en el genoma de *Drosophila simulans*, concretamente en el ARN de la región 5' UTR (D. G. Eickbush and Eickbush, 2010). El retrotransposón R2 es sitio-específico, integrándose las copias del mismo específicamente en genes del ARN ribosómico 28S. De esta forma, el elemento se cotranscribe junto a estos genes y la ribozima que contiene es la responsable de la liberación del transcrito del elemento del ARNr 28S.

L1TcRz fue la primera ribozima de tipo HDV descrita en tripanosomátidos, la segunda reportada en un retroelemento no-LTR y la cuarta caracterizada en profundidad (Sánchez-Luque et al., 2011). Además de en los retroelementos L1Tc y NARTc (Sánchez-Luque et al., 2011), se han encontrado ribozimas activas de tipo HDV asociadas a Pr77 en otros elementos móviles de tripanosomátidos, como en sus homólogos en *Trypanosoma congolense* (L1Tco y NARTco), así como en elementos

SIDER de *Trypanosoma vivax* y de la especie *Leishmania* como *L. infantum*, *L. donovani*, *L. major* y *L. mexicana* (Sánchez-Luque et al., 2014) por lo que se ha sugerido que sean los retroelementos los responsables de expandir este sistema dual en tripanosomátidos.

### **3. RESPUESTA INMUNE FRENTE A LA INFECCIÓN POR *T. cruzi***

La infección causada por *T. cruzi* desencadena múltiples mecanismos inmunitarios en el hospedador para combatir al parásito. Estos actúan a nivel innato y adaptativo, así como a escala humoral y celular (Tarleton, 2007), los cuales se caracterizan principalmente por procesos de proliferación celular, generación de citoquinas e inducción de mecanismos de muerte celular (de Meis et al., 2009). La disminución o ausencia de cualquier mecanismo efector de la respuesta innata o adaptativa conduce a un aumento de la parasitemia y una mayor susceptibilidad a enfermedades (Budzko et al., 1975; Kierszenbaum et al., 1983; Kumar and Tarleton, 1998; Rottenberg et al., 1992; Sardinha et al., 2006; Silva et al., 1992; Tarleton, 1990). Es por ello que la pérdida de equilibrio entre la respuesta del sistema inmunitario y la replicación del parásito es crucial para la progresión de la enfermedad (dos Santos Virgilio et al., 2014).

#### **3.1. Respuesta inmunológica innata**

La respuesta inmunológica innata tiene un importante papel en la eliminación rápida de patógenos. En la infección por *T. cruzi*, se considera una barrera esencial que controla la replicación del parásito, así como su propagación a los tejidos. Esta respuesta se basa en el reconocimiento de patrones moleculares asociados a patógenos (PAMP), los cuales están presentes en diversos organismos, pero ausentes en el huésped, actuando como una señal exógena que alerta al huésped de la presencia de patógenos (Schenten and Medzhitov, 2011).

Las células dendríticas, NK y macrófagos son esenciales en la contención de la replicación de *T. cruzi* al inicio de la infección, siendo activadas por distintas moléculas del parásito (Abrahamsohn, 1998; Savino et al., 2007). En los macrófagos y células dendríticas se activan mecanismos que reconocen al parásito a través de los receptores tipo Toll (TLRs), activando las vías de señalización dependiente de Myd88 o TRIF, mediante las que se genera una respuesta basada, principalmente, en la liberación de citoquinas proinflamatorias (Tarleton, 2007). Así, en la fase aguda de la enfermedad de Chagas, en respuesta al parásito, se liberan las citoquinas TNF- $\alpha$ , IL-2 e IL-12, que a su vez activan a las células NK, responsables de actividad citotóxica y de la secreción de IFN- $\gamma$  (Cardoni et al., 1999; la Fuente Zeraín, 2006; Trinchieri, 1995). Tanto el IFN- $\gamma$  como el TNF- $\alpha$  restringen la propagación del parásito en macrófagos, esencialmente por mecanismos que implican la participación del óxido nítrico (Silva et al., 2003). Así, se ha descrito que estas dos citoquinas participan en la activación de la enzima óxido nítrico sintasa inducible (iNOS) responsable de la generación de óxido nítrico (NO). El NO posee una gran capacidad de destrucción del parásito asociándose a protección frente a la infección, aunque también puede contribuir a la patogénesis de la enfermedad de Chagas si sus especies reactivas se producen en exceso o por largos periodos y no hay una respuesta antioxidante adecuada (Silva et al., 2003; Zacks et al., 2005).

La presencia de estas citoquinas, así como la activación de las células presentadoras de antígenos, ponen en acción a las células T y B, direccionando la respuesta frente a la infección por *T. cruzi* hacia una respuesta inmune adaptativa.

## **3.2. Respuesta inmunológica adaptativa**

### **3.2.1. Respuesta humoral**

La respuesta inmune humoral producida frente a los antígenos de *T. cruzi* permite controlar la propagación del parásito en el hospedador (Sepulveda et al., 2000). Así, los anticuerpos bloqueantes secretados por las células B participan en el control de la parasitemia por el aclaramiento de los parásitos circulantes, desempeñando, por tanto, esta repuesta un papel relevante en la respuesta protectora frente a *T. cruzi* (Acosta-Rodríguez et al., 2004; Krettli and Brener, 1976).

Es por ello que diferentes estudios han descrito que la producción de anticuerpos, aunque es secundaria en importancia a las respuestas celulares, es absolutamente necesaria. Así, se ha observado que ratones deficientes en linfocitos B, acaban sucumbiendo a la infección (Kumar and Tarleton, 1998). Por otra parte, se han implicado diferentes isotipos de IgG en la activación policlonal de las células B y en la respuesta de anticuerpos específicos contra el parásito (Bouhdidi et al., 1994; D'Imperio-Lima et al., 1986; Minoprio et al., 1987; Rowland et al., 1992).

Del mismo modo, se ha sugerido un papel protector para los anticuerpos "líticos" anti-*T. cruzi*, encontrándose niveles superiores de estos en el suero de pacientes indeterminados en comparación con pacientes cardiacos (Kierszenbaum and Lima, 1983) siendo capaces de fomentar la expansión de células T y B de estos pacientes.

Por todo ello, el conocimiento de la respuesta humoral resulta esencial, dado que la identificación en el suero de pacientes de Chagas de anticuerpos específicos frente a antígenos del parásito constituye una relevante herramienta diagnóstica de la infección, así como del estado patológico o la eficacia del tratamiento.

### **3.2.2. Respuesta celular**

Dada la replicación intracelular de *T. cruzi* en la célula infectada del hospedador, durante la evolución de la enfermedad de Chagas, es fundamental la respuesta celular de la inmunidad adaptativa. Esta la desempeñan, principalmente, los linfocitos T, que reconocen los antígenos del parásito e impulsan funciones para combatir la infección (de Moraes et al., 2015). Así, se ha descrito que los pacientes con Chagas crónico (indeterminados y cardiacos) presentan en su torrente sanguíneo altas frecuencias de células T activadas (Dutra et al., 1994).

La acción antiparasitaria más directa de estas poblaciones de células T está impulsada, principalmente, por las células T citotóxicas CD8<sup>+</sup> antígeno-específicas, que pueden reconocer y destruir las células infectadas del huésped a través de la secreción de moléculas citolíticas o por la vía Fas/FasL (Padilla et al., 2009; Ruiz and Becker, 2007). Del mismo modo, los linfocitos T CD8<sup>+</sup> con capacidad citotóxica, mediada por perforinas y granzimas, son fundamentales para controlar la infección



intracelular, aunque requieren la ayuda aportada por las células T CD4<sup>+</sup> para alcanzar el fenotipo de memoria y ser autónomos en una segunda expansión al reencontrarse con el antígeno (Novy et al., 2007). Cuando las células T CD8<sup>+</sup> se activan por el estímulo con el parásito, se secretan citoquinas como IFN- $\gamma$  y TNF- $\alpha$ . Esta capacidad de secreción de citoquinas inflamatorias, así como de actividad citotóxica, unido a la abundancia de estos linfocitos en infiltrados inflamatorios y tejidos infectados, ponen de manifiesto el crítico papel de estas células T en la resistencia a *T. cruzi* (Martin and Tarleton, 2005).

Respecto a la población de células T CD4<sup>+</sup>, las citoquinas producidas por ellas son esenciales en la regulación de la respuesta inmunitaria generada frente a *T. cruzi* (Montiel and Díaz, 2002). Según las moléculas liberadas, se dará una respuesta Th1, de la que son características el IFN- $\gamma$ , IL-2, IL-12 y TNF- $\alpha$ , que dan lugar a una respuesta T CD8<sup>+</sup> efectora, o Th2, caracterizada por IL-4, IL-5, IL-9 e IL-13, que promueven la activación de células B y la generación de anticuerpos, como la secreción de IgE a través de la producción de IL-4 (de Moraes et al., 2015). También se han descrito los linfocitos Th17, que producen IL-17. La expresión de esta citoquina proinflamatoria se ha correlacionado con una mejora en la función cardiaca, dado que en estudios en pacientes de Chagas indeterminados y cardiacos se pudo observar que aquellos con sintomatología cardiaca presentaban menores niveles de expresión de IL-17 y menor abundancia de células Th17 circulantes (Magalhães et al., 2013; Sousa et al., 2017).

De este modo, en fase aguda, la respuesta Th1 ejercería una función protectora (Brener and Gazzinelli, 1997; Briceño and Mosca, 1996). Por el contrario, el predominio de una respuesta Th2, con el correspondiente aumento de la producción de IL-4, aumentaría la susceptibilidad del huésped a la infección por *T. cruzi* (Montiel and Díaz, 2002). No obstante, el balance entre las respuestas Th1/Th2 es necesario, puesto que Th1 implica mecanismos que controlan la replicación del parásito y Th2 regula la respuesta inmune desencadenada, que en exceso genera patología (Abrahamsohn, 1998; Mosca and Briceño, 2000).

La citoquina supresora IL-10 es fundamental en la mediación de la enfermedad aguda dado su papel en la tolerancia inmunológica, el desarrollo de células T y células dendríticas, así como en la proliferación y diferenciación de

células B (Moore et al., 2001). IL-10 es secretada por numerosas células del sistema inmune, como los linfocitos T CD4<sup>+</sup> y CD8<sup>+</sup> reguladores, que reducen el daño ocasionado por la respuesta inmune generada (Shevach, 2001). Esta importante función inmunomoduladora se observó en ensayos con ratones deficientes de IL-10, los cuales presentaron una disminución de la parasitemia, pero un aumento de la mortalidad (Abrahamsohn and Coffman, 1996; Hunter et al., 1997).

### ***3.2.2.1. Respuesta celular en fase crónica de la enfermedad de Chagas***

Los mecanismos del sistema inmune activados durante la fase aguda de la enfermedad de Chagas controlan la replicación de *T. cruzi*, resolviendo la sintomatología clínica de esta etapa, de tal forma que se establece un equilibrio que da lugar a la cronicidad de la enfermedad. Es por ello que, en la fase crónica, los signos y síntomas clínicos de los pacientes se asocian a la eficacia de la respuesta inmune generada.

Mientras que en la forma crónica cardíaca predomina un ambiente inflamatorio, con la producción de citoquinas como TNF- $\alpha$ , IFN- $\gamma$  y otras moléculas citotóxicas en las que participan células T CD8<sup>+</sup>, en la forma crónica indeterminada predomina una respuesta inmune reguladora, caracterizada por la producción de IL-10 e IL-17 (Pérez-Molina and Molina, 2018). No obstante, también se han detectado niveles elevados de TNF- $\alpha$  e IFN- $\gamma$  en pacientes IND en comparación con sujetos sanos (Ferreira et al., 2003; Requena-Méndez et al., 2013). Sin embargo, existe controversia y otros autores han descrito una correlación opuesta entre la expresión de IFN- $\gamma$  y la enfermedad cardíaca, detectándose en estos pacientes menor número de linfocitos T CD8<sup>+</sup> productores de IFN- $\gamma$  en comparación a los pacientes IND o sujetos con un cuadro clínico leve (Laucella et al., 2004). Además, en pacientes asintomáticos, se ha encontrado la coproducción de citocinas TNF- $\alpha$ , IFN- $\gamma$  e IL-2 por parte de células T CD8<sup>+</sup> en alta proporción, que disminuye a medida que la enfermedad progresa hacia formas cardíacas (Mateus et al., 2015). Se ha detectado en pacientes crónicos con enfermedad severa, un incremento de linfocitos T CD8<sup>+</sup> altamente diferenciados, agotamiento clonal e incremento en la tasa de apoptosis de los linfocitos T CD4<sup>+</sup> periféricos (Albareda et al., 2009; Lasso et al., 2015). En este proceso de agotamiento, la actividad funcional de los linfocitos

antígeno-específicos del huésped está regulada por los receptores inhibitorios (Pérez-Antón et al., 2019).

Por otro lado, se ha demostrado que las células T activadas circulantes en sujetos asintomáticos y sintomáticos expresan tanto citoquinas inflamatorias como antiinflamatorias, lo que es consistente con una inmunoregulación activa en la fase crónica (Cunha-Neto et al., 2005; Dutra et al., 1997).

Cabe destacar que, durante la inmunidad adaptativa, se activan las células de memoria antígeno-específica, lo cual se traduce en una disminución de los signos clínicos severos tras volver a infectarse con *T. cruzi*. Es por ello que la reinfección suele presentarse con una menor y breve parasitemia, así como con una leve sintomatología.

A pesar de los avances realizados en el estudio de la respuesta inmunológica en los pacientes de Chagas, queda mucho por conocer de los mecanismos que resultan eficaces para el control de la enfermedad, así como de los marcadores inmunológicos que se asocian a cada fase clínica en los pacientes de Chagas.



## **OBJETIVOS**

---



1. Identificar los retroelementos portadores de la huella Pr77 y el material genético de interés contenido en el genoma de 6 cepas de *T. cruzi* pertenecientes a diferentes DTUs y analizar la variabilidad existente entre las 6 cepas a estudio.
2. Caracterizar los sitios de inserción de huellas Pr77 no asociadas a retroelementos (Pr77Sh) en el genoma de las distintas cepas de *T. cruzi* a estudio y la actividad ribozima asociada a la estructura secundaria tipo HDV de estas huellas.
3. Determinar el contenido de retroelementos y de genes de interés en el genoma de la cepa SOL de *T. cruzi* secuenciada por tecnología PacBio y su expresión diferencial entre las formas epimastigotas, amastigotas y tripomastigotas del parásito.
4. Analizar la sensibilidad y especificidad de los antígenos KMP11, HSP70, PFR2 y 3973<sup>d</sup> de *T. cruzi* para el diagnóstico serológico de pacientes asintomáticos con enfermedad de Chagas en fase crónica.
5. Evaluar a los antígenos KMP11, HSP70, PFR2 y 3973<sup>d</sup> como biomarcadores serológicos del impacto terapéutico en pacientes con enfermedad de Chagas en fase crónica.
6. Estudiar cómo responde el sistema inmunitario humano a la infección por *T. cruzi* mediante el mapeo de los cambios en los patrones de expresión génica (tecnología OpenArray®) en pacientes crónicos con enfermedad de Chagas en las distintas fases clínicas de la enfermedad y tras el tratamiento con benznidazol.





# **MATERIALES Y MÉTODOS**

---



## MATERIALES

### 1. Material y productos químicos

En la presente Tesis Doctoral se ha utilizado material y reactivos de grado molecular y con alta pureza. Estos pertenecen, principalmente, a las casas comerciales Sigma-Aldrich, Gibco, Thermo Fisher, Bio-Rad y Promega. Las placas (OpenArray® *plate*) personalizadas para el sistema *QuantStudio™ 12K Flex Real-Time PCR* que incluían sondas TaqMan® se obtuvieron de Thermo Fisher Scientific. Los productos usados para el marcaje radiactivo fueron adquiridos de Perkin Elmer.

### 2. Tampones y soluciones

- Laemmli: Tris-HCl 50 mM (pH 6,8), DTT 100 mM, SDS 2%, azul de bromofenol 0,1%, glicerol 10% (Laemmli, 1975).
- Tampón de electroforesis de proteínas (SDS-PAGE) 10X: Tris-base 30,3 g/l, glicina 144,2 g/l, SDS 10 g/l (pH 8,3).
- Tampón de lisis de eritrocitos: NH<sub>4</sub>Cl 0,15 M, KHCO<sub>3</sub> 1 mM, EDTA 0,1 mM (pH 7,4).
- Tampón de lisis de parásitos: Tris-HCl 50 mM (pH 7,4), Nonidet P-40 0,05%, NaCl 50 mM, PMSF 1 mM, leupeptina 1 µg/ml.
- Tampón fosfato salino (PBS) 1X: NaCl 0,13 M, Na<sub>2</sub>HPO<sub>4</sub> 70 mM, NaH<sub>2</sub>PO<sub>4</sub> 3 mM (pH 7,2).
- Tampón carbonato/bicarbonato: Na<sub>2</sub>CO<sub>3</sub> 0,02 M, NaHCO<sub>3</sub> 0,02 M (pH 9,6).
- Solución ortofenildiamina: 1 mg/ml ODP (Sigma-Aldrich), 1 µL/ml H<sub>2</sub>O<sub>2</sub> 30% en tampón citrato 0,1 M (pH 5).
- Tampón TAE 50X: Tris-HCl 2 M (pH 7,2), EDTA 50 mM (pH 8), ácido acético hasta pH 7,2.
- Tampón TBE 10X: Tris-HCl 890 mM (pH 8,3), ácido bórico 890 mM y EDTA 20 mM.
- TEN<sub>250</sub>: 10 mM Tris-HCl (pH 7,5), 1 mM EDTA, 250 mM NaCl.
- Tampón de carga geles de agarosa 6X: Azul de bromofenol 0,25% (p/v), xilencianol 0,25% (p/v) y glicerol 30% (v/v).
- Tampón de carga desnaturizante 2X: Formamida 92% (v/v), azul de bromofenol 0,025% (p/v), xilencianol 0,025% (p/v) y EDTA 17 mM.

- Tampón de corte 5X: 200 mM Tris-HCl (pH 7), 50 mM NaCl y 0,5 mM MgCl<sub>2</sub> (tampón 0,1 mM), 5 mM MgCl<sub>2</sub> (tampón 1 mM) ó 50 mM MgCl<sub>2</sub> (tampón 10 mM).
- Fenol: el fenol destilado se saturó y equilibró con Tris-HCl 0,1 M (pH 8). En el momento de utilizar se añadió 8-hidroxiquinoleína hasta 0,1% y se almacenó a 4°C.

### 3. Anticuerpos

- *Goat Anti-Human IgG (H+L)* conjugado a peroxidasa (Life Technologies).

### 4. Antibióticos

- Ampicilina 100 µg/ml (Amresco®).
- Gentamicina 50 µg/ml (Thermo Fisher Scientific).
- Tetraciclina 10 µg/ml (Sigma-Aldrich).

### 5. Medios de cultivo

- Medio LB: Bactotripton 1% (p/v), NaCl 0,5% (p/v), extracto de levadura 0,5% (p/v) (pH 7-7,2). LB-agar: medio LB líquido, agar al 15% (p/v).
- Medio RPMI 1640 con L-Glutamina (Gibco), suplementado con 10% de suero fetal bovino inactivado por calor (56°C x 30 min) (SFBi) (Gibco) y gentamicina.
- DMEM (*high glucose*, Gibco): (+) L-Glutamina, (+) 4,5 g/l glucosa y (-) piruvato, suplementado con 10% de SFBi (Gibco) y gentamicina.
- Medio LIT: NaCl 4 g, KCl 0,4 g, Na<sub>2</sub>HPO<sub>4</sub> 8 g, glucosa 2 g, infusión de hígado 5 g, triptosa 5 g, hemina 25 mg y agua destilada hasta 1 litro (pH 7,2).

### 6. Patrones de peso molecular

- Electroforesis de proteínas

- *Prestained SDS-PAGE Standards* (Bio-Rad). Tamaño aproximado de las bandas: 206.911, 118.961, 98.540, 56.695, 37.544, 29.494, 20.265 y 6.880 Da.
- *Precision Plus Protein All Blue Standards* (Bio-Rad). Tamaño aproximado de las bandas: 250.000, 150.000, 100.000, 75.000, 50.000, 37.000, 25.000, 20.000, 15.000 y 10.000 Da.

## - Electroforesis de ADN

- *100 bp DNA Ladder* (Bioron). Tamaño molecular de los fragmentos: 1.000, 900, 800, 700, 600, 500, 400, 300, 200, 100 pares de bases (pb).
- *Lambda DNA/HindIII DNA Molecular Weight Marker II* (Roche). Tamaño molecular de los fragmentos: 23.130, 9.416, 6.557, 4.361, 2.322, 2.027, 564, 125 pb.

**7. Oligonucleótidos**- **Universal**

- **M13Fw(-20) universal primer** (5'-GTAAAACGACGGCCAGT-3')

- **Diseñados**

- **Pr77f** (5'CCCTGGCTCAGCCGGCCACCTCAACGTGGTGCCAGGGTCTAGTACTC-3')
- **Pr77r** (5'-CAGCAGGCGCTTAGCTTCCTCTCTAGCAAAGAGTACTAGACCCTGGC-3')
- **51IKf** (5'-CCCTGGCAATGCCGGCCGCTCAGCGTGGTGCCAGGGTCGATTGCC-3')
- **51IKr** (5'-TGCTGGCTCCTGGCTTCCCCGACAATACGGGGCAATCGACCCTGGC-3')
- **38IKr** (5'-CTTCCGGCCTTTTGATTCTTTTGATTCTTTTCAAGCAAGAGTACTAGACACTGGCA-3')
- **8IKf** (5'-CCCTGGCGCAGCCGGCCACTTCAACGTTGTGCCAGTGTCTAGTACTC-3')
- **8IKr** (5'-CTTCCGGCCCTTTTGATTCTCCCCAAGCAAGAGTACTAGACACTGGCA-3')
- **1BMr** (5'-CAGCAGGCGCTTGGCTTCCTCTCTAGCAAAGAGTACTAGACCCTGGC-3')
- **2BMf** (5'-CCCTGGCGCAGCCAGACACCTCAACGTGGTGCCAGGGTCTAGTACTC-3')
- **2BMr** (5'-CAGCAGGCACTTGACTTCCCTCACTAGCAAAGAGTACTAGACCCTGGC-3')
- **15Yf** (5'-CCCTGGCAACGCCGGCCGCTCAGCGTGGTGCCAGGGTCAATTGC-3')
- **15Yr** (5'-TGCTGACACCTGGCTTCCCCGACAATACGGAGCAATTGACCCTGGCAC-3')
- **38SOLr** (5'-CTTTTGATTCTTTTGATTCTTTTCAAGCAAGAGTACTAGACACTGGCA-3')
- **102CLf** (5'-GCCTGGCAATGCCGGCCGCGCAGCGTGGTGCCAGGGTCGATTGCCCC-3')
- **102CLr** (5'-TGCTGGCTCCTGGCTTCCCCGACAATACGGGGCAATCGACCCTG-3')

**8. Cepas de *Trypanosoma cruzi***

- Cepa SOL (MHOM/ES/2008/SOL; DTU V), formas epimastigote, tripomastigote y amastigote.
- Cepa B. M. López (MHOM/CO/87; DTU Ia), forma epimastigote.
- Cepa Dm28 (MDID/CO/00/Dm28, DTU Id), forma epimastigote.
- Cepa Y (MHOM/BR/1950/Y; DTU II), forma epimastigote.
- Cepa Ikiakarora (IRHO/CO/95; DTU III), forma epimastigote.
- Cepa CL Brener (FIOCRUZ/MG/1963; DTU VI), forma epimastigote.

### 9. Líneas celulares

- LLC-MK2: línea epitelial procedente de riñón de mono Rhesus.

### 10. Cepas bacterianas

- *Escherichia coli* XL1-Blue (Genotipo: recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, F' proAB lacI<sup>a</sup> M15, Tn10 (Tet<sup>r</sup>)).

### 11. Proteínas recombinantes y péptidos

- KMP11 (*Kinetoplastid Membrane Protein 11 kDa*).
- HSP70 (*Heat Shock Protein 70 kDa*).
- PFR2 (*Paraflagellar Rod protein 68 kDa*).
- Péptido 3973<sup>d</sup> (FGQAAAGDKPSLGGFGQAAAGDKPSL).

### 12. Población de estudio y consideraciones éticas

La participación de todos los pacientes y donantes sanos fue voluntaria y todos firmaron el consentimiento informado para participar en el estudio.

- **Análisis multicéntrico para evaluación de diferentes biomarcadores de *T. cruzi* en pacientes con enfermedad de Chagas**

Este estudio multicéntrico cumplió con los principios de la Declaración de Helsinki. Los protocolos fueron aprobados por los comités éticos de la Fundación CEADES (ref.: NHEPACHA-16,032,013; Cochabamba, Bolivia); Hospital General de Agudos “Dr. I. Pirovano” (ref.: DI-2015-522-HGAIP) y Hospital Interzonal General de Agudos “Eva Peron” (refs.: 24/14 y 11/17) de Buenos Aires (Argentina); Consejo Superior de Investigaciones Científicas (ref.: 094/2016; Madrid, España); Instituto de Medicina Tropical (ref.: CEC-IMT 12/2016; Caracas, Venezuela); Nucleo de Estudos da doença de Chagas (ref.: 2.245.763; Goiânia, Brasil) y Hospital Clínic de Barcelona (ref.: HCB/2013/8031; Barcelona, España). La población a estudio incluye sujetos con enfermedad crónica de Chagas diagnosticados en el país de origen con, al menos, dos pruebas serológicas distintas, de acuerdo a las recomendaciones de la OMS (WHO, 2002) y sujetos sanos procedentes de zonas endémicas de la enfermedad de Chagas. El estado serológico de los participantes en el estudio se confirmó mediante la prueba ELISA disponible en el mercado (Biokit). Los sujetos infectados procedían de seis países diferentes de Sudamérica y los del grupo control de Bolivia (**Tabla M1**).

**Tabla M1. Cohorte de pacientes empleada para el análisis de la sensibilidad y especificidad de los antígenos KMP11, HSP70, PFR2 y 3973<sup>d</sup> de *T. cruzi* en pacientes con enfermedad de Chagas.**

Grupo de individuos	Origen (%)	Edad (años)	Sexo
		Media ( $\pm$ SD)	[% Mujeres (M)/ Hombres (H)]
<b>Donantes sanos</b> (n=20)	100% Bolivia	32,6 (12,0)	62% M 38% H
<b>Pacientes infectados con <i>T. cruzi</i></b> (n=221)	56% Bolivia 15% Argentina 16% Brasil 3% Colombia 2% Paraguay 8% Venezuela 0,5% ND	40,1 (12,3)	59% M 41% H

ND: no determinado.

- **Evaluación del set de antígenos KMP11, HSP70, PFR2 y 3973<sup>d</sup> de *T. cruzi* como biomarcadores de eficacia terapéutica en pacientes con enfermedad de Chagas. Establecimiento del algoritmo para evaluación de la eficacia terapéutica**

Las muestras de suero de los 66 pacientes estudiados se recogieron en los hospitales Virgen de la Arrixaca (Murcia) y Virgen de las Nieves (Granada). La enfermedad de Chagas se diagnosticó mediante ELISA (Bioelisa Chagas Biokit, España) e inmunofluorescencia indirecta (Inmunofluor Chagas, Biocientífica, Argentina), siguiendo los criterios de la OMS (WHO, 2002). Los pacientes no habían recibido ningún tratamiento para la enfermedad de Chagas antes de su inclusión en este estudio. Todos eran adultos en fase crónica indeterminada (IND) y tenían entre 22 y 53 años de edad. Sesenta y cinco de los pacientes eran de Bolivia y uno de El Salvador. Todos los pacientes residían en España, donde el Chagas no es endémico, y el 86,4% de ellos había residido en España durante, al menos, 2 años. Los pacientes con enfermedad de Chagas se trataron con benznidazol (5 mg/kg de peso corporal al día durante 60 días) y se les realizó seguimiento clínico durante 48 meses. Se recogieron y analizaron muestras de todos los pacientes (n=66) antes del tratamiento (día 0, T0) y a los 9 y 24 meses después del tratamiento (T9 y T24) y de 32 pacientes a los 48 meses después del tratamiento (T48), ya que los 34 restantes no acudieron a la consulta médica a los 48 meses. Los protocolos fueron aprobados por los Comités Éticos del Consejo Superior de Investigaciones Científicas (ref.: 094/2016; Madrid, España) y el Hospital Virgen de la Arrixaca (ref.: MTR-02/2014; Murcia, España).

- **Expresión diferencial de genes de respuesta inmune en pacientes con enfermedad de Chagas**

Los pacientes adultos con enfermedad de Chagas crónica originarios de zonas endémicas y residentes en España incluidos en este estudio se reclutaron, diagnosticaron y evaluaron clínicamente en el Hospital Virgen de la Arrixaca de Murcia (España). Además, se incluyeron donantes sanos de zonas endémicas y no endémicas de la enfermedad de Chagas, reclutados en el IPBLN-CSIC (Granada). Los protocolos fueron aprobados por los Comités Éticos del Consejo Superior de Investigaciones Científicas (ref.: 094/2016; Madrid, España) y el Hospital Virgen de la Arrixaca (ref.: MTR-05/2016; Murcia, España). Los pacientes se diagnosticaron según los criterios de la OMS en base a las pruebas serológicas (ELISA Chagas, Ortho Clinical Diagnosis, e Inmunofluor Chagas, Biocientífica, Argentina) y caracterizaron como indeterminados (IND) debido a la ausencia de manifestaciones cardíacas o digestivas (G0 según la clasificación de Kuschnir), o cardíacos (CCC) por la presencia de las mismas (GI según la clasificación de Kuschnir). Del mismo modo, se incluyeron pacientes IND y CCC que habían recibido tratamiento antiparasitario (benznidazol; 5 mg/kg de peso corporal al día durante 60 días). Los datos referentes a la edad, sexo y el país de origen de cada uno de los sujetos incluidos en este estudio se detallan en la **Tabla M2**.

En total, se incluyeron 39 muestras de 71 pacientes IND pre-tratamiento, 46 de 61 pacientes IND post-tratamiento, 18 de 38 pacientes CCC pre-tratamiento, 21 de 24 pacientes CCC post-tratamiento y 30 muestras de 34 donantes sanos (14 de zonas endémicas y 20 de zonas no endémicas). Debido a la cantidad de ARN necesaria para la síntesis de ADNc para la RT-qPCR de alto rendimiento y al número limitado de células aisladas de la muestra de sangre de determinados pacientes, en ciertos casos, fue necesario mezclar células de algunos pacientes.



**Tabla M2. Cohorte de pacientes empleada para el estudio de la expresión diferencial de genes de respuesta inmune en pacientes con enfermedad de Chagas mediante OpenArrays®. IND: Enfermedad de Chagas indeterminada; CCC: Enfermedad de Chagas cardiaca. ND: No determinado.**

Grupo de pacientes	Origen (%)	Edad (años)		Sexo [% Mujeres (M)/ Hombres (H)]	
		Media ( $\pm$ SD)	Rango		
Donantes sanos	<b>Procedente de zonas no endémicas</b> (n=20)	100% España	37,6 (12,8)	22-56	60% M 40% H
	<b>Procedentes de zonas endémicas</b> (n=14)	28,6% Colombia 14,3% Venezuela 7,1% Chile 7,1% Panamá 7,1% Ecuador 35,7% ND	37,5 (8,4)	21-54	64,3% M 35,7% H
	<b>Pacientes IND pre-tratamiento</b> (n=71)	93% Bolivia 1,2% Salvador 1,2% Paraguay 4,7% ND	36,7 (9,6)	18-59	70,9% M 29,1% H
	<b>Pacientes IND post-tratamiento</b> (n=61)	96,7% Bolivia 1,6% Salvador 1,6% Paraguay	37,1 (8,6)	21-57	70,5% M 29,5% H
	<b>Pacientes CCC pre-tratamiento</b> (n=38)	73,7% Bolivia 2,6% Ecuador 2,6% Paraguay 21,1% ND	36,4 (8,8)	21-55	57,9% M 42,1% H
	<b>Pacientes CCC post-tratamiento</b> (n=24)	83,3% Bolivia 4,2% Paraguay 12,5% ND	32,8 (9)	19-55	62,5% M 37,5% H



## MÉTODOS

### 1. ANÁLISIS COMPARATIVO DEL CONTENIDO GÉNICO DE LOS GENOMAS DE DIFERENTES CEPAS DE *T. cruzi*

#### 1.1. Cultivo de formas epimastigotes y purificación de ADN genómico

Las formas epimastigotes de las cepas B. M. López (DTU Ia), Dm28 (DTU Id), Y (DTU II), Ikiakarora (DTU III), SOL (DTU V) y CL Brener (DTU VI) de *T. cruzi* se cultivaron en medio LIT suplementado con un 10% (v/v) de SFBi (Flow Laboratory, Irvine, Reino Unido), a 28°C y con agitación suave. Los cultivos en fase logarítmica de crecimiento ( $10\text{-}20 \times 10^6$  parásitos/ml), se recolectaron por centrifugación a 2.500 rpm durante 15 min, lavaron 2 veces con PBS 1X frío y se lisó la membrana externa añadiendo PBS 1X conteniendo NP40 al 1%. Los núcleos se recogieron por centrifugación a 13.000 rpm durante 5 min, en frío, y se suspendieron en PBS 1X frío conteniendo SDS al 0,5%. El ADN genómico se purificó tras extracción con fenol/cloroformo (una extracción fenólica, dos extracciones con fenol:cloroformo:alcohol isoamílico (25:24:1) y una con cloroformo:alcohol isoamílico (24:1)). El ADN genómico se precipitó en 1/10 volumen de acetato de sodio (AcONa) 3 M (pH 6,25) y 2,5-3 volúmenes de etanol absoluto frío (EtOH). El ADN se recogió centrifugando a 13.000 rpm durante 30 min en frío (4°C) y se lavó con EtOH al 70% frío.

#### 1.2. Genotipado de las cepas

La identificación del linaje de los parásitos la llevaron a cabo otros miembros del laboratorio por PCR para la amplificación del mini-exón, el fragmento A10-e, y los genes ribosomales 18S y 24S. La amplificación por PCR del mini-exón y ARNr 24S $\alpha$  se realizó como describieron Brisse et al., 2001 (Brisse et al., 2001). La reacción de PCR para la amplificación del dominio de tamaño variable de la secuencia de ARNr 18S se ejecutó como describieron Clark et al., 1994 (Clark and Pung, 1994). La amplificación del fragmento A10-e se realizó mediante PCR como describieron Brisse et al., 2000 (Brisse et al., 2000). Por otro lado, la caracterización del haplotipo de TcI se basó en los polimorfismos de nucleótidos de la región intergénica del mini-exón como lo describieron Falla et al., 2009 (Falla et al., 2009).

### 1.3. Preparación de librerías y secuenciación de los genomas

#### - Secuenciación Ion Torrent

Los genomas completos de las cepas de *T. cruzi* B. M. López, Dm28, Y, Ikiakarora, SOL y CL Brener se secuenciaron mediante tecnología Ion Torrent (Thermo Fischer Scientific) por otros miembros del laboratorio, tal como se ha reportado en (Gómez, López, et al., 2021; Gómez, Rastrojo, Lorenzo-Díaz, et al., 2020; Gómez, Rastrojo, Sánchez-Luque, et al., 2020). Brevemente, se emplearon 500 ng de ADN genómico en tampón Tris 10 mM (pH 7,6) a una concentración de 10 ng/μl para la construcción automática de librerías en el sistema AB Library Builder utilizando el kit '*Ion Xpress Plus Library*' (Thermo Fischer Scientific) siguiendo las instrucciones del fabricante. Las librerías se seleccionaron por tamaño hasta un rango de longitud óptimo de 450 pb mediante el sistema *E-Gel* y geles de agarosa *SizeSelect 2%* (Invitrogen). Para el control de calidad de las muestras de ADN, la determinación de la concentración y la estimación de la distribución de la longitud de los fragmentos, se empleó el sistema *Agilent 2100 Bioanalyzer* utilizando el kit '*High-sensitivity DNA kit*' (Agilent Technologies). También se determinó la calidad del ADN empleando el kit '*Quant-IT dsDNA assay*' y por fluorometría usando *Qubit* (Invitrogen). A continuación, las muestras de ADN se diluyeron en serie hasta una concentración final de 23 pM y se sometieron a la PCR en emulsión y al enriquecimiento mediante el sistema *One Touch 2 (Ion OneTouch 400 Template kit;* Life Technologies). Las muestras enriquecidas se cargaron en los chips *Ion 316(v2)* y se secuenciaron en el sistema *Ion Torrent PGM™ con Hi-Q™ Sequencing Chemistry* (Thermo Fischer Scientific) en el Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias (IUETSPC).

#### - Secuenciación Pacific Bioscience (PacBio).

El ADN genómico de la cepa SOL de *T. cruzi* se secuenció y ensambló en *contigs* por tecnología de secuenciación *Pacific Biosciences* (PacBio) en el centro Noruego de Secuenciación (NSC, *Norwegian Sequencing Centre*) ([www.sequencing.uio.no](http://www.sequencing.uio.no)) por otros miembros del laboratorio. La librería se preparó usando el protocolo de preparación de librería de 20 kb de PacBio. La selección del tamaño de la librería final se realizó utilizando el sistema *BluePippin* (tamaño mínimo 7 kb). La librería obtenida se secuenció con el equipo *Pacific Biosciences RS II* utilizando química P6-C4, con 240 min de tiempo de ejecución. Para

la secuenciación, se utilizaron un total de siete células SMRT (*Single Molecule Real Time sequencing*), lo que arrojó 692.620 lecturas crudas.

#### 1.4. Control de calidad de las lecturas crudas

Tras la secuenciación, las lecturas crudas se analizaron con el programa FastQC v0.10.1 (Andrews, 2010), con parámetros por defecto, obteniéndose parámetros de calidad para cada nucleótido secuenciado. Seguidamente se usó el programa Prinseq v0.20.4 (Schmieder and Edwards, 2011) para el filtrado por calidad aplicando múltiples ejecuciones con los siguientes parámetros:

- -derep 14, -ns\_max\_p 1 -ns\_max\_n 3 -trim\_ns\_left 1 -trim\_ns\_right 1
- -trim\_qual\_right 20 -trim\_qual\_type mean -trim\_qual\_window 5 -trim\_qual\_step 1
- -trim\_qual\_right 20 -trim\_qual\_type mean -trim\_qual\_window 1 -trim\_qual\_step 1
- -trim\_qual\_left 20 -trim\_qual\_type mean -trim\_qual\_window 5 -trim\_qual\_step 1
- -trim\_qual\_left 20 -trim\_qual\_type mean -trim\_qual\_window 1 -trim\_qual\_step 1
- -lc\_method entropy -lc\_threshold 50, -min\_qual\_mean 25, -min\_len 50.

#### 1.5. Ensamblaje de los genomas y visualización de los *contigs* resultantes

Los genomas de las seis cepas de *T. cruzi* secuenciadas por tecnología Ion Torrent se ensamblaron empleando CLC Genomic Workbench v8.0 (Qiagen). Se aplicaron los parámetros definidos por defecto, modificando la longitud y similitud de la fracción, que se estableció en 0,90 y 0,97, respectivamente. Se exigió, además, una longitud mínima de *contig* de 500 pb.

El genoma de *T. cruzi* de la cepa SOL secuenciado por PacBio y ensamblado, se reensambló de nuevo por nosotros mediante el alineamiento de las lecturas resultantes de la secuenciación, empleando el programa Canu v2.0 (Koren et al., 2017) seleccionando parámetros de sensibilidad extra (corMhapSensitivity=high corMinCoverage=0 genomeSize=32.5m). Posteriormente, para mejorar el ensamblaje, se realizaron tres rondas de refinamiento con lecturas cortas de *Illumina* de esta misma cepa (procedentes de la secuenciación Ion Torrent y ARN-seq) utilizando el programa Pilon v1.23 (Walker et al., 2014) ya que se encontró que este método aumentaba, de manera importante, la calidad del ensamblaje (Chen et al., 2021).

Para visualizar los alineamientos de las lecturas sobre los ensamblajes genómicos se usó el programa IGV (*Integrative Genomics Viewer*) (Thorvaldsdottir et al., 2013).

### 1.6. Evaluación de la integridad de los ensamblajes

La integridad del ensamblaje de los genomas se evaluó mediante análisis BUSCO (*Benchmarking Universal Single-Copy Orthologs*) v.4.0.5 (Seppey et al., 2019). Esta herramienta compara los genomas ensamblados con un conjunto de ortólogos predefinido, en este caso el de Euglenozoa (v10), el cual incluye 130 BUSCOs. Los parámetros ejecutados fueron: -m genome, -l euglenozoa\_odb10. Se repitió el análisis permitiendo que el programa seleccionase automáticamente el conjunto de ortólogos del linaje más cercano (parámetros: -m genome --autolineage). Los resultados de ambas ejecuciones coincidieron al detectarse como linaje más cercano a Euglenozoa.

### 1.7. Anotación de los genomas

La anotación de cada genoma se realizó de manera individual usando la herramienta online de predicción AUGUSTUS disponible en: (<http://bioinf.uni-greifswald.de/webaugustus/>) tal y como describe (Hoff and Stanke, 2013). Dado que *T. cruzi* no se encontraba entre los organismos elegibles para este proceso y con el fin de que la predicción fuese lo más precisa posible, se llevó a cabo un paso previo en el que a partir de un genoma de referencia se generó un archivo de parámetros a aportar en la predicción. En el caso de los genomas de *T. cruzi* secuenciados por tecnología Ion Torrent, la selección de los genomas de referencia se basó en un análisis de *Best Hit* en el que se enfrentaron los genomas ensamblados a una base de datos constituida por las proteínas identificadas de varios genomas de *T. cruzi* descargadas de la base de datos TriTrypDB (<http://tritrypdb.org/>), usando un script en Python personalizado. De esta forma, los genomas de referencia para cada cepa se seleccionaron en función a la mayor identidad obtenida en cada caso. Así, el genoma Dm28c (v.46) se seleccionó como genoma de referencia para la anotación de los genomas de B. M. López y Dm28, CL Brener Esmeraldo-like (v.46) para la de los genomas de Y e Ikiakarora y CL Brener non-Esmeraldo-like (v.46) para los

genomas de la cepa SOL y CL Brener. Una vez realizada la predicción, se usó la herramienta OrthoMCL (Li et al., 2003) disponible en VEuPathDB Galaxy Site (<https://veupathdb.globusgenomics.org/>) para asignar el set de proteínas predichas en los genomas a grupos de OrthoMCL usando la base de datos orthoMCL\_6r1 con parámetros por defecto. La descripción de las funciones de los grupos de OrthoMCL se descargaron de ([https://orthomcl.org/common/downloads/release-6.1/defines\\_OrthoMCL\\_6\\_1.txt.gz](https://orthomcl.org/common/downloads/release-6.1/defines_OrthoMCL_6_1.txt.gz)). Finalmente, se usó un script en Python personalizado para fusionar los genes predichos, grupos de OrthoMCL asignados y sus respectivas descripciones al archivo de anotaciones de cada genoma.

Esta misma metodología se aplicó para la anotación del genoma de *T. cruzi* de la cepa SOL secuenciado por el sistema PacBio. En este caso, el archivo de parámetros se generó usando como referencia el genoma de CL Brener non-Esmeraldo-like (v.50). En OrthoMCL se usó la versión orthoMCL\_6r4 de la descripción de los grupos, descargable en [https://orthomcl.org/common/downloads/release-6.4/defines\\_OrthoMCL-6.4.txt.gz](https://orthomcl.org/common/downloads/release-6.4/defines_OrthoMCL-6.4.txt.gz).

### 1.8. Análisis de las características de los genomas ensamblados

Las características y propiedades de los genomas ensamblados se determinaron mediante el uso de comandos en la terminal de Linux en Unix y scripts personalizados en Python ([www.python.org](http://www.python.org)) y Perl (*Practical Extraction and Reporting Language*) ([www.perl.org](http://www.perl.org)).

### 1.9. Identificación de secuencias de retroelementos

- Identificación de retroelementos en los 6 genomas de *T. cruzi* secuenciados por tecnología Ion Torrent.

Para la identificación de retrotransposones en los genomas de *T. cruzi* secuenciados por tecnología Ion Torrent, se emplearon sondas correspondientes a los extremos 5' y 3' de los elementos L1Tc o NARTc. Se realizaron dos búsquedas independientes en las lecturas resultantes de la secuenciación ejecutando un script de Python personalizado. La sonda 5' correspondía a la huella Pr77 (secuencia en **Anexo 1**). Esta se fragmentó en semillas de 8 nt de longitud con 7 nt solapantes. Las secuencias que contenían la huella Pr77 se extrajeron de las lecturas y se alinearon

con la sonda Pr77, estableciendo un umbral mínimo de identidad del 70%. Además, este umbral de identidad se ajustó en función de la longitud de la secuencia identificada según la siguiente fórmula:  $mic = mi + (1 - mi) * (1 - (\frac{hit\_length}{probe\_length}))$ , donde mic es la identidad mínima corregida, mi es la identidad mínima (0,7), hit\_length es la longitud del fragmento de la huella Pr77 encontrado en la lectura, y probe\_length es la longitud de la sonda buscada. Tan sólo se analizaron las secuencias superiores a 15 nt. Tras la identificación de la huella Pr77, se analizaron las secuencias corriente abajo de Pr77 para clasificarlas como L1Tc o NARTc en función de su identidad con las secuencias de 40 nt de longitud empleadas como sondas identificadoras (sonda L1Tc: 5'-CCCATCCGCTGCCCCGAGAGGCAGAGGCGCACAA-3' y sonda NARTc: 5'-TTCATGCTTCAAACCCGATGAGTAGTTGTTAGTTTT-3'). En este alineamiento, también se exigió una identidad mínima del 70% y un tamaño de secuencia de, al menos, 15 nt. Para la identificación de los extremos 3', se utilizaron como sonda los últimos 55 nt de L1Tc (5'- CCACCTTCGGCACTCAGATGGCACTGTAGCTAGACGCTGTAAGTAG-3') y de NARTc (5'-TACGCTTACCAGAGCTGCATTGAGGTGTAGTTACTACGGACTGTAAGAAG-3'), situados corriente arriba de las colas de poli(A) en cada caso, para identificar y clasificar los elementos como L1Tc o NARTc. En este caso, sólo se analizaron las secuencias con una identidad superior al 75% y mayores de 15 nt. Independientemente de la estrategia seguida para su identificación, las secuencias clasificadas se validaron mediante BLASTN (Altschul et al., 1990) frente a una base de datos constituida por las diferentes secuencias correspondientes a elementos L1Tc o NARTc (secuencias en **Anexo 1**) utilizando un script de Python personalizado. En este caso, se requirió una identidad de secuencia mínima del 70% y un tamaño de alineamiento de, al menos, el 70% para confirmar su clasificación como elementos L1Tc o NARTc.

- Identificación de retroelementos en genoma de SOL (secuenciación PacBio).

La identificación de las copias de retroelementos presentes en el genoma de la cepa SOL de *T. cruzi* (secuenciación PacBio) se realizó mediante BLASTN (Altschul et al., 1990) con parámetros por defecto. En concreto, se realizó la búsqueda de la secuencia de la huella Pr77, L1Tc, NARTc y 5 DIREs distintos (secuencias en **Anexo 1**). Una vez localizadas en el genoma ensamblado, se extrajo su secuencia a partir de sus coordenadas ejecutando el programa BEDTools v2.29.1 (Quinlan and Hall, 2010)



en la terminal de Linux. Se exigió un tamaño de alineamiento mínimo del 70% del tamaño del elemento a estudio. Adicionalmente, a las secuencias identificadas se le aplicaron filtros de tamaños superiores (> 80%, 90% y 95%). Para la identificación de elementos NARTc fue necesario el establecimiento de un umbral de identidad superior al 80%.

### **1.10. Identificación del contexto genómico de los retroelementos**

El análisis del contexto genómico de los retroelementos identificados en los genomas de *T. cruzi* secuenciados por tecnología Ion Torrent se llevó a cabo mediante la ejecución de un script de Python personalizado. Aplicando este, se extrajeron las secuencias situadas corriente arriba de la huella Pr77 y corriente abajo de los extremos 3' de los retroelementos identificados, exigiendo una longitud mínima de 15 nt fuera de la secuencia del retroelemento para continuar con el análisis. La identidad de las secuencias se obtuvo empleando la herramienta bowtie2 (Langmead et al., 2009) v2.2.0 (parámetros por defecto), mediante la cual se alinearon las secuencias extraídas con el archivo en formato *fasta* del genoma ensamblado y, una vez obtenidas sus coordenadas, con el archivo en formato *gtf* de anotaciones para hallar la descripción correspondiente a dicha región del genoma.

### **1.11. Determinación de los sitios de inserción de las secuencias Pr77Sh e identidad con Pr77 de L1Tc/NARTc**

Para analizar la localización genómica de las huellas Pr77 únicas o *single Pr77 hallmarks* (Pr77Sh), se extrajeron aquellas lecturas del genoma que estaban flanqueadas por una región de, al menos, 45 nt corriente arriba a Pr77 y 100 nt corriente abajo y contenían el motivo CGTG en su secuencia, mediante la ejecución de un script de Python personalizado. La identidad de dichas secuencias flanqueantes se determinó mediante análisis BLASTN (Altschul et al., 1990) frente a las secuencias disponibles en las bases de datos NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) y TriTrypDB (<http://tritrypdb.org/>).

La identidad de cada secuencia Pr77Sh con la secuencia Pr77 de los retroelementos L1Tc/NARTc se calculó mediante la herramienta ClustalW (Madeira et al., 2019) y el ajuste manual de los alineamientos.

### 1.12. Estimación del contenido génico en los genomas de *T. cruzi*

Para la estimación del número de copias de genes codificantes para las distintas proteínas o familias de proteínas y el porcentaje que representaban dichos genes en cada genoma ensamblado, se ejecutó un script en Perl (<https://bitbucket.org/ipbln/AnnotationStats/>), sobre el archivo de anotaciones (en formato *gtf*) correspondiente a cada genoma. Los genes analizados codificaban para las proteínas trans-sialidasas (TS), mucinas, proteínas de superficie asociadas a mucinas (MASP), la metaloproteasa de zinc GP63 (GP63), proteínas RHS (*Retrotransposon Hot Spot*), las proteínas TASV (Tripomastigote, Alanina, Serina, Valina), DGF-1 (*Dispersed Gene Family protein 1*), quinesinas, helicasas de ARN, proteínas quinasas, glicosiltransferasas, cisteína peptidasas, factor de elongación 1-gamma (EF1-gamma), la proteína de membrana de kinetoplástidos de 11 kDa (KMP11), proteínas de choque térmico de 70 kDa (HSP70), proteínas paraflagelares (PFR) y proteínas de unión a ARN. También se analizó el contenido de genes que codifican proteínas hipotéticas en todos los genomas.

## 2. DETERMINACIÓN DE LA FUNCIÓN RIBOZIMA (TIPO-HDV) DE LAS SECUENCIAS Pr77Sh

### 2.1. Obtención de células competentes de *E. coli* XL1-Blue

A partir de un inóculo cultivado durante toda la noche en LB conteniendo tetraciclina (10 µg/ml de medio), a 37°C se hizo una dilución 1:50 en medio de cultivo fresco, y se incubó a 37°C en agitación hasta que el cultivo alcanzara una absorbancia de densidad óptica (D.O.) de 0,2-0,4 a 600 nm. El cultivo se incubó en hielo 10 min y se centrifugó a 1.500 rpm durante 15 min a 4°C. A continuación, los pellets se resuspendieron con la mitad del volumen inicial de una solución de MgCl<sub>2</sub> 0,1 M fría y se recogieron por centrifugaron en las mismas condiciones. Los pellets se resuspendieron en 1 volumen (v/v) de CaCl<sub>2</sub> 0,1 M que representaba 1/20 parte del volumen inicial y se incubaron en hielo durante 30 min. Para su almacenamiento, se añadió 20% de glicerol a la mezcla que se alicuoteó y conservó a -80°C hasta su uso.

## 2.2. Transformación bacteriana por choque térmico

Para la obtención de ADN de plásmido se transformaron células de *Escherichia coli* XL1-Blue. Para ello, se empleó una alícuota de 100  $\mu$ l de células competentes por transformación, a la que se le añadió el ADN a transformar mezclando suavemente por pipeteo e incubó en hielo durante 30 min. La mezcla de células y ADN se sometió a un choque térmico, 42°C durante 45 seg, seguido de enfriamiento en hielo. A continuación, se añadieron 800  $\mu$ l de medio LB atemperado y se incubaron durante 1 hora a 37°C en agitación (~200 rpm). Las células se sembraron en placas LB-agar conteniendo el antibiótico de selección del plásmido correspondiente (ampicilina a 100  $\mu$ g/ml de medio) e incubaron en estufa, a 37°C, durante 12-16 horas.

## 2.3. Selección de bacterias transformadas por el gen de la $\beta$ -galactosidasa

Para la selección de los clones realizados en el vector pGEM-T (Promega) se empleó el método conocido como “azules y blancas” dado que el marcador genético de inserción del gen de la  $\beta$ -galactosidasa de *E. coli* forma parte del plásmido pGEM-T, de manera que al transformar bacterias lac<sup>-</sup> con pGEM-T, éstas pasan a ser lac<sup>+</sup> y las colonias se colorean de azul cuando se cultivan en placas LB-agar conteniendo la sustancia cromogénica X-Gal y el inductor de expresión de proteínas IPTG. Por ello, tras el clonaje de fragmentos de ADN en pGEM-T y transformación de la bacteria con la ligación, las células se siembran en placas LB-agar con el antibiótico de selección, 0,2 mM de IPTG y 0,195 mM de X-Gal. Si el plásmido contiene el inserto, se interrumpe del gen LacZ, por lo que el X-Gal no puede ser metabolizado y la colonia se mantiene de color blanco. De esta forma, se seleccionaron las colonias blancas obtenidas y se crecieron en medio LB líquido, suplementado con ampicilina, con la finalidad de obtener masa del cultivo para poder extraer el ADN plasmídico correspondiente.

## 2.4. PCR colonia

Para verificar que las colonias blancas contienen el inserto, se analizó la presencia del inserto mediante PCR de cada colonia seleccionada (PCR-colonia). Para ello, cada una de las colonias seleccionadas se inoculan en 20  $\mu$ l de agua estéril

y se lisan tras calentar a 100°C durante 5 min. Tras centrifugación a 13.000 rpm durante 2 min, se recoge el sobrenadante que se emplea como ADN molde para la PCR. Para ello, se empleó 4 µl de cada sobrenadante y 10 pmol de cada par de oligonucleótidos usando 0,2 U de Taq ADN polimerasa (Biotools), 2,5 mM de dNTPs, DMSO 0,01% y tampón de reacción 1X en un volumen final de 10 µl. Las condiciones de PCR fueron 94°C 2 min; 94°C 30 s, 52°C 30 s, 72°C 30 s (x20); 72°C 7 min.

### 2.5. Electroforesis de ADN en geles de agarosa

La resolución de productos amplificados en la PCR, se realizó en geles de agarosa al 2% (p/v). Para la visualización de los fragmentos amplificados, a cada gel se le añadió la solución intercalante de ácidos nucleicos *REALSAFE Nucleic Acid Staining Solution* (REAL) a concentración final de 1X. A las muestras se les añadió tampón de carga (*'Materiales', 'Tampones y soluciones'*) y se cargaron en los geles para llevar a cabo la electroforesis en cubetas de desarrollo horizontal (Bio-Rad), en tampón de electroforesis TAE 1X (*'Materiales', 'Tampones y soluciones'*). Se aplicó un voltaje de 100 V durante aproximadamente 60 min. Para la visualización de los fragmentos de ADN se utilizó el transiluminador *Universal Hood II* (BioRad).

### 2.6. Purificación de ADN de plásmidos y secuenciación

Para purificación del ADN plasmídico de cada clon se realizaron minipreparaciones empleando el kit comercial *'Real Miniprep Turbo kit'* (REAL) siguiendo las recomendaciones del fabricante. Para identificación de la secuencia, el ADN de los distintos clones se secuenció siguiendo la metodología de *Sanger* (Sanger et al., 1977).

### 2.7. Construcción de moldes de ADN para transcripción *in vitro*

Para determinar la actividad ribozima tipo HDV de cada huella Pr77 seleccionada se realizó transcripción *in vitro* de moldes de ADN generados tras clonar cada una de las huellas en el vector pGEM-T Easy (Promega), amplificación por PCR de cada inserto empleando como cebador en 5' para todos los clones el oligonucleótido universal M13Fw(-20) *universal primer* y un cebador antisentido que mapea en el extremo 3' de cada una de las huellas, transcripción *in vitro* empleando T7 ARN polimerasa y visualización del/los productos generados en gel

desnaturalizante de acrilamida al 8%. Cada clon/molde de ADN conteniendo cada una de las huellas se generó tras anillamiento de dos oligonucleótidos cuya secuencia correspondía a los extremos 5' (cebador sentido) y 3' (cebador antisentido) de cada una de las huellas a analizar los cuales solapaban en, al menos, 20 nucleótidos, por sus extremos 3' (ver secuencias en '*Materiales*', '*Oligonucleótidos*') y posterior amplificación por PCR. Dicha PCR se realizó en un volumen de 50  $\mu$ l conteniendo 1X *High Fidelity buffer*, 5  $\mu$ l dNTP (2,5 mM cada uno), 30 pmol de cada cebador, DMSO 0,01% y 3,5 U de Taq ADN polimerasa (*Expand High Fidelity*, Roche). Las condiciones de PCR fueron: desnaturalización a 94°C x 2 min; 20 ciclos de amplificación (94°C x 30 s, 52°C x 30 s, 72°C x 30 s) y extensión a 72°C por 7 min.

Los productos de PCR se purificaron del gel de agarosa al 2% empleando el kit *GeneClean* (Promega) y clonaron en el vector pGEM-T Easy (Promega). La identidad de la secuencia de cada clon se verificó tras secuenciación de cada clon [IK-51 (51IKf y 51IKr), IK-38 (8IKf y 38IKr), IK-8 (8IKf y 8IKr), BM-1 (Pr77f y 1BMr), BM-2 (2BMf y 2BMr), Y-15 (15Yf y 15Yr), SOL-38 (8IKf y 38SOLr) y CL Brener-102 (102CLf y 102CLr)].

Para la transcripción *in vitro* de cada huella se generaron los moldes de ADN mediante PCR empleando el oligonucleótido M13Fw(-20) *universal primer* y un cebador antisentido que mapea en el extremo 3' de cada una de las huellas. Las condiciones de la PCR fueron: 94°C x 2 min; 94°C x 30 s, 52°C x 30 s, 72°C x 30 s (x25 ciclos); 72°C x 7 min. Los productos de PCR, que contenían el promotor de la T7 ARN polimerasa del vector pGEM-T Easy, secuencia del plásmido pGEM-T Easy de 61 nt y la secuencia de cada Pr77Sh, se purificaron del gel de agarosa al 2% mediante extracción fenólica. Como control se empleó el molde correspondiente a L1TcRz (oligonucleótidos Pr77f y Pr77r).

## 2.8. Ensayos de corte co-transcripcional de huellas Pr77

22 ng de los moldes de ADN generados por PCR en la sección de *Métodos 2.7* se transcribieron empleando el kit T7 ARN polimerasa (Promega) a 42°, 37° y 25°C, durante 2 h, siguiendo las recomendaciones de la casa comercial con la excepción de que las reacciones se realizaron en un volumen de 10  $\mu$ l, conteniendo 1X tampón de transcripción, 0,7  $\mu$ Ci de  $\alpha$ -32P UTP, 1 mM de cada rNTP (0,4 mM de UTP frío), 10

mM DTT, 0,4 U RNAsa-in y 0,4 U de la enzima T7 ARN polimerasa. También se transcribieron a 37°C en las mismas condiciones descritas, a una concentración de rNTP hasta 100 veces menor. A cada reacción se le añadieron 10 µl de tampón de carga desnaturizante 2X (*'Materiales', 'Tampones y soluciones'*) resolviéndose los productos en geles de poliacrilamida al 8%, urea 7 M desarrollados en tampón 1X TBE como se describe en la sección *Métodos 2.10*.

### 2.9. Cinética de corte post-transcripcional de huellas Pr77

La cinética de corte post-transcripcional se realizó sobre los productos de transcripción *in vitro* no cortados en la reacción de corte co-transcripcional. Para ello, se realizó una transcripción *in vitro* empleando, aproximadamente, 45 ng de cada uno de los moldes de ADN (generados en la sección *Métodos 2.7*) y el kit T7 ARN polimerasa (Promega) en un volumen de 100 µl conteniendo 1X tampón de transcripción, 10 mM DTT, 0,4 U RNAsa-in (Promega), 0,5 mM de cada rNTP (salvo UTP, 0,2 mM), 30 µCi de  $\alpha$ -<sup>32</sup>P UTP y 0,4 U de la enzima T7 ARN polimerasa (Promega). La reacción se incubó a 37°C durante 2 h y paró añadiendo 100 µl de tampón de carga desnaturizante 2X. Los productos se resolvieron en geles de poliacrilamida al 8%, urea 7 M en buffer 1X TBE, a 36 mA durante 2 horas. Tras la electroforesis, empleando la señal detectada en la autorradiografía del transcrito marcado sin cortar, se escindió el transcrito sin cortar del gel y se eluyó en tampón TEN<sub>250</sub> (*'Materiales', 'Tampones y soluciones'*) toda la noche a 4°C, purificándose por fenolización el día posterior.

250 cps de cada ARN no cortado correspondiente a cada uno de los moldes se renaturalizó en tampón Tris-HCl 1 mM (pH 7,5) y EDTA 0,1 mM, a 85°C durante 5 min seguido de incubación a 25°C durante 10 min. Las reacciones se atemperaron a 37°C durante 2 min considerándose este como tiempo 0. Así, se tomaron 8 µl de cada reacción, se le añadió 1 volumen de tampón desnaturizante 2X y se conservó a -20°C hasta su carga en gel (tiempo 0). Al volumen restante, se le añadieron 12 µl de tampón de corte 5X (*'Materiales', 'Tampones y soluciones'*) a una concentración final de magnesio (0, 0,1, 1 o 10 mM MgCl<sub>2</sub>). En ese momento comenzó la cinética de corte, tomándose 10 µl de cada reacción en diferentes tiempos. Las alícuotas obtenidas se conservaron a -80°C, con la previa adicción de 1 volumen de tampón de carga desnaturizante 2X, hasta el momento de ser cargadas en un gel. Los

productos de las reacciones se resolvieron en geles desnaturizante de poliacrilamida 8%, urea 7 M como se describe en la sección *Métodos 2.10*.

### **2.10. Electroforesis en geles de poliacrilamida-urea**

Se emplearon geles desnaturizantes de poliacrilamida al 8%, urea 7 M y TBE 1X de diferente grosor (0,8 mm para ensayos de actividad y 1,5 mm para purificar fragmentos). Para un gel de 60 ml corresponden: 25,2 g urea, 6 ml TBE 10X, 12 ml acrilamida:bisacrilamida (19:1), 375  $\mu$ l APS 10%, 37,5  $\mu$ l de TEMED y H<sub>2</sub>O hasta 60 ml. Antes de la electroforesis, se llevó a cabo una pre-electroforesis de aproximadamente 30 min en las mismas condiciones. Se cargaron las muestras preparadas en tampón de carga desnaturizante 2X para, seguidamente, aplicar un amperaje de 36 mA por gel durante 60-120 min con TBE 0,5X como tampón de electroforesis (*'Materiales', 'Tampones y soluciones'*). Una vez finalizada, los geles se secaron al vacío a 80°C en el equipo *Slab Gel Dryer SGD 4050* (SAVANT), y se expusieron en pantallas *Storage Phosphor Screen* (Amersham Biosciences) toda la noche a -80°C, revelándose el día posterior empleando el equipo *Typhoon Fla 9400 Phosphorimager* (Amersham Biosciences). Las imágenes resultantes se analizaron empleando el programa ImageQuant v5.2 (GE Healthcare) y los datos se representaron usando GraphPad Prism v6.0 (GraphPad Software, San Diego, CA).

### **2.11. Elución de ARN desde gel de poliacrilamida**

Los ARN transcritos *in vitro* tras la resolución electroforética en geles al 8% de poliacrilamida-urea se localizaron mediante una película autorradiográfica que permitió detectar las bandas de interés. Estas se escindieron empleando una cuchilla estéril, se trocearon y transfirieron a tubos de 1,5 ml. Se incubaron en agitación suave con tampón de elución TEN<sub>250</sub> durante toda la noche a 4°C. A continuación, se llevó a cabo una extracción con fenol y, seguidamente, con cloroformo:alcohol isoamílico (24:1) (v/v). El ARN se precipitó añadiendo 0,1 volumen de AcONa 3 M (pH 5,2), 2 - 2,5 volúmenes de EtOH absoluto frío y 1  $\mu$ l de glicógeno. Tras precipitar durante toda la noche a -80°C, se aisló centrifugando a 13.000 rpm durante 1 hora a 4°C. El pellet resultante se dejó secar al aire y se resuspendió en 20  $\mu$ l de agua MilliQ estéril para ser seguidamente divididas en 4 alícuotas de 5  $\mu$ l antes de ser conservadas a -20°C.

### **3. ANÁLISIS DE EXPRESIÓN GÉNICA EN LOS DIFERENTES ESTADIOS DE *T. cruzi* MEDIANTE SECUENCIACIÓN DEL TRANSCRIPTOMA DE NUEVA GENERACIÓN (ARN-Seq)**

#### **3.1. Cultivo de formas epimastigotes**

Las formas epimastigotes se cultivaron en medio líquido LIT, suplementado con un 10% de SFBi, en frascos de cultivo estériles no ventilados. Los parásitos se cultivaron en agitación suave a 28°C, partiendo de una concentración inicial de  $1 \times 10^6$  parásitos/ml. Para mantener los cultivos se realizaron pases cada 4 o 5 días y se recolectaron cuando se alcanzó el volumen deseado de cultivos en fase de crecimiento  $4 \times 10^6$  parásitos/ml, aproximadamente, mediante centrifugación a 2.500 rpm durante 15 min. Se conservaron a -80°C hasta su uso.

#### **3.2. Obtención de formas tripomastigotes y amastigotes**

Se cultivaron células LLC-MK2 en medio DMEM suplementado con 2 mM L-glutamina, 10% de SFBi y 50 µg/ml de gentamicina a 37°C en atmósfera con 5% de CO<sub>2</sub>. Cuando la monocapa de células alcanzó el 50% de confluencia, estas se infectaron con tripomastigotes metacíclicos de *T. cruzi* (cepa SOL), aislados de ratones infectados con *T. cruzi*, en una proporción parásito:célula 4:1 durante 12 horas. Entre 96-120 horas tras la infección, comenzaron a aparecer formas tripomastigotes y amastigotes de *T. cruzi* en el sobrenadante de las células LLC-MK2 infectadas. Este sobrenadante se recolectó y se eliminaron las células LLC-MK2 y amastigotes centrifugando a 800 rpm durante 1 min. El sobrenadante se volvió a centrifugar a 2.500 rpm durante 15 min e incubó 2 h a 37°C. Pasado ese tiempo, los tripomastigotes se recolectaron del sobrenadante y se centrifugaron 15 min a 3.000 rpm, el pellet resultante se congeló a -80°C hasta su uso.

Las formas amastigotes se aislaron del pellet obtenido tras las 2 h de incubación previamente descritas. A este pellet se le añadieron 3 ml de medio DMEM y se depositó en una placa. Tras 30 min, se retiró el medio con objeto de eliminar restos celulares y los tripomastigotes que no habían sedimentado en el fondo de la placa. Seguidamente se añadieron 3 ml de medio DMEM y se incubó 4 h a 37°C. Transcurrido ese tiempo, las células vivas se encuentran adheridas a la placa, los amastigotes pegados a las células y los tripomastigotes en el sobrenadante. Se retiró



el medio para eliminar la mayoría de tripomastigotes y se volvió a añadir medio DMEM para recolectar los amastigotes por centrifugación a 3.000 rpm durante 15 min y congelaron a -80°C hasta su uso.

### 3.3. Secuenciación del transcriptoma de nueva generación (ARN-Seq)

Este ensayo se realizó sobre las tres formas de *T. cruzi* de la cepa SOL: epimastigotes, tripomastigotes y amastigotes. Se tomaron entre 80 y 104x10<sup>6</sup> parásitos de cada forma. La extracción de ARN de los parásitos se llevó a cabo lisando los parásitos con NP40 1% y empleando el kit '*RNeasy Plus mini kit*' (Qiagen) siguiendo las instrucciones del fabricante. El ARN obtenido, libre de ADN, se cuantificó mediante espectrofotometría (*NanoDrop 1000*, Thermo Scientific), la calidad del ARN se determinó mediante el análisis de su integridad por electroforesis microfluídica en chip (*Bioanalyzer 2100*, Agilent Technologies) y confirmó su concentración por fluorometría usando el equipo *Qubit* (Invitrogen).

Para la secuenciación ARN-Seq se utilizaron tres réplicas diferentes de las diferentes formas epimastigote, tripomastigote y amastigote. Así, se prepararon nueve bibliotecas de ARNm utilizando 1 µg de material de partida de ARN y el kit de preparación de librerías de ARNm '*TruSeq Stranded mRNA Library Prep kit*' (Illumina, CA, USA) de acuerdo a las instrucciones del fabricante. El tamaño y la distribución de la calidad de las bibliotecas de ARNm indexadas se validaron mediante el ensayo de ADN de alta sensibilidad *Bioanalyzer* empleando el equipo *Bioanalyzer 2100*. La concentración de la muestra se midió en el fluorómetro *Qubit* (Invitrogen). Las bibliotecas finales se combinaron equimolecularmente y luego se diluyeron y desnaturalizaron como se recomienda. La secuenciación "2x75nt paired-end" se realizó en un secuenciador *NextSeq 500* (Illumina, CA, USA) (*mid output mode*: 150 ciclos).

### 3.4. Análisis de datos procedentes de secuenciación ARN-Seq

Para el análisis de muestras transcriptómicas se usó el flujo de trabajo de miARma-Seq (Andrés-León and Rojas, 2019). Este flujo de trabajo incluye todos los pasos, desde los datos crudos sin procesar hasta el cálculo de genes con expresión diferencial. En primer lugar, los datos crudos se evaluaron con el software FastQC

para analizar la calidad de las lecturas (Andrews, 2010). El 94% de las lecturas mostró calidad  $> 30$  y no se observó acumulación de adaptadores. Posteriormente, la herramienta miARma-Seq alineó todas las secuencias usando HISAT2 (Kim et al., 2015) lo que resultó en un 83% de lecturas alineadas correctamente. Para ello, se utilizó el genoma secuenciado, ensamblado y anotado por nosotros descrito previamente en la sección de *Resultados 1.7*.

### ***3.4.1. Análisis de expresión génica diferencial de datos ARN-Seq***

Para realizar el análisis de expresión génica diferencial se utilizó el paquete edgeR (Nikolayeva and Robinson, 2014). Los genes de baja expresión se eliminaron y los restantes se normalizaron mediante el método TMM (*Trimmed Mean of M values*) (Robinson and Oshlack, 2010). Además, se calcularon lecturas por kilobase por millón de lecturas mapeadas (RPKM), cuentas por millón (CPM) y  $\log_2$  de cuentas por millón ( $\log$ -CPM) por gen en cada muestra (Nikolayeva and Robinson, 2014).

Para obtener una descripción general de la similitud de las muestras de secuenciación de ARN, con el fin de inferir la replicabilidad de las muestras, se empleó el análisis de componentes principales (PCA) y la agrupación jerárquica de muestras normalizadas (Reeb et al., 2015; Ritchie et al., 2015). Posteriormente, se analizaron todas las muestras, ya que no se detectó ningún valor atípico evidente.

Los genes con expresión génica diferencial entre los diferentes estadios del parásito se obtuvieron comparando las tres réplicas de muestras de amastigote frente a muestras de epimastigote, muestras de tripomastigote frente a muestras de epimastigote y, finalmente, muestras de amastigote frente a muestras de tripomastigote. Se utilizó el valor  $\log_2$  FC para evaluar el cambio en la expresión de un gen entre muestras pertenecientes a diferentes grupos, estableciendo un umbral de  $\log_2$  FC  $> 1$  o  $< -1$  para aceptar la existencia de expresión génica diferencial, junto con un valor de FDR (*False Discovery Rate*)  $< 0,05$ . Estas mismas comparaciones y criterios se emplearon para determinar la expresión diferencial entre estadios de las regiones del genoma correspondientes a retroelementos. Los resultados de ambos análisis comparativos se representaron en gráficos de volcán y gráficos de barras empleando los programas SPSS Statistics v25 (SPSS Inc., Chicago, Illinois, Estados Unidos) y GraphPad Prism v8.0 (GraphPad Software, San Diego, CA).

#### 4. DETERMINACIÓN DEL NIVEL DE ANTICUERPOS ESPECÍFICOS DE ANTÍGENOS EN EL SUERO DE PACIENTES CON ENFERMEDAD DE CHAGAS

##### 4.1. Aislamiento de antígenos solubles de *T. cruzi* (STcA)

Para la obtención del lisado de antígenos solubles de la cepa SOL de *T. cruzi*, las formas tripomastigotes y amastigotes se obtuvieron siguiendo el protocolo que se describe en la sección de *Métodos 3.2*.

Los tripomastigotes:amastigotes (en proporción 1:1) presentes en el sobrenadante de las células LLC-MK2, se centrifugaron a 2.500 rpm y se lavaron dos veces con PBS 1X y lisaron añadiendo tampón de lisis de parásitos (1 µl por cada  $1 \times 10^6$  parásitos) (*'Materiales', 'Tampones y soluciones'*). Para completar la lisis, las muestras se sometieron a 3 ciclos de sonicación de 40 seg con pulsos de 50-62 kHz, descansando 20 seg entre ciclos. El extracto se centrifugó a 10.000 rpm durante 20 min a 4°C. El sobrenadante obtenido, donde se encontraban los antígenos solubles de *T. cruzi*, se alicuoteó y conservó a -80°C hasta su uso.

La concentración de las proteínas del extracto obtenido se determinó usando el método del ácido bicinconínico (BCA). El perfil de proteínas e integridad se analizó por electroforesis SDS-PAGE.

##### 4.2. Electroforesis de proteínas mediante SDS-PAGE

La electroforesis de proteínas se realizó en geles desnaturalizantes de poliacrilamida SDS-PAGE empleando el sistema *Mini-Protean III Cell* (Bio-Rad). Se usaron geles con diferente concentración de poliacrilamida, entre un 8-10%, en función del tamaño de las proteínas a resolver. Antes de ser cargadas en el gel, a las muestras se les añadió un volumen de tampón Laemmli (*'Materiales', 'Tampones y soluciones'*) y se desnaturalizaron a 95°C durante 5 min. La electroforesis se realizó en tampón de electroforesis (*'Materiales', 'Tampones y soluciones'*) aplicando 100 V durante 120 min. En cada gel, se incluyó el marcador de peso molecular *Precision Plus Protein All Blue Standards* (Bio-Rad). Los geles se tiñeron con azul *Coomassie* siguiendo el método *Fairbanks* de tinción rápida (Sambrook et al., 1989). En este método se utilizan 4 soluciones diferentes de forma secuencial. Brevemente, el gel se sumerge en la primera solución que contiene isopropanol al 25%, ácido acético al 10% y azul *Coomassie* al 0,05% (solución A). A continuación, se calentó en el microondas hasta cerca de ebullición y se agitó en balancín 5 min. Estos pasos se

repiten con las soluciones B (isopropanol al 10%, ácido acético 10% y azul *Coomassie* 0,005%), C (ácido acético 10% y azul *Coomassie* 0,002%) y D (ácido acético 10%).

### 4.3. Cuantificación de proteínas

La concentración de las proteínas purificadas, así como de las proteínas totales extraídas de *T. cruzi*, se determinó mediante el método del ácido bicinconínico (BCA), empleando el kit '*Micro BCA™ Protein Assay kit*' (Thermo Scientific). En cada determinación se utilizó una curva patrón creada a partir de diferentes concentraciones de la proteína albúmina de suero bovino (BSA).

### 4.4. Determinación del nivel de anticuerpos antígeno-específicos mediante la técnica ELISA

La presencia y nivel de anticuerpos específicos frente las proteínas recombinantes identificadas y caracterizadas en el laboratorio: KMP11, HSP70, PFR2, el péptido 3973<sup>d</sup> y los antígenos solubles totales de *T. cruzi* (*STcA*) se determinaron en suero de pacientes con Chagas mediante ELISA. Para ello, se fijaron placas de 96 pocillos de fondo plano con 0,5 µg de proteínas, purificadas por otros miembros del laboratorio, o 1 µg en caso de *STcA*, por pocillo las cuales se diluyeron en 100 µl de tampón carbonato/bicarbonato ('*Materiales*', '*Tampones y soluciones*') y se almacenaron a -20°C hasta su uso. En el momento de uso, las placas se atemperaron por 15 min, se incubaron a 37°C por 30 min y se lavaron los pocillos 5 veces con 200 µl de PBS 0,05% Tween-20. Se añadieron 300 µl/pocillo de solución de bloqueo (PBS 1X + 5% leche desnatada)/pocillo y las placas se incubaron a 37°C durante 2 h. Se añadieron los sueros de los pacientes diluidos en PBS 1X + 0,05% Tween + 5% leche desnatada) en un volumen final de 100 µl/pocillo y se incubaron a 37°C por 2 h. Tras lavar los pocillos 5 veces con PBS 1X + 0,05% Tween, se añadió el anticuerpo secundario (anti-IgG acoplado a peroxidasa (Life Technologies) a una dilución 1:1.000, diluido en PBS 1X + 0,05% Tween + 5% leche desnatada incubándose a 37°C durante una hora. Tras lavar 5 veces con la solución de lavado, la reacción se reveló añadiendo 100 µl/pocillo de solución ortofenildiamina ('*Materiales*', '*Tampones y soluciones*') durante 5 min, en oscuridad, a temperatura ambiente. La reacción se detuvo con ácido sulfúrico 2 N (50 µl/pocillo) y se midió la

densidad óptica a 492 nm en un fotómetro para microplacas (*Multiskan™*, Thermo Scientific).

En todos los ensayos se incluyeron sueros control, positivos y negativos, y pocillos blancos, los cuales indicaban los valores de fondo al encontrarse fijados con el antígeno y sufrir todo el proceso en paralelo salvo la adicción de los sueros de los pacientes. Las muestras se ensayaron por triplicado y a 2 y 3 diluciones diferentes. Para el análisis de la significancia estadística de las diferencias en las D.O. entre grupos, se empleó el método no paramétrico de Wilcoxon o la prueba de Friedman seguida de la prueba Post Hoc de Dunn para comparaciones múltiples. Se consideraron estadísticamente significativas aquellas con  $p < 0,05$ .

- Algoritmo para la determinación de la eficacia terapéutica

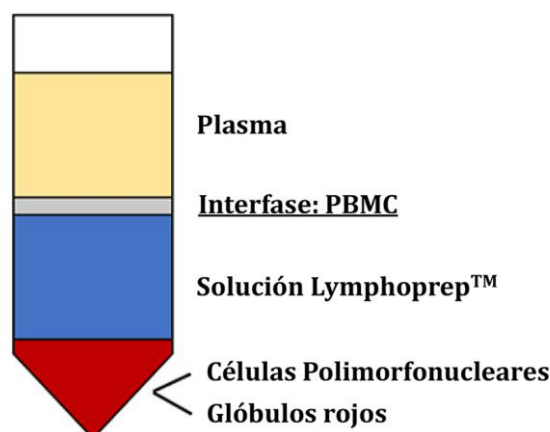
Para establecer el criterio de eficacia terapéutica en los pacientes tratados, se diseñó un script personalizado en Perl mediante el cual se procesaron los datos de reactividad de los pacientes frente a los antígenos ensayados. El programa se diseñó para incluir los datos de densidad óptica de, al menos, 3 tiempos por cada paciente (incluyendo el T0). Aplicando este algoritmo, se podía valorar si el tratamiento había sido eficaz o se había producido fracaso terapéutico, tal como se describe en (Egui et al., 2019).

El criterio de eficacia terapéutica estándar (STEC) se definió como una disminución continua y constante de la reactividad de los sueros de los pacientes frente a los cuatro biomarcadores (BMKs) después del tratamiento ( $T9 \leq T0$  o  $T9$  no más del 20% de  $T0$ ;  $T24/T48 < T9$  y  $T24/T48 < T0$ ) y que la misma fuese notable ( $> 40\%$  para KMP11, PFR2 y 3973<sup>d</sup> y  $> 30\%$  para HSP70) en relación con la detectada antes del tratamiento para, al menos, dos de los cuatro BMK. También se diseñó un criterio más restrictivo (RTEC) en el que, además de la caída continua y constante en la reactividad frente a los 4 BMKs, se exigía que la reducción en la reactividad fuese  $> 40\%$  para KMP11, PFR2 y 3973<sup>d</sup> y  $> 30\%$  para HSP70 para, al menos, tres de los cuatro antígenos.

### 5. ANÁLISIS DE LA EXPRESIÓN GÉNICA RELATIVA EN PBMC DE PACIENTES CON ENFERMEDAD DE CHAGAS

#### 5.1. Aislamiento de células mononucleares de sangre periférica (PBMC)

A cada uno de los sujetos incluidos en este estudio se les extrajo aproximadamente 30 mililitros de sangre mediante punción venosa, en tubos con anticoagulante EDTA. Entre 16 - 18 h después de la recolección de las muestras de sangre, otros miembros del laboratorio aislaron células mononucleares de sangre periférica (PBMC). Para ello, se vertió la sangre sobre un volumen de Lymphoprep™ (Axis-Shield), evitando que ambos se mezclasen y se centrifugó sin aceleración ni freno durante 35 min a 800 x g a temperatura ambiente. Una vez finalizada la centrifugación, debido a este medio con gradiente de densidad, se generaron diferentes fases (esquema en **Figura M1**). Las PBMC se recolectaron de la interfase, donde forman un anillo, con una pipeta Pasteur y centrifugaron a 1.500 rpm a temperatura ambiente durante 5 min. Con objeto de lisar los eritrocitos que pudiese contener la muestra, al pellet obtenido tras la centrifugación, se le añadieron 5 ml de tampón de lisis de eritrocitos (*'Materiales', 'Tampones y soluciones'*) incubándose 3 min a temperatura ambiente. Pasado este tiempo, rápidamente se añadió PBS 1X y se centrifugó a 1.500 rpm, 5 min a 4°C, seguido de 2 lavados con PBS 1X aplicando centrifugaciones de 1.500 rpm, 10 min a 4°C. Las células obtenidas se diluyeron en suero bovino fetal inactivado por calor (SFBi) con DMSO al 10% (Sigma-Aldrich) para ser criopreservadas en nitrógeno líquido hasta su uso.



**Figura M1.** Fases obtenidas tras la centrifugación de la muestra de sangre periférica con la solución Lymphoprep™ durante la técnica de aislamiento de células mononucleares de sangre periférica (PBMC).

## 5.2. Descongelación, cultivo y estimulación de PBMC

Las PBMC se descongelaron en un baño de agua a 37°C, se transfirieron a tubos que contenían 10 ml de medio RPMI 1640 con 2 mM de L-glutamina (Gibco), 10% de SFBi y 50 µg/ml de gentamicina y centrifugaron a 1.500 rpm durante 10 min a temperatura ambiente. Tras la centrifugación, al pellet de PBMC se le añadieron 2 ml de medio RPMI 1640 suplementado y se procedió al conteo del número de células viables de forma manual usando azul tripán (Sigma-Aldrich). El fundamento de la tinción con azul tripán se basa en que las células vivas, con membrana citoplasmática íntegra, excluyen el colorante y no se tiñen, mientras que, en las células muertas con lesión de la membrana celular, este colorante penetra, lo que hace que se vean de color azul. Seguidamente, todas las PBMC se añadieron a placas de 12 pocillos a una concentración final entre 7,5 y 8,5x10<sup>6</sup> células en un volumen máximo de 3,5 ml por pocillo, cultivadas a 37°C durante, aproximadamente, 4 h en atmósfera de 5% de CO<sub>2</sub> para permitir el equilibrio de la expresión génica basal bajo condiciones de crecimiento *in vitro*. Posteriormente, las células se estimularon con STcA (10 µg/ml) y cultivaron durante 14 – 14,5 h a 37°C en atmósfera con 5% de CO<sub>2</sub>.

## 5.3. Aislamiento de ARN, cuantificación y determinación de la calidad

De las PBMC estimuladas, se obtuvo ARN total usando el kit '*RNeasy Plus mini kit*' (Qiagen) siguiendo las instrucciones del fabricante. Así, se obtuvieron muestras de ARN total enriquecidas en ARNm y libres de ADN, dado que las columnas del kit empleado desechaban aquellos fragmentos de ARN con un tamaño inferior a 200 nucleótidos (como los ARN ribosómicos y ARN transferentes), así como el ADN genómico.

El ARN extraído se cuantificó mediante espectrofotometría (*NanoDrop 1000*, Thermo Scientific). Su concentración se confirmó por fluorometría (*Qubit*, Invitrogen), así como se determinó la calidad del ARN mediante el análisis de su integridad por electroforesis microfluídica en chip (*Bioanalyzer 2100*, Agilent Technologies). Con esta técnica se obtiene un parámetro denominado RIN (*RNA Integrity Number*), el cual basándose en la ratio de las áreas de los ARN 28S y 18S, permite determinar la integridad del ARN, adquiriendo mayor valor mientras mayor integridad posea la muestra. Las muestras incluidas en el ensayo poseían valores de

RIN de entre 7,4 y 10, siendo el RIN promedio de 9,1, y se conservaron a -80°C hasta su uso.

### 5.4. Transcripción inversa

La síntesis de ADNc por transcripción inversa se llevó a cabo partiendo de 2 µg de ARN total mediante el uso del kit '*High-Capacity cDNA Reverse Transcription kit*' (Applied Biosystems) siguiendo las indicaciones del fabricante. Se añadieron 10 µl de ARN total (200 ng/µl) a 10 µl de '*reverse transcription mix*', obteniéndose un volumen final de 20 µl. Las condiciones del termociclador (T100, BioRad) se configuraron en 10 min a 25°C, 2 h a 37°C y 5 min a 85°C seguido de un enfriamiento rápido a 4°C. Las muestras de ADNc resultantes se almacenaron a -20°C hasta su uso.

### 5.5. Diseño del ensayo. Selección de genes

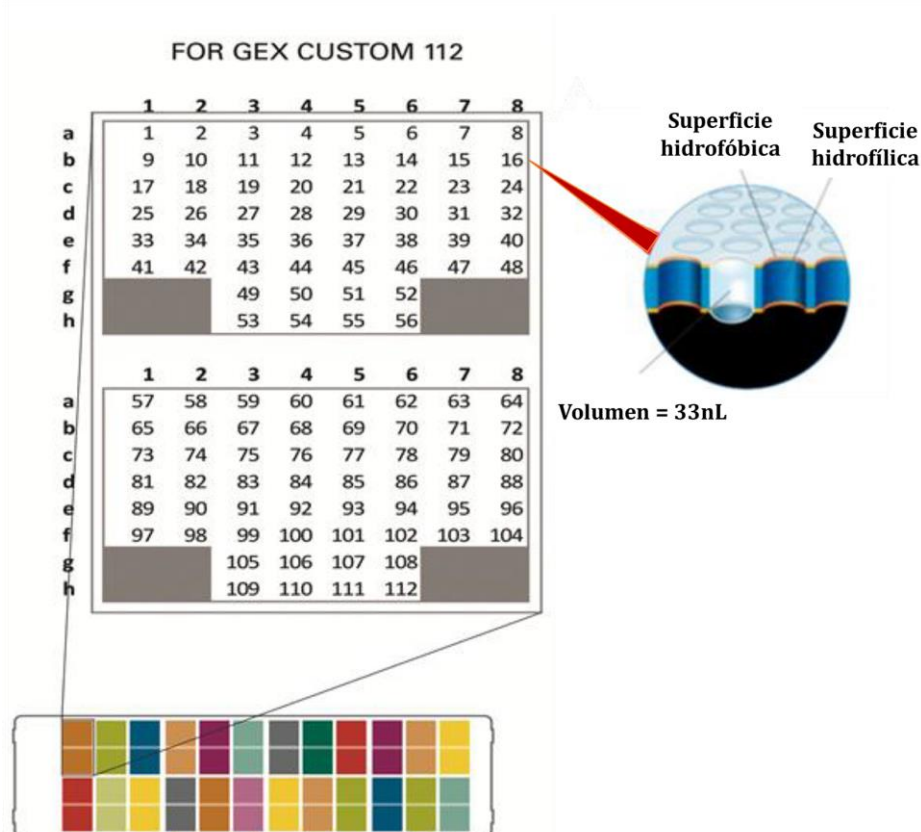
En el estudio se analizó la expresión de 106 genes implicados en diferentes mecanismos del sistema inmune. Así, genes codificantes de citoquinas, interleuquinas y sus receptores, factores de transcripción, receptores inhibitorios y sus ligandos, marcadores de fenotipo, moléculas coestimuladoras, marcadores de células dendríticas, moléculas de adhesión, citotóxicas, con implicación en apoptosis o senescencia y otras moléculas con implicación inmunológica (enzimas, marcador de proliferación...). Además, se incluyeron 6 genes, de los cuales se esperaba una expresión constitutiva (genes *housekeeping*). En la **Tabla M3** se muestra el listado de genes ensayados y la clasificación de estos atendiendo a la naturaleza o función del producto que codifican. Todos los genes se encontraban disponibles en forma de sonda '*TaqMan® Gene Expression Assays*' (Applied Biosystems). Las sondas seleccionadas fueron de la especie *Homo sapiens* y cumplían los siguientes requisitos: amplificar regiones exón-exón, no amplificar ADN genómico, no amplificar secuencias similares y cuyo colorante fuese FAM-MGB. La relación de genes seleccionados y sus correspondientes sondas TaqMan® se muestran en el **Anexo 2**.



**Tabla M3. Genes a estudio (n=112), identidad y papel funcional.** Se muestran los 106 genes seleccionados por su implicación en diversos procesos inmunológicos y los 6 genes con expresión constitutiva. En color se distinguen las diferentes categorías. *IL*: interleuquinas.

Nº	Gen	Clasificación	Nº	Gen	Clasificación
1	ACTB	Genes constitutivos	57	IL10	Citoquinas / IL y sus receptores
2	B2M	Genes constitutivos	58	IL10RA	Citoquinas / IL y sus receptores
3	B3GAT1	Apoptosis / Senescencia	59	IL10RB	Citoquinas / IL y sus receptores
4	BCL2	Apoptosis / Senescencia	60	IL12A	Citoquinas / IL y sus receptores
5	BTLA	Receptores inhibitorios y sus ligandos	61	IL12B	Citoquinas / IL y sus receptores
6	CASP3	Apoptosis / Senescencia	62	IL12RB1	Citoquinas / IL y sus receptores
7	CCL5	Citoquinas / IL y sus receptores	63	IL12RB2	Citoquinas / IL y sus receptores
8	CCR1	Marcadores de células dendríticas	64	IL13	Citoquinas / IL y sus receptores
9	CCR5	Marcadores de células dendríticas	65	IL17A	Citoquinas / IL y sus receptores
10	CCR7	Marcadores de fenotipo	66	IL17RA	Citoquinas / IL y sus receptores
11	CD160	Receptores inhibitorios y sus ligandos	67	IL18	Citoquinas / IL y sus receptores
12	CD19	Marcadores de fenotipo	68	IL18R1	Citoquinas / IL y sus receptores
13	CD1A	Marcadores de células dendríticas	69	IL1B	Citoquinas / IL y sus receptores
14	CD2	Moléculas coestimuladoras	70	IL2	Citoquinas / IL y sus receptores
15	CD27	Marcadores de fenotipo	71	IL23A	Citoquinas / IL y sus receptores
16	CD274	Receptores inhibitorios y sus ligandos	72	IL23R	Citoquinas / IL y sus receptores
17	CD28	Marcadores de fenotipo	73	IL25	Citoquinas / IL y sus receptores
18	CD3E	Marcadores de fenotipo	74	IL27	Citoquinas / IL y sus receptores
19	CD4	Marcadores de fenotipo	75	IL2RA	Citoquinas / IL y sus receptores
20	CD40	Moléculas coestimuladoras	76	IL2RG	Citoquinas / IL y sus receptores
21	CD40LG	Moléculas coestimuladoras	77	IL4	Citoquinas / IL y sus receptores
22	CD48	Moléculas coestimuladoras	78	IL4R	Citoquinas / IL y sus receptores
23	CD58	Moléculas coestimuladoras	79	IL5	Citoquinas / IL y sus receptores
24	CD69	Moléculas coestimuladoras	80	IL5RA	Citoquinas / IL y sus receptores
25	CD80	Moléculas coestimuladoras	81	IL6	Citoquinas / IL y sus receptores
26	CD83	Marcadores de células dendríticas	82	IL6R	Citoquinas / IL y sus receptores
27	CD86	Moléculas coestimuladoras	83	IL7	Citoquinas / IL y sus receptores
28	CD8A	Marcadores de fenotipo	84	IL7R	Citoquinas / IL y sus receptores
29	CLEC9A	Marcadores de células dendríticas	85	ITGA4	Marcadores de fenotipo
30	CR2	Marcadores de fenotipo	86	ITGAL	Moléculas de adhesión
31	CSF1	Citoquinas / IL y sus receptores	87	ITGAX	Moléculas de adhesión
32	CSF2	Citoquinas / IL y sus receptores	88	ITGB2	Moléculas de adhesión
33	CTLA4	Receptores inhibitorios y sus ligandos	89	KLRG1	Marcadores de fenotipo
34	CXCR3	Citoquinas / IL y sus receptores	90	LAG3	Receptores inhibitorios y sus ligandos
35	FAS	Apoptosis / Senescencia	91	LGALS9	Receptores inhibitorios y sus ligandos
36	FASLG	Apoptosis / Senescencia	92	MKI67	Otras moléculas
37	FCER1A	Marcadores de fenotipo	93	NCAM1	Marcadores de fenotipo
38	FCER2	Marcadores de fenotipo	94	NOS2	Otras moléculas
39	FOXP3	Marcadores de fenotipo	95	PDCD1	Receptores inhibitorios y sus ligandos
40	GAPDH	Genes constitutivos	96	PDCD1LG2	Receptores inhibitorios y sus ligandos
41	GATA3	Factores de Transcripción	97	PGK1	Genes constitutivos
42	GZMA	Moléculas citotóxicas	98	PRF1	Moléculas citotóxicas
43	GZMB	Moléculas citotóxicas	99	SELL	Marcadores de fenotipo
44	GZMH	Moléculas citotóxicas	100	STAT1	Factores de Transcripción
45	GZMK	Moléculas citotóxicas	101	STAT3	Factores de Transcripción
46	GZMM	Moléculas citotóxicas	102	TBP	Genes constitutivos
47	HAVCR2	Receptores inhibitorios y sus ligandos	103	TBX21	Factores de Transcripción
48	HPRT1	Genes constitutivos	104	TGFB1	Citoquinas / IL y sus receptores
49	ICAM1	Moléculas de adhesión	105	TGFB2	Citoquinas / IL y sus receptores
50	ICOS	Moléculas coestimuladoras	106	TGFBR1	Citoquinas / IL y sus receptores
51	ICOSLG	Moléculas coestimuladoras	107	TGFBR2	Citoquinas / IL y sus receptores
52	IDO1	Otras moléculas	108	TNF	Citoquinas / IL y sus receptores
53	IFNAR1	Citoquinas / IL y sus receptores	109	TNFRSF14	Receptores inhibitorios y sus ligandos
54	IFNG	Citoquinas / IL y sus receptores	110	TNFRSF1A	Citoquinas / IL y sus receptores
55	IFNGR1	Citoquinas / IL y sus receptores	111	TNFSF10	Apoptosis / Senescencia
56	IFNGR2	Citoquinas / IL y sus receptores	112	XCR1	Citoquinas / IL y sus receptores

Tras la selección de las sondas, se diseñaron las placas (*Custom TaqMan® OpenArray® Real-Time PCR plates*) para el sistema *QuantStudio™ 12K Flex Real-Time PCR* (Life Technologies). Cada placa incluía los 112 ensayos de expresión génica, organizados en 48 subarrays, cada uno de ellos con 64 nanopocillos, permitiendo realizar en cada placa 3,072 reacciones de amplificación (**Figura M2**).



**Figura M2.** Diseño del ‘*TaqMan® OpenArray® Real-Time PCR plate*’. En la ampliación pueden visualizarse 2 subarrays (con 64 nanopocillos cada uno) en los que se distribuyen las 112 sondas *TaqMan®* correspondientes a los 112 genes seleccionados. Cada nanopocillo posee una capacidad de 33 nl). Los espacios en gris representan nanopocillos sin cebador (controles). Cada color corresponde con una muestra, pudiendo ensayarse hasta 24 muestras por *plate*. Dado que cada paciente se ensaya por triplicado y se incluye un control de placa, cada *plate* puede incluir muestras de hasta 7 pacientes y el control de placa.

### 5.6. PCR cuantitativa en tiempo real de alto rendimiento (RT-qPCR)

Las reacciones de RT-qPCR se llevaron a cabo empleando el equipo *QuantStudio™ 12K Flex Real-Time PCR* (Applied Biosystems). Las muestras de ADNc se cargaron en los ‘*Custom TaqMan® OpenArray® Real-Time PCR plates*’. A cada subarray de la placa se añadió 2,5 µl de *2X OpenArray® Real-Time Master Mix* (Applied Biosystems), 1,3 µl de agua tratada con DEPC y 1,2 µl de ADNc. Cada una de las reacciones se preparó en una placa de 386 pocillos, correspondiendo cada pocillo de dicha placa con un subarray en cuyos 64 nanopocillos se cargaron 33 nl

usando el instrumento *OpenArray® AccuFill™* (Applied Biosystems) de acuerdo con los protocolos del fabricante. Las reacciones de RT-qPCR y la detección de la fluorescencia se llevaron a cabo mediante el sistema *QuantStudio™ 12K Flex Real-Time PCR* (Applied Biosystems) tras la configuración previa del equipo siguiendo las instrucciones del fabricante. Cada muestra se ensayó por triplicado. En cada placa, se incluyó una muestra control, necesaria para el posterior intercalibrado entre placas de los valores de expresión obtenidos. Se obtuvieron dos muestras control, que procedían de 3 sujetos sanos, con un valor de RIN promedio de 9,9, en ambos casos. La idoneidad de las mismas se evaluó tras el ensayo del primer *plate* demostrándose que ambas eran buenas candidatas a control detectándose amplificación en la mayoría de los genes a estudio. No obstante, se seleccionó para incluir en todas las placas aquella que presentó menor número de genes sin valor numérico de amplificación (*Undetermined*).

## 5.7. Procesamiento de los datos de RT-qPCR

### 5.7.1. Limpieza de réplicas, cálculo de promedios e intercalibrado entre placas

Los resultados del equipo *QuantStudio™ 12K Flex Real-Time PCR* se exportaron para cada placa como archivos Excel. Este archivo contenía los valores de amplificación (Cq) para cada uno de los 112 genes, por triplicado, para cada muestra ensayada. El primer paso fue la limpieza de réplicas mediante la ejecución de macros en *Visual Basic* en Excel. Esta macro hizo posible la automatización del proceso, estableciéndose los siguientes criterios para el análisis (**Figura M3**):

- Eliminación de réplicas sin valor numérico (*Undetermined*).
- Conservación de los valores de Cq de las tres réplicas si no diferían entre sí en más de 0,7 unidades.
- Si las diferencias entre 2 réplicas eran inferiores a 0,7, pero mayores con la réplica restante, eran aceptados los valores de las 2 réplicas similares y rechazado el de la tercera.
- Si los valores de todas las réplicas diferían entre sí en más de 0,7 unidades, pero entre el valor mínimo y el intermedio, o el intermedio y el máximo, las diferencias no superaban la unidad, se aceptaba el valor intermedio y los otros dos eran descartados.
- En caso de que los valores de Cq de todas las réplicas variasen entre sí en más de 1 unidad, se escogía únicamente aquella que presentase el valor más similar al promedio de los valores de Cq del grupo biológico al que pertenecía dicha muestra.
- Eliminación de los valores Cq procedentes de una única réplica.

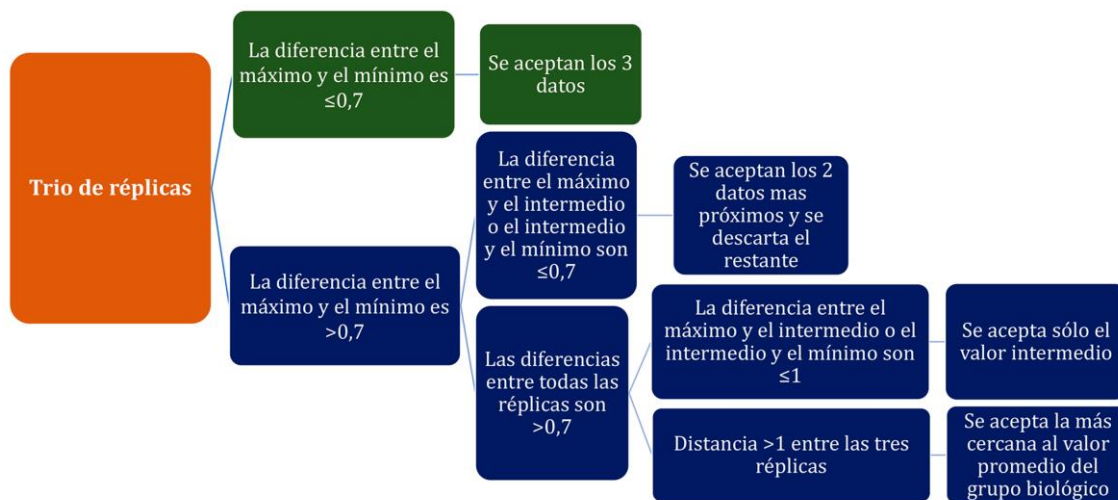


Figura M3. Representación esquemática de los criterios establecidos en el algoritmo empleado para la limpieza de réplicas en *Visual Basic*.

Una vez realizada la limpieza de réplicas, se hallaron los valores promedios obteniéndose un solo valor de Cq por muestra y gen analizado. Con objeto de evitar diferencias en los valores de expresión entre placas, se usó el programa GenEx v6 para aplicar un intercalibrado entre todas las placas ensayadas (*interplate-calibration*) tomando los valores de expresión obtenidos para la muestra control incluida en todas ellas.

### 5.7.2. Selección de genes de normalización. Normalización de datos de RT-qPCR y aplicación de cantidades relativas

#### 5.7.2.1. Selección de genes de normalización

Para normalizar los valores de expresión obtenidos tras el intercalibrado se identificaron los genes candidatos a genes normalizadores aplicando los algoritmos específicos geNorm, NormFinder y RefFinder.

#### - Normalización con geNorm

El algoritmo geNorm evalúa la estabilidad en la expresión de los genes seleccionados y valora la conveniencia de emplear múltiples genes, definiendo el número idóneo de genes para la normalización (Vandesompele et al., 2002). La estabilidad está representada por el valor M, este valor se define como la variación media por pares de un gen específico frente al resto de genes del ensayo candidatos a normalizadores, siendo, por tanto, menor cuanto mayor estabilidad en la expresión presente el gen evaluado. La aplicación de este algoritmo se realizó

usando el programa GenEx v6 (MultiD Analyses), el cual establece un valor M máximo de 0,5 para la selección de los genes candidatos.

### - Normalización con NormFinder

NormFinder evalúa la variación general en la expresión de cada gen candidato a gen de referencia. Este algoritmo estima la estabilidad de un gen en particular, considerando tanto las variaciones en la expresión entre los subgrupos de las muestras (variación inter-grupo), como su variación dentro de los grupos (variación intra-grupo). NormFinder también proporciona información sobre el número óptimo de genes de referencia para llevar a cabo la normalización. Tras aplicar el algoritmo, los genes se ordenan en función a su estabilidad, así los genes con una variación inter e intra-grupo mínima se consideran los más estables (Andersen et al., 2004). Para aplicar este algoritmo, también se empleó el programa GenEx v6.

### - Normalización con RefFinder

Esta herramienta online compara e integra los resultados de cuatro algoritmos diferentes (geNorm, NormFinder, BestKeeper (Pfaffl et al., 2004) y el método comparativo  $\Delta$ Ct (Silver et al., 2006) para clasificar los genes candidatos a genes normalizadores. RefFinder (Xie et al., 2012) evalúa cada gen individual basándose en la media geométrica de los cuatro métodos. Se encuentra disponible en: <https://www.heartcure.com.au/reffinder/>.

#### **5.7.2.2. Normalización de datos de RT-qPCR y cantidades relativas**

Una vez seleccionados los mejores candidatos a genes de referencia según los resultados de los diferentes algoritmos, todo el conjunto de datos se normalizó con dichos genes y, posteriormente, se aplicaron cantidades relativas, obteniéndose tras este procesamiento en GenEx v6, valores de expresión génica relativa normalizados (NRQ).

#### **5.7.3. Detección de outliers. Criterios de exclusión**

Dentro de cada grupo biológico se identificaron valores atípicos denominados *outliers*. Para ello, se aplicó un módulo en *Visual Basic* de Excel cuyo código se basaba en los cuartiles y el rango intercuartílico (RI).

Los cuartiles se definen como tres valores que dividen un conjunto de datos ordenados en cuatro porciones iguales:

- 1<sup>er</sup> cuartil (Q1): el 25% de los datos es menor o igual a este valor.
- 2<sup>do</sup> cuartil (Q2): el 50% de los datos es menor o igual a este valor (mediana).
- 3<sup>er</sup> cuartil (Q3): el 75% de los datos es menor o igual a este valor.

Y, el rango intercuartílico, definido como la distancia entre el 1<sup>er</sup> cuartil y el 3<sup>er</sup> cuartil (Q3-Q1), abarcando el 50% central de los datos.

Según los criterios establecidos en este módulo, se consideraron *outliers* aquellos valores que superasen el Q1 menos 1,5 veces el RI, o el Q3 más 1,5 veces el RI. Estos valores atípicos detectados se excluyeron del análisis de gráficos de volcán.

### 5.8. Análisis estadístico

Los análisis estadísticos se llevaron a cabo usando los programas SPSS Statistics v25 (SPSS Inc., Chicago, Illinois, Estados Unidos) y GraphPad Prism v8.0 (GraphPad Software, San Diego, CA).

La significancia de las diferencias observadas en la expresión génica entre los grupos comparados se determinó empleando el test paramétrico T de Student o el test no paramétrico de Mann-Whitney, atendiendo a si cumplían o no la condición de normalidad, la cual se evaluó previamente aplicando los test Shapiro-Wilk y Kolmogorov-Smirnov ( $\alpha=0,05$ ). En SPSS v25, se emplearon análisis de componentes principales (PCA) sobre los valores NRQ para determinar la estructura del conjunto de datos. Las diferencias en las puntuaciones de las componentes principales representadas entre los grupos se confirmaron mediante la prueba de Mann-Whitney o prueba T de Student, según fuese apropiado.

Se estableció que las diferencias observadas eran estadísticamente significativas cuando adquirirían un valor  $p$  inferior a 0,05 ( $p < 0,05$ ).

### 5.9. Obtención de valores de expresión génica diferencial y representación

Para la obtención de los valores de expresión génica diferencial, en cada grupo biológico se halló el *fold change* (FC) definido como la ratio entre el promedio de los valores de expresión génica relativa normalizada para cada gen entre los grupos a comparar (grupo de estudio / grupo control). A este valor se le aplicó el logaritmo en base 2 ( $\text{Log}_2$  FC) obteniéndose valores positivos, correspondientes a genes con mayor expresión en el grupo de estudio (genes sobreexpresados), o

negativos, genes con menor expresión en este grupo respecto al grupo control (genes infraexpresados). Se estableció un punto de corte de  $\log_2 FC > 0,6$ ;  $\log_2 FC < -0,6$ , que correspondía a un FC de 1,5 o 0,66 y otro punto de corte más estricto de  $\log_2 FC > 1$ ;  $\log_2 FC < -1$ , correspondientes a un FC de 2 o 0,5, respectivamente.

### **5.9.1. Representación en gráfico de volcán**

Un gráfico de volcán es una representación gráfica de la relación que existe entre la diferencia en los valores de expresión génica entre las muestras de dos grupos y su significancia estadística. En el eje de abscisas (eje X) se representa el  $\log_2 FC$  (promedio grupo 1 / promedio grupo 2) y en el eje de ordenadas (eje Y) el  $-\log_{10}$  p-valor para cada gen analizado. De esta forma, cuanto mayor sean las diferencias de expresión génica en los grupos comparados, más alejados del 0 se encontrarán los puntos en el eje de la X (mostrándose con valores negativos los genes infraexpresados (azul) y con valores positivos los genes sobreexpresados (rojo) en el grupo de estudio respecto al grupo control). En el eje Y, mientras mayor sea el valor de  $-\log_{10}$  p-valor, mayor es la significancia estadística.

En los gráficos se marcó el valor 0 en el eje de abscisas, indicando que los genes situados en dicha posición no varían su expresión entre los grupos comparados. Del mismo modo, se trazó una línea horizontal en  $-\log_{10} p = 1,3$  y  $3$  correspondiente a  $p = 0,05$  y  $p = 0,001$ , respectivamente. Los valores superiores a  $1,3$  correspondían a genes cuya expresión génica presentaba diferencias estadísticamente significativas entre el grupo biológico de estudio y el grupo control. En determinados casos en los que un grupo a comparar no disponía de valores de expresión para un gen determinado, para poder realizar el cálculo de la ratio y obtener el FC se sustituyó el valor 0 por  $10^{-6}$ . Estos genes poseen valores de  $\log_2 FC$  extremos, indicando cuando son positivos la no expresión de dicho gen en el grupo control o cuando son negativos que no se expresan en el grupo de estudio.

### **5.9.2. Representación en gráfico de barras**

Mediante gráficos de barras se representaron los valores de expresión génica diferencial entre los grupos comparados. Cada barra correspondía a un gen expresado diferencialmente y cuyas diferencias, a su vez, eran estadísticamente significativas, siendo la altura de la barra equivalente al valor de  $\log_2 FC$  obtenido para dicho gen. Estos gráficos permitieron la visualización rápida y comparativa de los valores obtenidos para los diferentes genes analizados.

### 5.10. Predicción de interacciones funcionales proteína-proteína

La herramienta STRING (Szklarczyk et al., 2019) disponible en (<https://string-db.org>) se empleó para la predicción de las interacciones funcionales entre las proteínas codificadas por los genes con expresión diferencial entre los grupos de estudio (individuos IND *versus* DS). Para la generación de las redes, la especie se limitó a *Homo sapiens* y a una puntuación mínima de la interacción de 0,9 (definida como máxima confianza). Como fuentes de predicción de interacción se aplicaron fusión génica, bases de datos, experimentos, coocurrencia, vecindad y coexpresión.

### 5.11. Análisis de enriquecimiento de conjuntos de genes (GSEA)

El conjunto de genes enriquecidos en el grupo de pacientes IND *versus* DS se determinó empleando el software GSEA (*Gene Set Enrichment Analysis*) v4.1.0 (Mootha et al., 2003; Subramanian et al., 2005). Este método determina, a partir de los valores de expresión génica, el conjunto de genes enriquecidos en un grupo biológico respecto al grupo control, indicando a su vez las rutas biológicas en los que participan dichos genes. Para este análisis se usaron las rutas incluidas en la base de datos BioCarta (*Canonical Pathways gene sets derived from the BioCarta pathway database*) incluida en "C2: curated gene sets collection" de la base de datos MSigDB (*Molecular Signatures Database*) (Liberzon et al., 2011, 2015; Subramanian et al., 2005). Los parámetros establecidos en GSEA fueron los siguientes: permutaciones: 100.000; tipo de permutación: fenotipo ( $n^{\circ}$  muestras > 7); estadística de enriquecimiento: ponderado; métrica de clasificación: Prueba T de Student; tamaño máximo: 500; tamaño mínimo: 3.



## **RESULTADOS**

---



## A. BIOMARCADORES GENÉTICOS

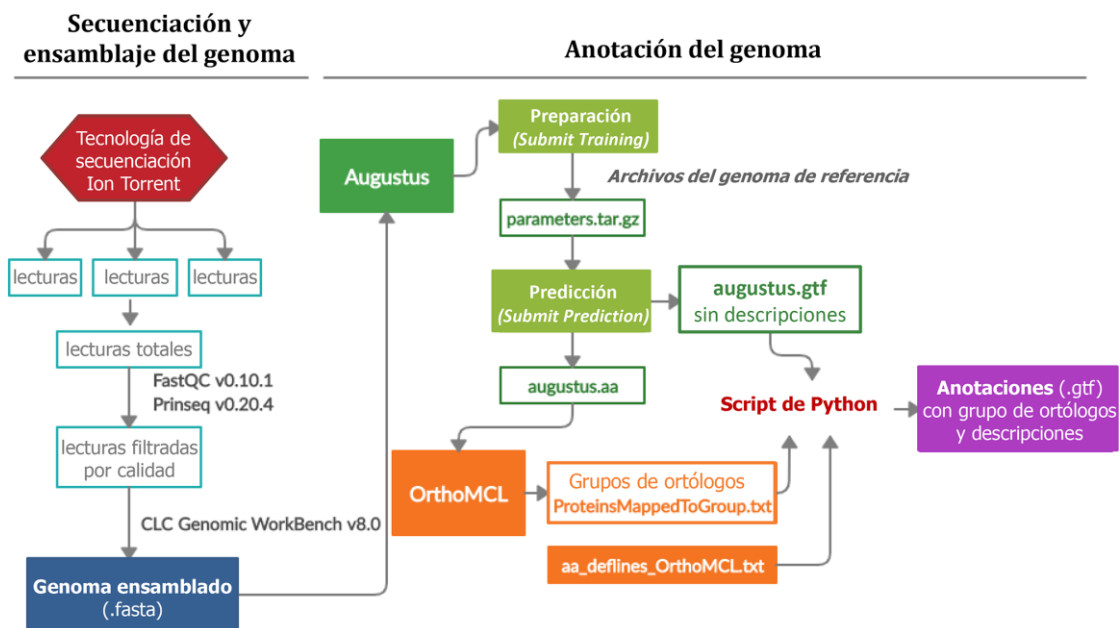
### A.1. SECUENCIACIÓN, ENSAMBLAJE Y ANOTACIÓN DE LOS GENOMAS DE DIFERENTES CEPAS DE *T. cruzi*. ANÁLISIS COMPARATIVO DEL CONTENIDO GÉNICO Y DE RETROELEMENTOS PORTADORES DE LA HUELLA Pr77

Con objeto de identificar la existencia de diferencias en el contenido genómico en el genoma de *T. cruzi*, se secuenciaron en el laboratorio los genomas de diferentes cepas de *T. cruzi* pertenecientes a diferentes DTUs empleando la tecnología Ion Torrent. Las cepas seleccionadas fueron B. M. López (DTU Ia), Dm28 (DTU Id), Y (DTU II), Ikiakarora (DTU III), SOL (DTU V) y CL Brener (DTU VI). En la **Tabla R1** se indica el DTU al que pertenecen, su origen geográfico y el hospedador/vector del que se aisló cada una de ellas. Las secuencias resultantes se ensamblaron en *contigs* y anotaron para, posteriormente, llevar a cabo estudios genómicos comparativos.

**Tabla R1. DTU, origen geográfico y hospedador/vector de las cepas de *T. cruzi* analizadas en este estudio** (Hamuy et al., 2013; Higuera et al., 2013; Murcia et al., 2013; Rodríguez et al., 1998). Adaptada de Gómez, López, et al., 2021.

Cepa	DTU	Localización geográfica	Hospedador/Vector
<b>B. M. López</b>	Ia	Paratebueno, Cundinamarca (Colombia)	<i>Homo sapiens</i>
<b>Dm28</b>	Id	Carabobo (Venezuela)	<i>Didelphis marsupialis</i>
<b>Y</b>	II	Sao Paulo (Brasil)	<i>Homo sapiens</i>
<b>Ikiakarora</b>	III	Catatumbo, Norte Santander (Colombia)	<i>Rhodnius prolixus</i>
<b>SOL</b>	V	Murcia (España)	<i>Homo sapiens</i>
<b>CL Brener</b>	VI	Encruzilhada, Rio Grande do Sul (Brasil)	<i>Triatoma infestans</i>

El ensamblaje se realizó empleando el programa CLC Genomic WorkBench v8.0 y la anotación mediante el uso de las herramientas Augustus y OrthoMCL disponibles en (<http://bioinf.uni-greifswald.de/webaugustus/> y <https://veupathdb.globusgenomics.org/>, respectivamente) siguiendo el flujo de trabajo que se muestra en la **Figura R1** (más detalles en la sección de *Métodos 1.5 y 1.7*).



**Figura R1. Esquema que describe los procesos de secuenciación, ensamblaje y anotación llevados a cabo en los genomas de las diferentes cepas de *T. cruzi* descritas en este estudio. Adaptada de Gómez, López, et al., 2021.**

En la **Tabla R2** se muestra el resumen de las métricas de secuenciación y ensamblaje. Se obtuvieron un total de  $5,4 \times 10^6$ ,  $3,4 \times 10^6$ ,  $2,3 \times 10^6$ ,  $3,9 \times 10^6$ ,  $3,3 \times 10^6$  y  $3,5 \times 10^6$  lecturas crudas de las librerías genéticas derivadas de las cepas B. M. López, Dm28, Y, Ikiakarora, SOL y CL Brener, respectivamente. Después de filtrar por calidad de secuencia, las lecturas restantes ( $4,6 \times 10^6$  (B. M. López),  $2,9 \times 10^6$  (Dm28),  $2 \times 10^6$  (Y),  $3,3 \times 10^6$  (Ikiakarora),  $2,9 \times 10^6$  (SOL) y  $3 \times 10^6$  (CL Brener) se ensamblaron en 5.923 *contigs* (B. M. López), 6.541 (Dm28), 6.942 (Y), 11.096 (Ikiakarora), 11.946 (SOL) y 11.101 (CL Brener), presentando el *contig* más largo un tamaño de 45.876 (B. M. López), 33.325 (Dm28), 17.707 (Y), 33.607 (Ikiakarora), 26.881 (SOL) y 23.056 (CL Brener) pares de bases (pb). Los tamaños promedio de los *contigs* (en pb) fueron de 3.124 para B. M. López, 2.633 para Dm28, 2.239 para Y, 1.666 para Ikiakarora, 1.679 para SOL y 1.759 para CL Brener. Tras el ensamblaje, los genomas resultantes mostraron tener un tamaño de 18,5 Mb (B. M. López), 17,2 Mb (Dm28), 15,5 Mb (Y), 18,5 Mb (Ikiakarora), 20,1 Mb (SOL) y 19,5 Mb (CL Brener). Se calcularon los valores de N50 (en pb) para cada cepa que resultaron 5.125 (B. M. López), 3.659 (Dm28), 2.891 (Y), 2.193 (Ikiakarora), 2.171 (SOL) y 2.298 (CL Brener).

Tabla R2. Resumen de métricas de secuenciación y ensamblaje de genomas de las cepas B. M. López, Dm28, Y, Ikiakarora, SOL y CL Brener de *T. cruzi*. Adaptada de Gómez, López, et al., 2021.

Cepa	Secuenciación		Ensamblaje					Bases totales
	Lecturas crudas	Lecturas filtradas <sup>a</sup>	contigs	N50 (pb)	Menor tamaño <sup>b</sup>	Promedio	Mayor tamaño	
								contig (pb)
B. M. López	5415819	4591877	5923	5125	500	3124	45876	18508455
Dm28	3354685	2947182	6541	3659	500	2633	33325	17227559
Y	2345376	1972462	6942	2891	500	2239	17707	15548466
Ikiakarora	3928712	3338764	11096	2193	500	1666	33607	18492845
SOL	3332449	2888011	11946	2171	500	1679	26881	20061745
CL Brener	3468705	2955230	11101	2298	500	1759	23056	19533022

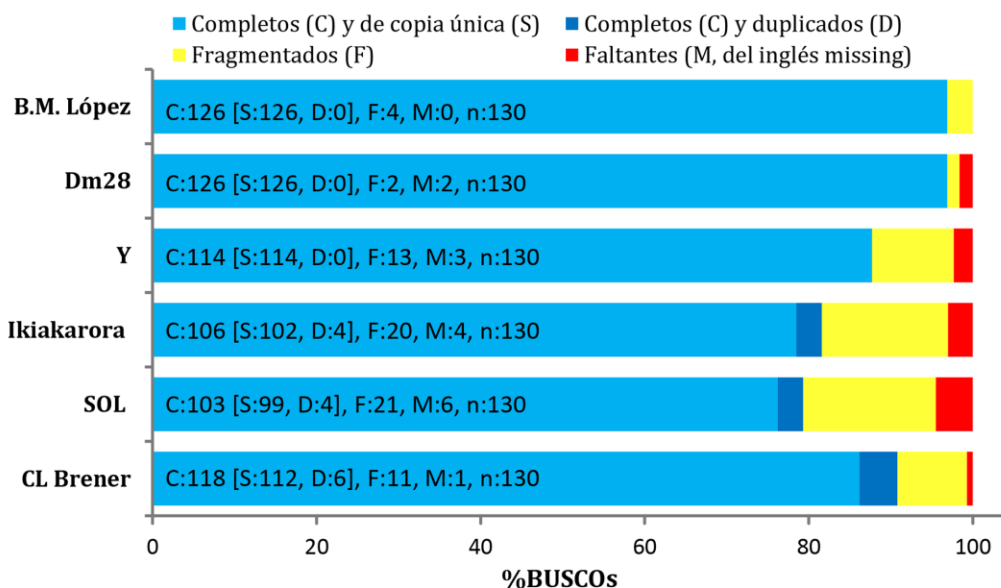
a: Número de lecturas tras el análisis de calidad con FastQC y filtrado de calidad empleando la herramienta Prinseq.

b: El tamaño mínimo de *contig* se estableció en 500 pb durante el proceso de ensamblaje.

### A.1.1. Análisis de la integridad de los ensamblajes

La integridad de los ensamblajes obtenidos para los seis genomas de *T. cruzi* se evaluó mediante el análisis comparativo de ortólogos universales de copia única (BUSCO, *Benchmarking Universal Single-Copy Orthologs*) para el conjunto de datos de Euglenozoa como predeterminado, el cual contiene 130 proteínas BUSCO que se espera estén presentes en el genoma como genes de copia única. Los conjuntos de genes de los genomas ensamblados se clasificaron de acuerdo con los parámetros de análisis de BUSCO (**Figura R2**) como genes completos (C), de copia única (S), duplicados (D), fragmentados (F) o faltantes (M). Los resultados obtenidos, representados gráficamente en la **Figura R2**, reflejaron la buena integridad de los genomas ensamblados al indicar que el porcentaje de genes BUSCO completos y de copia única identificados fue del 96,9% (B. M. López), 96,9% (Dm28), 87,7% (Y), 78,5% (Ikiakarora), 76,2% (SOL) y 86,2% (CL Brener) de los 130 grupos de búsqueda de BUSCO (**Figura R2**).

### Análisis BUSCO



**Figura R2. Análisis BUSCO de la integridad del ensamblaje del genoma.** Para la evaluación del ensamblaje del genoma se empleó el conjunto de datos BUSCO (*Benchmarking set of Universal Single-Copy Orthologues*) de Euglenozoa odb\_10, que incluye 130 BUSCO. El gráfico de barras muestra para cada genoma: % de genes BUSCO completos y de copia única (azul claro), completos y duplicados (azul oscuro), fragmentados (amarillo) y faltantes (rojo) en los ensamblajes. *Adaptada de* Gómez, López, et al., 2021.

#### A.1.2. Estimación del tamaño de los genomas

Para estimar el tamaño del genoma de las cepas a estudio, se utilizó el número de bases totales identificadas en cada genoma ensamblado (**Tabla R2**) y el porcentaje de genes BUSCO completos y de copia única (**Figura R2**) teniendo en cuenta el tamaño del genoma publicado de la cepa CL Brener como genoma de referencia (55 Mb (El-Sayed, Myler, Blandin, et al., 2005)). Estos cálculos, descritos en la **Tabla R3**, llevaron a una estimación del tamaño del genoma para cada cepa de 49,7 Mb para B. M. López, 46,3 Mb para Dm28, 43,5 Mb para Y, 54,2 Mb para Ikiakarora y 59,5 Mb para SOL (**Tabla R3**).

**Tabla R3. Estimación del tamaño del genoma de las cepas de *T. cruzi*.** La estimación del tamaño de cada genoma se calculó en base al número total de bases en cada genoma ensamblado y el porcentaje de genes BUSCO identificados y, posteriormente, se referenció al tamaño del genoma de CL Brener como genoma de referencia. *Adaptada de Gómez, López, et al., 2021.*

Cepa	pb totales del genoma ensamblado	Estimación de pb totales <sup>a</sup>	Estimación de pb perdidas durante el ensamblaje <sup>b</sup>	Tamaño estimado del genoma <sup>c</sup>
CL Brener	19.533.022	22.660.118 bp	32.339.882 (55Mb-22.660.118)	55 Mb*
B. M. López	18.508.455	19.100.572 bp	30.643.555	49,7 Mb
Dm28	17.227.559	17.778.698 bp	28.522.838	46,3 Mb
Y	15.548.466	17.729.151 bp	25.742.844	43,5 Mb
Ikiakarora	18.492.845	23.557.764 bp	30.617.711	54,2 Mb
SOL	20.061.745	26.327.749 bp	33.215.263	59,5 Mb

\*Tamaño del genoma de CL Brener según (El-Sayed, Myler, Blandin, et al., 2005).

a: Se obtiene de la relación de pb totales identificadas en los genomas ensamblados entre el porcentaje de genes BUSCO completos y de copia única de cada genoma multiplicado por 100: Estimación del número de bases totales =  $\frac{\text{N}^\circ \text{ bases totales}}{\% \text{ BUSCO}} \times 100$

b: El n° de bases perdidas durante el ensamblaje en CL Brener se calculó restando el n° de bases totales estimadas en el genoma de CL Brener al tamaño publicado del genoma de CL Brener (55 Mb). La estimación del número de bases perdidas en cada genoma se calculó en función de las bases totales de cada genoma, bases perdidas en CL Brener y bases totales del genoma ensamblado de CL Brener siguiendo la fórmula:

$$\text{N}^\circ \text{ de bases perdidas} = \frac{\text{N}^\circ \text{ bases totales} \times 32.339.882 \text{ (bases perdidas en CL Brener)}}{19.533.022 \text{ (pb totales CL Brener)}}$$

### A.1.3. Anotación de los genomas

La anotación de los genes se llevó a cabo para cada genoma utilizando la herramienta de predicción de genes AUGUSTUS y el flujo de trabajo OrthoMCL para asignar el conjunto de proteínas predichas a los grupos OrthoMCL como se indica en la **Figura R1**. En primer lugar, para la predicción de genes mediante AUGUSTUS fue necesario establecer un genoma de referencia para cada una de las cepas cuyo genoma se deseaba anotar. Para ello, se realizó un análisis de *Best Hit* (BH) frente a una base de datos de proteínas constituida por las proteínas descargadas de diferentes genomas de *T. cruzi* disponibles en la base de datos de kinetoplástidos TriTrypDB. El genoma de referencia a usar para cada uno de los seis genomas secuenciados se seleccionó en base a la puntuación de identidad más alta obtenida en cada caso (**Tabla R4**). Así, se seleccionó Dm28c para la anotación de los genomas B. M. López y Dm28 con 3.402 y 4.329 BH, CL Brener Esmeraldo-like se usó para la

## RESULTADOS

anotación del genoma Y e Ikiakarora dado que CL Brener Esmeraldo-like fue la cepa con la que, Y e Ikiakarora mantuvieron la identidad más alta con 6.892 y 5.485 BH, respectivamente. Sobre esta base, se utilizó CL Brener non-Esmeraldo-like para la anotación de los genomas SOL y CL Brener con 6.120 y 5.717 BH, respectivamente.

**Tabla R4. Resumen del análisis de *Best Hit*.** La tabla muestra el número de proteínas de los seis genomas secuenciados con *Best Hit* para cada genoma de referencia. Los genomas de referencia utilizados fueron CL Brener, CL Brener non-Esmeraldo-like (CL Brener NEL), CL Brener Esmeraldo-like (CL Brener EL), Dm28c, SylvioX10-1 y la subespecie Marinkellei B7. *Adaptada de* Gómez, López, et al., 2021.

Cepa	Cepas de referencia					
	CL Brener	CL Brener NEL	CL Brener EL	Dm28c	Sylvio X10-1	Marinkellei B7
B. M. López	62	922	474	3.402	4.266	62
Dm28	52	795	391	4.329	3.753	65
Y	108	1.859	6.892	572	704	148
Ikiakarora	139	4.427	5.485	685	972	111
SOL	191	6.120	5.078	852	1.212	158
CL Brener	214	5.717	4.983	699	968	122

### A.1.4. Determinación de las características de los genomas

Tras el ensamblaje y anotación, se determinaron parámetros referentes a los genomas obtenidos. Así, el contenido de G+C de los genomas fue del 48,3%, 48,6%, 49,8%, 48,7%, 49,5% y 49,5% para las cepas B. M. López, Dm28, Y, Ikiakarora, SOL y CL Brener, respectivamente (**Tabla R5**). El número de genes identificados y anotados en cada genoma fue de 7.661 (B. M. López), 7.720 (Dm28), 8.297 (Y), 11.358 (Ikiakarora), 12.829 (SOL) y 11.909 (CL Brener), lo que representa el 62,9% (B. M. López), 64,1% (Dm28), 77,7% (Y), 73,2% (Ikiakarora), 72,6% (SOL) y 73,3% (CL Brener) del total de pares de bases de cada genoma ensamblado. El tamaño medio de los genes fue de 1.518 (B. M. López), 1.429 (Dm28), 1.456 (Y), 1.191 (Ikiakarora), 1.135 (SOL) y 1.201 (CL Brener) pb. Del número total de genes anotados, pudo asignarse información a 5.036 (B. M. López), 5.145 (Dm28), 5.590 (Y), 7.739 (Ikiakarora), 8.600 (SOL) y 8.019 (CL Brener) genes basada en la identidad de su secuencia con proteínas homólogas descritas en otros organismos.



Los genes restantes se anotaron como genes codificantes de proteínas hipotéticas (Tabla R5).

**Tabla R5. Características de los genomas ensamblados y genes anotados en las cepas de *T. cruzi* a estudio.** Adaptada de Gómez, López, et al., 2021.

Cepa	%GC <sup>a</sup>	Genes	Tamaño medio de los genes <sup>b</sup> (pb)	% genoma genes anotados <sup>c</sup>	Homología con proteínas de función conocida <sup>d</sup>	Proteínas hipotéticas <sup>e</sup>
B. M. López	48,26	7.661	1.518	62,86	5.036	2.625
Dm28	48,59	7.720	1.429	64,05	5.145	2.575
Y	49,79	8.297	1.456	77,74	5.590	2.707
Ikiakarora	48,72	11.358	1.191	73,20	7.739	3.619
SOL	49,48	12.829	1.135	72,63	8.600	4.229
CL Brener	49,53	11.909	1.201	73,26	8.019	3.890

a: Contenido en guanina-citosina.

b: Suma de la longitud de todos los genes identificados dividido entre el total de genes identificados.

c: Porcentaje del genoma ocupado por los genes anotados.

d: Número de genes que poseen anotación basado en la identidad con proteínas de función conocida.

e: Número de genes con función desconocida (codifican proteínas hipotéticas).

#### A.1.5. Análisis del contenido en retroelementos

Con el fin de identificar los retrotransposones L1Tc y NARTc, que portan la huella Pr77, en las lecturas de los seis genomas secuenciados, se realizaron dos búsquedas para identificar los extremos 5' y 3' de los elementos L1Tc y NARTc. En primer lugar, se usó una sonda del extremo 5', la cual correspondía a la huella Pr77 presente en ambos retroelementos. Seguidamente, se emplearon las secuencias de 40 nt ubicadas corriente abajo de Pr77 para identificar la clase de elemento (L1Tc o NARTc), solicitando, al menos, un 70% de identidad para la asignación. En el caso de los extremos 3', se utilizó una sonda correspondiente a los últimos 55 nt de L1Tc o NARTc, para llevar a cabo identificación y clasificación en un sólo paso. Finalmente, los candidatos se validaron mediante búsqueda BLASTN frente una base de datos compuesta por 12 y 7 secuencias diferentes correspondientes a elementos L1Tc o NARTc (secuencias en Anexo 1), respectivamente. Solo se tomaron para el análisis las secuencias que mantuvieron una identidad superior al 70% y con una longitud de alineamiento de, al menos, el 70% de la secuencia del elemento validado. En todas las cepas estudiadas, se encontraron secuencias que contenían la huella Pr77 que no

## RESULTADOS

correspondían ni a L1Tc ni a NARTc, en consecuencia, estas secuencias se clasificaron como huellas Pr77 únicas o *single Pr77 hallmarks* (Pr77Sh) (**Tabla R6**).

Para comparar la frecuencia de las huellas Pr77 identificadas entre las cepas, se calculó un índice Pr77 dividiendo el número de secuencias Pr77 identificadas en cada genoma entre el total de lecturas filtradas obtenidas en cada genoma secuenciado. Como se muestra en la **Tabla R6** (índice Pr77<sup>a</sup>), B. M. López fue la cepa donde se identificó un menor número de huellas Pr77 (índice 9,5) y la cepa Ikiakarora aquella con el mayor número de ellas (índice 14,4). Se calculó, también, el número de huellas Pr77 en cada cepa relativo al detectado en la cepa B. M. López para lo cual se aplicó la siguiente fórmula:

$$\text{Abundancia relativa de huellas Pr77} = \frac{(\text{Índice cepa} - \text{índice B. M. López})}{\text{índice B. M. López}} + 1$$

Los resultados, incluidos en la **Tabla R6**, mostraron un enriquecimiento de 1,4 veces (Dm28), 1,2 (Y), 1,5 (Ikiakarora), 1,3 (SOL) y 1,5 (CL Brener) de las huellas Pr77 en los genomas en relación con el genoma de B. M. López.

**Tabla R6. Comparación del contenido en elementos transponibles.** Adaptada de Gómez, López, et al., 2021.

	B. M. López	Dm28	Y	Ikiakarora	SOL	CL Brener
<b>Índice Pr77<sup>a</sup></b>	9,5	13,3	11,7	14,4	12,5	13,8
<b>Presencia Pr77 respecto a cepa con menor índice<sup>b</sup></b>	1	1,4	1,2	1,5	1,3	1,5
<b>Índice NARTc completos (con Pr77 y extremo 3')<sup>c</sup></b>	0,88	1,11	1,18	1,11	1,06	1,21
<b>Índice Pr77 no asociados a L1Tc o NARTc<sup>d</sup></b>	0,33	0,17	0,42	0,93	0,87	1,17

**a:** Índice obtenido dividiendo el número de secuencias Pr77 identificadas entre el total de lecturas filtradas de cada genoma y multiplicando por 10<sup>4</sup>.

**b:** Abundancia de las secuencias Pr77 en las diferentes cepas relativa a la abundancia de Pr77 en el genoma de B. M. López (cepa con menor índice Pr77<sup>a</sup>).

**c:** Índice calculado dividiendo el número de NARTc de longitud completa entre el total de lecturas filtradas de cada genoma y multiplicando por 10<sup>4</sup>.

**d:** Índice calculado dividiendo el número de Pr77 identificados no asociados a elementos L1Tc o NARTc entre el total de lecturas filtradas de cada genoma y multiplicando por 10<sup>4</sup>.

El análisis del número de elementos NARTc de longitud completa (conteniendo huella Pr77 y extremo 3') en relación con el total de lecturas en cada genoma, mostró un índice de 0,88 (B. M. López), 1,11 (Dm28), 1,18 (Y), 1,11 (Ikiakarora), 1,06 (SOL) y 1,21 (CL Brener) (**Tabla R6, Índice c**). El número de huellas Pr77, del total de lecturas filtradas de cada genoma, que no se asociaron ni con L1Tc ni con NARTc fue de 0,33 (B. M. López), 0,17 (Dm28), 0,42 (Y), 0,93 (Ikiakarora), 0,87 (SOL) y 1,17 (CL Brener) (**Tabla R6, Índice d**).

**A.1.5.1. Proporción de retroelementos L1Tc y NARTc**

La abundancia de los retroelementos L1Tc y NARTc se analizó comparativamente en los seis genomas de *T. cruzi*. Para ello, se consideraron los datos procedentes de las dos búsquedas realizadas con la huella Pr77 (sonda 5') y las sondas 3' correspondientes a L1Tc o NARTc. Los resultados indicaron que todos los genomas contenían un mayor número de copias del retrotransposón L1Tc que de NARTc. Además, en todas las cepas analizadas, la relación de elementos L1Tc/NARTc identificados usando la sonda 3' fue mayor que la detectada cuando se usó la sonda 5' (secuencia Pr77). Los datos indican que en todos los genomas hay, al menos, cuatro veces más copias de elementos L1Tc que de elementos NARTc (**Tabla R7**).

**Tabla R7. Abundancia relativa de retroelementos identificados por (a) huella Pr77 más 40 nt ubicados corriente abajo de Pr77, (b) extremos 3' (55 últimos nt de L1Tc o NARTc) o (c) promedio de elementos identificados tanto por el extremo 5' como 3'. Adaptada de Gómez, López, et al., 2021.**

Proporción	B. M. López	Dm28	Y	Ikiakarora	SOL	CL Brener
L1Tc / NARTc (Identificadas a partir de Pr77) <sup>a</sup>	4	5,02	3,39	4	4,37	4,19
L1Tc / NARTc (Identificadas a partir de 3') <sup>b</sup>	5,64	7,08	5,59	6,73	4,80	6,38
L1Tc / NARTc (Total) <sup>c</sup>	4,88	6,09	4,44	5,27	4,62	5,26

**a:** Secuencia de la huella Pr77 más 40 nt localizados corriente abajo de Pr77.

**b:** Extremo 3' (55 últimos nucleótidos de L1Tc o NARTc).

**c:** Promedio de proporción de elementos identificados a partir de su extremo 5' y 3'.

### A.1.6. Contexto genómico de los retroelementos

Para determinar la ubicación y el contexto genómico en el que se encuentran insertados los retroelementos portadores de la huella Pr77, se extrajeron de las lecturas las secuencias localizadas corriente arriba de las huellas Pr77 y corriente abajo de los extremos 3' de los retroelementos y se alinearon con los *contigs* de los genomas ensamblados y, posteriormente, con los genomas anotados para analizar la identidad de las secuencias.

Los resultados, resumidos en la **Tabla R8**, indicaron que los retroelementos L1Tc y NARTc estaban presentes en todas las cepas, frecuentemente posicionados cerca de genes codificantes de trans-sialidasas (TS), proteínas de superficie asociadas a mucinas (MASP), RHS (*Retrotransposon Hot Spot*) y genes que codifican proteínas hipotéticas (HP). Además, los resultados indicaron que estos genes podrían encontrarse tanto corriente arriba como corriente abajo de los elementos L1Tc, NARTc y huellas Pr77 no asociadas a retroelementos (Pr77Sh), aunque en diferentes proporciones según el tipo de retroelemento, la clase de gen y la cepa del parásito. Estas asociaciones no eran aleatorias dado que se encontraban frecuentemente representadas en todos los genomas analizados, tal y como indica el porcentaje que representa cada asociación (retroelemento y secuencia particular) con respecto al número total de retroelementos portadores de Pr77 identificados en cada genoma (**Tabla R8**).

El análisis reveló que el porcentaje de elementos asociados con genes codificantes de TS varió entre cepas, demostrando una mayor frecuencia en la cepa SOL (40,28%), localizadas principalmente corriente arriba (UP) de las huellas Pr77. No obstante, las secuencias Pr77 se asociaron con mayor frecuencia con genes que codifican proteínas hipotéticas que se localizaban tanto corriente arriba como corriente abajo (DW) de las huellas Pr77 en todas las cepas, oscilando de 407 y 368 en B. M. López a 76 y 113 secuencias en la cepa Y, respectivamente (**Tabla R8**).

El número de retroelementos flanqueados por MASP resultó ser menor que los flanqueados por TS. También se observaron diferencias entre cepas, presentando la cepa CL Brener un porcentaje elevado de elementos flanqueados por estas proteínas (7,58% UP y 0,25% DW) y la cepa Y el porcentaje más bajo (0,76% UP), en la que no se localizaron MASP corriente abajo de los retroelementos

identificados. El número de retroelementos flanqueados por genes codificantes de RHS fue mayor en las cepas SOL e Ikiakarora, representadas por un total de 22,34% y 16,53%, respectivamente. Este porcentaje fue de 5,7%, 2,5% y 0,24% en las cepas CL Brener, Y y B. M. López. No se identificaron retroelementos flanqueados por RHS en Dm28 y, en la cepa B. M. López, solo se identificaron RHS corriente arriba de los retroelementos (**Tabla R8**). Los porcentajes correspondientes a elementos flanqueados por proteínas hipotéticas oscilaron de 28,9% a 60,26% (UP) y del 34,59% a 67,66% (DW), según la cepa. La cepa en la que mayores porcentajes se alcanzaron fue Dm28 (60,26% UP y 53,06% DW), así como SOL fue aquella cepa en la que se detectaron los menores valores (39,23% UP y 34,59% DW) (**Tabla R8**).

Tabla R8. Contexto genómico donde se insertan los retroelementos identificados en las diferentes cepas a estudio.

Adaptada de Gómez, López, et al., 2021.

	B. M. López	Dm28	Y	Ikiakarora	SOL	CL Brener
(b)	L1Tc/NARTc/Sh/NC	L1Tc/NARTc/Sh/NC	L1Tc/NARTc/Sh/NC	L1Tc/NARTc/Sh/NC	L1Tc/NARTc/Sh/NC	L1Tc/NARTc/Sh/NC
TS	UP 2 / 1 / 22 / 7 (a) 32 (3,87%)	11 / 0 / 13 / 4 28 (4,64%)	13 / 3 / 28 / 16 60 (22,81%)	2 / 36 / 0 / 3 41 (6,43%)	13 / 41 / 95 / 39 188 (34,31%)	31 / 13 / 0 / 10 54 (13,33%)
	DW 20 / 68 / nd / nd 88 (12,79%)	14 / 9 / nd / nd 23 (6,71%)	-	15 / 6 / nd / nd 21 (7,37%)	17 / 2 / nd / nd 19 (5,97%)	10 / 1 / nd / nd 11 (5,56%)
MASP	UP 13 / 0 / 0 / 1 14 (1,69%)	5 / 0 / 0 / 6 11 (1,82%)	1 / 0 / 0 / 1 2 (0,76%)	6 / 3 / 7 / 0 16 (2,51%)	1 / 0 / 0 / 0 1 (0,18%)	1 / 0 / 0 / 0 1 (0,25%)
	DW 24 / 3 / nd / nd 27 (3,92%)	11 / 0 / nd / nd 11 (3,21%)	-	1 / 1 / nd / nd 2 (0,70%)	14 / 1 / nd / nd 15 (4,72%)	8 / 7 / nd / nd 15 (7,58%)
RHS	UP 0 / 1 / 0 / 1 2 (0,24%)	-	4 / 1 / 0 / 0 5 (1,90%)	52 / 30 / 0 / 10 92 (14,42%)	13 / 4 / 0 / 2 19 (3,47%)	7 / 8 / 0 / 4 19 (4,69%)
	DW - -	-	0 / 1 / nd / nd 1 (0,60%)	3 / 3 / nd / nd 6 (2,11%)	60 / 0 / nd / nd 60 (18,87%)	2 / 0 / nd / nd 2 (1,01%)
HP	UP 101 / 157 / 42 / 107 407 (49,27%)	210 / 68 / 20 / 66 364 (60,26%)	8 / 33 / 25 / 10 76 (28,9%)	98 / 80 / 24 / 34 236 (36,99%)	99 / 40 / 28 / 48 215 (39,23%)	49 / 72 / 29 / 33 183 (45,19%)
	DW 202 / 166 / nd / nd 368 (53,49%)	104 / 78 / nd / nd 182 (53,06%)	97 / 16 / nd / nd 113 (67,66%)	107 / 70 / nd / nd 177 (62,11%)	51 / 59 / nd / nd 110 (34,59%)	69 / 59 / nd / nd 128 (64,65%)

a: Número de retrotransposones L1Tc y NARTc que portan la huella Pr77 (L1Tc y NARTc), huellas Pr77 no asociadas con retrotransposones L1Tc y NARTc (Pr77Sh) y huellas Pr77 no clasificadas debido a la longitud de lectura (NC).

b: TS (Trans-sialidasa), MASP (proteína de superficie asociada a mucina), RHS (*Retrotransposon Hot Spot*) y HP (proteína hipotética) identificada corriente arriba (UP, *upstream*) o corriente abajo (DW, *downstream*) de retrotransposones portadores de Pr77 o Pr77Sh. El número de veces que se encontró cada asociación en cada genoma se representa en negrita y el porcentaje que representa el total de asociaciones identificadas se muestra entre paréntesis. nd: No determinado por la longitud de lectura (inferior a 15 nt).

### A.1.7. Variaciones en el contenido génico en las diferentes cepas de *T. cruzi*

Para estimar el contenido génico y el grado de similitud y divergencias entre los genomas de las cepas a estudio, se determinó el número de copias presentes de genes codificantes para 13 familias de proteínas de interés, así como el porcentaje que representaban del total de genes identificados. También se determinó el número de copias de genes codificantes para proteínas hipotéticas en todas las cepas. Las familias de genes analizadas fueron aquellas que codificaban para transalidasas (TS), mucinas, proteínas de superficie asociadas a mucinas (MASP), proteínas RHS (*Retrotransposon Hot Spot*) y TASV (Tripomastigote Alanina, Serina, Valina). Estas proteínas se asocian con procesos de virulencia del parásito y con retroelementos. Además, se analizaron genes de expresión constitutiva como quinesinas, helicasas de ARN, proteínas quinasas, glicosiltransferasas, cisteína peptidasas, proteínas de choque térmico de 70 kDa (HSP70), proteínas paraflagelares (PFR) y proteínas de unión a ARN.

Los resultados, resumidos en la **Tabla R9**, mostraron que los seis genomas de las cepas de *T. cruzi* a estudio presentaban un contenido génico similar de los genes de expresión constitutiva analizados, así como de los codificantes para proteínas hipotéticas (oscilando del 31,86% al 34,26% del total de genes identificados). Sin embargo, se identificaron diferencias entre cepas en el número de copias presentes de genes particulares y, por tanto, en el porcentaje que representaban en sus genomas. Así, la familia de genes codificantes para transalidasas tuvo una mayor representación en los genomas de las cepas Y (0,87%), CL Brener (0,85%) y SOL (0,83%) que contenían más del doble de genes TS que el resto de cepas analizadas, con valores de 0,48%, 0,34% y 0,19% para los genomas de Ikiakarora, B. M. López y Dm28, respectivamente. El número de copias de genes que codifican las proteínas MASP y TASV también fue superior en los genomas de las cepas Y, CL Brener y SOL. Respecto a los genes que codifican mucinas, se encontraron en mayor proporción en el genoma de CL Brener (0,34%), seguido de los genomas Y (0,23%), Ikiakarora (0,20%) y SOL (0,19%), con porcentajes muy superiores a los encontrados en los genomas de las cepas Dm28 (0,04%) y B. M. López (0,02%). Por último, el porcentaje de genes identificados codificantes de RHS fue más elevado en cuatro de las cepas (0,29% en Y, 0,26% en CL Brener y 0,17% en

## RESULTADOS

Ikiakarora y SOL), presentando valores superiores a los detectados en los genomas de las cepas Dm28 (0,08%) y B. M. López (0,13%).

**Tabla R9. Comparación del contenido génico entre cepas.** Número de copias de cada gen identificadas en el genoma ensamblado de cada cepa de *T. cruzi* y cuánto representan del total de genes identificados (en porcentaje) en cada caso. *Adaptada de* Gómez, López, et al., 2021.

	B. M. López	Dm28	Y	Ikiakarora	SOL	CL Brener
<b>TS</b>	26 (0,34)	15 (0,19)	72 (0,87)	54 (0,48)	107(0,83)	101 (0,85)
<b>Quinesina</b>	47 (0,61)	48 (0,62)	61 (0,74)	81 (0,71)	99 (0,77)	99 (0,83)
<b>Mucina</b>	2 (0,02)	3 (0,04)	19 (0,23)	23 (0,20)	24 (0,19)	40 (0,34)
<b>Helicasa de ARN</b>	41 (0,54)	43 (0,56)	49 (0,59)	69 (0,61)	71 (0,55)	72 (0,60)
<b>MASP</b>	14 (0,18)	12 (0,16)	108 (1,30)	99 (0,87)	201(1,57)	195(1,64)
<b>RHS</b>	10 (0,13)	6 (0,08)	24 (0,29)	19 (0,17)	22 (0,17)	31 (0,26)
<b>Quinasa</b>	276 (3,60)	296 (3,83)	317 (3,82)	412 (3,63)	456 (3,55)	423 (3,55)
<b>Glicosil-transferasa</b>	15 (0,2)	14 (0,18)	15 (0,18)	23 (0,28)	27 (0,21)	26 (0,22)
<b>Cisteína peptidasa</b>	21 (0,27)	21 (0,27)	32 (0,39)	47 (0,41)	52 (0,41)	43 (0,36)
<b>TASV</b>	3 (0,04)	1 (0,01)	7 (0,08)	3 (0,03)	7 (0,05)	7 (0,06)
<b>HSP70</b>	7 (0,09)	6 (0,08)	8 (0,10)	10 (0,09)	15 (0,12)	9 (0,07)
<b>PFR</b>	21 (0,27)	20 (0,26)	23 (0,28)	29 (0,26)	41 (0,32)	35 (0,29)
<b>Proteínas de unión a ARN</b>	79 (1,03)	69 (0,89)	73 (0,88)	96 (0,85)	115 (0,90)	104 (0,87)
<b>Proteína hipotética</b>	2.625 (34,3)	2.575 (33,3)	2.707 (32,6)	3.619 (31,9)	4.229 (32,9)	3.890 (32,7)

**TS:** trans-sialidasa; **MASP:** proteína de superficie asociada a mucina; **RHS:** *Retrotransposon hot spot*; **TASV:** Tripomastigote, Alanina, Serina, Valina; **HSP70:** proteína de choque térmico de 70 kDa; **PFR:** proteínas paraflagelares.



## **A.2. LOCALIZACIÓN GENÓMICA Y PREDICCIÓN DE ESTRUCTURA SECUNDARIA DE TIPO RIBOZIMA HDV EN SECUENCIAS Pr77Sh. ESTUDIO DE SU FUNCIÓN CATALÍTICA**

### **A.2.1. Identificación de huellas Pr77 no asociadas a retroelementos en el genoma de diferentes cepas de *T. cruzi***

La secuenciación de los genomas de las seis cepas de *T. cruzi*, descrita en la sección de *Resultados A.1*, permitió, además de analizar la presencia y abundancia de los retrotransposones L1Tc y NARTc, portadores de Pr77, en los genomas, la identificación de huellas Pr77 no asociadas a retroelementos, denominadas huellas Pr77 únicas o *single Pr77 hallmarks* (Pr77Sh). Estas secuencias se hallaron en todas las cepas a estudio mostrando diferencias en su abundancia relativa y en los porcentajes de identidad con la huella Pr77 de los retroelementos L1Tc/NARTc. En total, se encontraron 311 secuencias Pr77Sh en la cepa Ikiakarora, 153 en B. M. López, 83 en Y, 250 en SOL, 345 en CL Brener y 51 en la cepa Dm28.

### **A.2.2. Determinación de la localización genómica de las secuencias Pr77Sh**

Con objeto de conocer el entorno genómico en el que se encontraban las huellas Pr77Sh en las distintas cepas, se llevó a cabo un análisis de identidad de las secuencias que se encontraban corriente arriba y abajo de las diferentes huellas. Se consideraron aquellas lecturas de los genomas en las que Pr77Sh se encontraba flanqueada por una secuencia de, al menos, 45 nt de longitud, corriente arriba, y 100 nt corriente abajo de Pr77Sh y que contuvieran el motivo CGTG (descrito como motivo importante en la actividad promotora de Pr77 de L1Tc (Macías et al., 2016). De este modo, se extrajeron 86 Pr77Sh de la cepa Ikiakarora, 53 de B. M. López, 23 de Y, 66 de SOL, 108 de CL Brener y 11 de Dm28. Cabe destacar, que no se encontró ninguna Pr77Sh de longitud completa con una identidad con Pr77 de L1Tc/NARTc inferior a 64,9% en ninguna cepa. Estos porcentajes de identidad de las secuencias Pr77Sh con la huella Pr77 de L1Tc/NARTc oscilaron de 68,8% a 81,8% en la cepa Ikiakarora, de 64,9% a 98,7% en B. M. López, de 68,8% a 83,2% en Y, de 68,8% al 100% en SOL, de 68,8% a 85,7% en CL Brener y, de 68,8 a 77,9% en la cepa Dm28.

Para determinar la identidad de las secuencias flanqueantes de las huellas Pr77Sh se empleó la herramienta BLAST (Altschul et al., 1990) frente a las

secuencias incluidas en las bases de datos NCBI y TriTrypDB. Los resultados obtenidos, mostrados en la **Tabla R10**, indicaron que las secuencias Pr77Sh se insertaron en diferentes regiones de los genomas, en secuencias, en su mayor parte, de función conocida, siendo algunas de ellas comunes entre varias cepas y muchas otras específicas de cepa. La inserción más frecuente se observó en la secuencia codificante de trans-sialidasa (TS), la cual se detectó en todas las cepas en diferentes proporciones (que iban desde el 11,3% del total de secuencias Pr77Sh en B. M. López hasta el 42,4% en la cepa SOL). Así, se detectaron 27 inserciones en la secuencia TS en la cepa Ikiakarora, 6 en B. M. López, 5 en Y, 28 en SOL, 41 en CL Brener y 2 en la cepa Dm28. Se hallaron inserciones de Pr77Sh específicas de cepa en el ARNr 5S en diferentes proporciones en las cepas Ikiakarora (38), Y (1), SOL (14) y CL Brener (47), no detectándose en las cepas B. M. López y Dm28. Del mismo modo, se encontró una pequeña proporción de Pr77Sh insertadas en la secuencia del gen codificante de tirosina aminotransferasa en tres de las cepas analizadas. Así, se observaron 8 en Ikiakarora y B. M. López y 2 en la cepa SOL. Además, se identificaron inserciones específicas de Pr77Sh en un componente de la telomerasa encontrando 3 de ellas en Ikiakarora, 6 en Y, 4 en SOL y 2 en CL Brener y en la proteína vacuolar que se observó en las cepas Y (1), SOL (1) y CL Brener (6). Se detectó una única inserción de Pr77Sh en la secuencia repetitiva DGF-1 y también en la subunidad  $\beta$  del coatómero, únicamente en la cepa B. M. López. Curiosamente, también se encontraron 24 inserciones en el retroelemento VIPER sólo en la cepa B. M. López. Se hallaron, también, inserciones quiméricas constituidas por Pr77Sh flanqueadas por diferentes secuencias conservadas y en diferentes proporciones en las cepas Ikiakarora, Y, SOL y CL Brener, no detectándose ninguna de ellas en las cepas B. M. López y Dm28. Este es el caso de las secuencias Pr77Sh flanqueadas por el gen FG-GAP en el extremo 5' y por TS en el extremo 3' (6 (Ikiakarora y SOL), 1 (Y), 9 (CL Brener)) y por el ARN pseudouridilato sintasa en 5' y TS en el extremo 3' (2 en Ikiakarora y 1 en las cepas Y, SOL y CL Brener).

**Tabla R10. Número y sitios de inserción de *single Pr77 hallmarks (Pr77Sh)* en el genoma de las seis cepas de *T. cruzi* a estudio.** El número de huellas Pr77Sh en cada cepa de *T. cruzi* y en cada sitio de inserción genómica se indica en negrita, así como el porcentaje (entre paréntesis) que representa cada inserción en cada sitio específico del número total de Pr77Sh identificadas por cepa. El número de huellas Pr77Sh idénticas insertadas en secuencias idénticas se indica entre corchetes.

		<b>Ikiakarora (n= 86)</b>	<b>B. M. López (n=53)</b>	<b>Y (n=23)</b>	<b>SOL (n=66)</b>	<b>CL Brener (n=108)</b>	<b>Dm28 (n=11)</b>
<b>Homología en 5'</b>	<b>Homología en 3'</b>						
TS	TS	<b>27</b> (31,4%) [3,2,2]	<b>6</b> (11,3%) [2]	<b>5</b> (21,7%) [2]	<b>28</b> (42,4%) [3,5,2,2,2]	<b>41</b> (37,9%) [4,3,4,2,2,2,4]	<b>2</b> (18,2%) [2]
ARNr 5S	ARNr 5S	<b>38</b> (44,2%) [24]		<b>1</b> (4,3%)	<b>14</b> (21,2%) [7]	<b>47</b> (43,5%) [34,3,2]	
TAT	TAT	<b>8</b> (9,3%)	<b>8</b> (15,1%) [3, 2]		<b>2</b> (3%)		
Componente telomerasa	Componente telomerasa	<b>3</b> (3,5%)		<b>6</b> (26,1%) [3, 3]	<b>4</b> (6,1%)	<b>2</b> (1,9%)	
Proteína vacuolar	Proteína vacuolar			<b>1</b> (4,3%)	<b>1</b> (1,5%)	<b>6</b> (5,6%) [4]	
DGF-1	DGF-1		<b>1</b> (1,9%)				
Coatómero subunidad β	Coatómero subunidad β		<b>1</b> (1,9%)				
VIPER	VIPER		<b>24</b> (45,3%) [19]				
FG-GAP	TS	<b>6</b> (7%) [2]		<b>1</b> (4,3%)	<b>6</b> (9,1%) [5]	<b>9</b> (8,3%) [4, 3]	
ARN pseudo-uridilato sintetasa	TS	<b>2</b> (2,3%)		<b>1</b> (4,3%)	<b>1</b> (1,5%)	<b>1</b> (0,9%)	
Canal de cloruro	TAT			<b>5</b> (21,7%) [3]			
TS	HP				<b>3</b> (4,5%)		
α y β tubulina	HP		<b>1</b> (1,9%)				
Cromosoma Sylvio	ADNk		<b>1</b> (1,9%)				
Glicina DHG	DGF-1				<b>1</b> (1,5%)		
RHS	DGF-1				<b>1</b> (1,5%)		
DGF-1	HP				<b>1</b> (1,5%)		
RHS	HP				<b>1</b> (1,5%)		
HP	DGF-1				<b>1</b> (1,5%)		
HP	TS				<b>1</b> (1,5%)		
TAT	DGF-1				<b>1</b> (1,5%)		
HP o Cromosoma Sylvio	HP o Cromosoma Sylvio	<b>2</b> (2,3%)	<b>11</b> (20,8%) [2]	<b>3</b> (13%)		<b>2</b> (1,9%)	<b>9</b> (81,8%) [2]

**TS** (Trans-sialidasa); **TAT** (Tirosina aminotransferasa); **DGF-1** (*Dispersed gene family protein 1*); **ADNk** (minicírculos de ADN de kinetoplasto); **Glicina-DHG** (Glicina deshidrogenasa); **RHS** (*Retrotransposon Hot Spot*); **HP** (proteína hipotética).

También se identificaron inserciones específicas de las Pr77Sh en determinadas cepas (véase la **Tabla R10** para más detalles), como 5 de ellas flanqueadas por el gen del canal de cloruro (en el extremo 5') y el de la tirosina aminotransferasa (en el extremo 3') en la cepa Y y 3 de ellas flanqueadas por TS (en el extremo 5') y por una proteína hipotética (en el extremo 3') en la cepa SOL. Asimismo, se detectaron inserciones de Pr77Sh específicas de cepa que daban lugar a moléculas quiméricas en las cepas B. M. López y SOL. Estas moléculas quiméricas involucraban a los genes  $\alpha$  y  $\beta$  tubulina, minicírculo de ADN del kinetoplasto, glicina deshidrogenasa, DGF-1, RHS, tirosina aminotransferasa y genes codificantes de proteínas hipotéticas. Además, se detectó un número variable de copias de Pr77Sh en secuencias codificantes de identidad aún desconocida identificándose 2, 11, 3, 2 y 9 en las cepas Ikiakarora, B. M. López, Y, CL Brener y Dm28, respectivamente.

Agregado a lo anterior, los análisis comparativos de las huellas Pr77Sh realizados entre las seis cepas de *T. cruzi* mostraron que había varias inserciones de la misma Pr77Sh en las mismas secuencias genómicas dentro de la misma cepa y también entre diferentes cepas del parásito.

### **A.2.3. Análisis de la estructura secundaria de las secuencias Pr77Sh. Compatibilidad con la estructura ribozima tipo HDV**

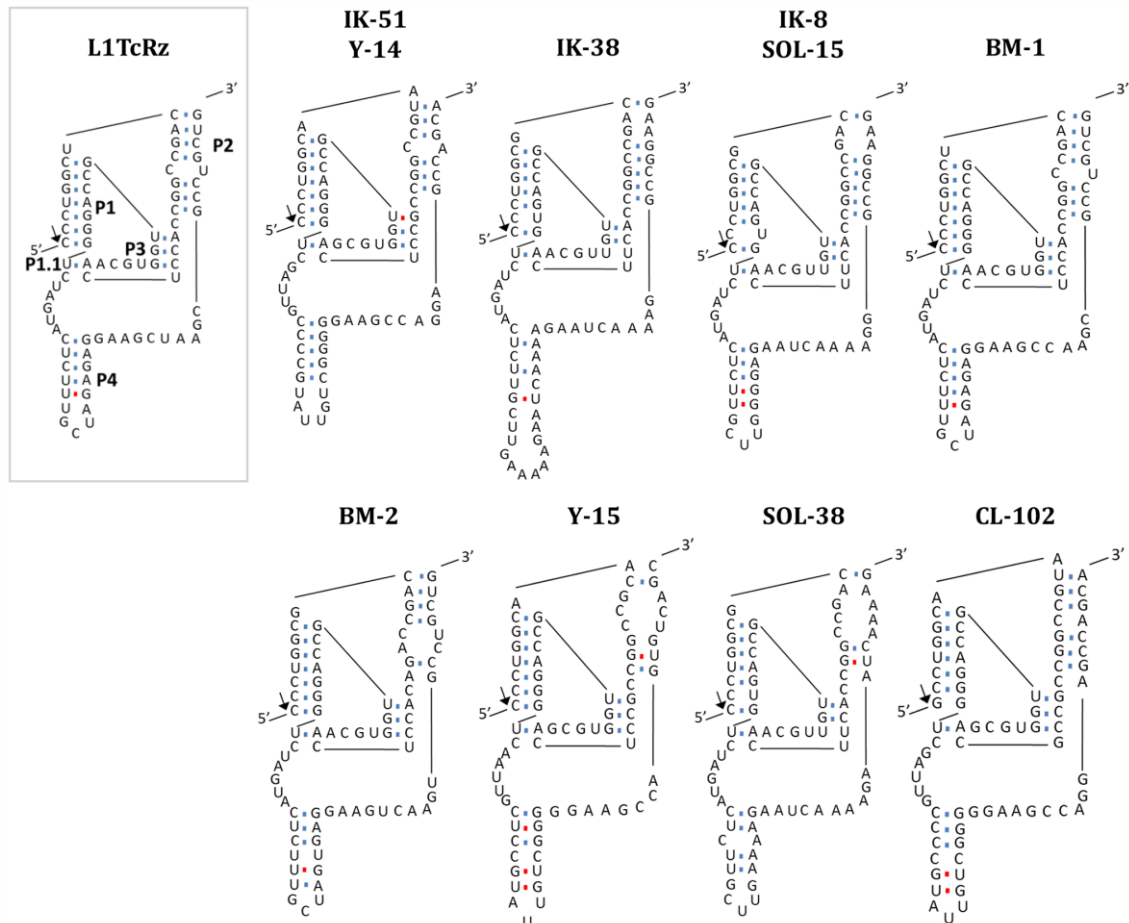
La predicción de la estructura ribozima tipo HDV se determinó mediante el plegado manual de todas las secuencias Pr77Sh de las diferentes cepas, utilizando como guía los plegamientos predichos para la ribozima L1TcRz. Las estructuras obtenidas se categorizaron en cuatro niveles dependiendo de si presentaban, independientemente de la secuencia, una estructura secundaria compatible con la ribozima tipo-HDV de L1Tc (I), había una divergencia en una hélice (II), presentaban un pseudonudo alterado (III), o no presentaban un plegamiento compatible con L1TcRz (IV). Los resultados obtenidos indicaron que todas las cepas contenían ribozimas tipo HDV con estructura tipo I, con la excepción de la cepa Dm28, cuyas secuencias Pr77Sh carecían de un plegamiento adecuado y compatible con el de L1TcRz. Interesantemente, la mayoría de las Pr77Sh de Ikiakarora, Y, SOL y CL Brener adoptaban una estructura ribozima de tipo I o II, alcanzando un número de hasta 74 (86%) en Ikiakarora, 20 (87%) en Y, 64 (97%) en SOL y 100 (92,6%) en la

cepa CL Brener. En la cepa B. M. López, 17 Pr77Sh (32,1%) adoptaron estructura tipo I, mientras que 36 de ellas (67,9%) no adoptaron una estructura compatible con la de L1TcRz. De igual manera, se observó que no había una correlación directa entre el grado de identidad de la secuencia Pr77Sh con la de Pr77 de L1Tc/NARTc y la estructura secundaria L1TcRz. Así, hubo secuencias que manteniendo un 74% de identidad con Pr77 de L1Tc/NARTc adquirieron una estructura ribozima tipo I y secuencias que con un 85,7% de identidad, no adoptaron una estructura compatible, teóricamente, con función HDV.

#### **A.2.4. Identificación de candidatos a ribozimas tipo HDV**

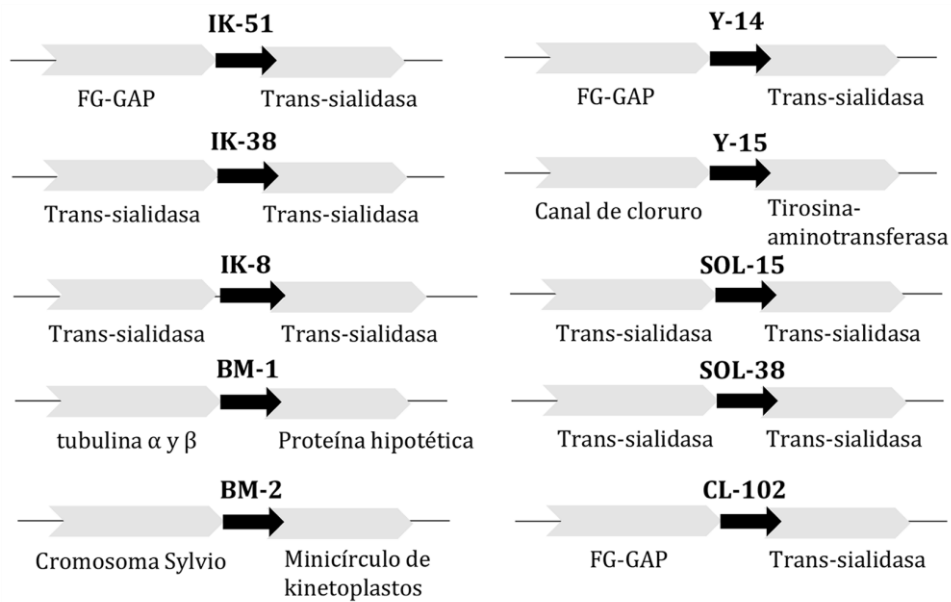
Dado que la secuencia DPE (CGTG) se ha descrito como un motivo importante requerido para la actividad promotora de la secuencia Pr77 de L1Tc (Macías et al., 2016), se analizó la presencia de la secuencia CGTG en las huellas Pr77Sh extraídas de los genomas de las 6 cepas de *T. cruzi*. Los resultados obtenidos indicaron la presencia del motivo DPE conservado en todas las cepas en cuanto a composición de la secuencia y localización (+25 a +28 respecto al nucleótido +1 de la secuencia Pr77) con la excepción de la cepa Dm28 en la que ninguna Pr77Sh portaba el motivo DPE. Se observó que existían algunas huellas Pr77Sh que, a pesar de mantener un alto grado de identidad de secuencia con Pr77 de L1Tc/NARTc, no conservaban el motivo DPE. En cuanto a la identidad de estas huellas Pr77Sh sin DPE respecto a la huella Pr77 de L1Tc/NARTc, esta llegó a ser de 81,8% en la cepa Ikiakarora, 84,4% en Y, 79,2% en B. M. López, 81,8% en SOL, 85,7% en CL Brener y 77,9% en Dm28.

Para determinar la presencia de actividad ribozima tipo-HDV en las secuencias Pr77Sh y basándonos en el plegamiento compatible con el de ribozimas tipo HDV adoptado por las huellas Pr77Sh y su contexto genómico, se seleccionaron diez secuencias Pr77Sh con estructura tipo I. Así, se seleccionaron las secuencias Ikiakarora-51Rz (IK-51) (idéntica a Y-14), Ikiakarora-38Rz (IK-38), Ikiakarora-8Rz (IK-8) (idéntica a SOL-15), B. M. López-1Rz (BM-1), B. M. López-2Rz (BM-2), Y-15Rz (Y-15), SOL-38Rz (SOL-38) y CL Brener-102Rz (CL-102) (**Figura R3**).



**Figura R3. Estructura secundaria de ribozimas seleccionadas teóricamente activas según su plegamiento.** Todas las *single Pr77 hallmarks* examinadas se plegaron manualmente tomando como referencia la estructura ribozima tipo HDV de L1TcRz (Sánchez-Luque et al., 2011) y se seleccionaron para el estudio aquellas compatibles con los requisitos de las ribozimas de tipo HDV descritas previamente o con mínimas alteraciones. En azul se señala el emparejamiento de bases adecuado según Watson y Crick y en rojo los emparejamientos oscilantes (*wobbles*). La flecha indica el punto de corte autocatalítico en 5' sobre el nucleótido +1 de la secuencia.

Las secuencias Pr77Sh seleccionadas presentaron un porcentaje de identidad con la secuencia Pr77 de L1Tc/NARTc que osciló de 74% (CL-102) a 98,7% (BM-1). En cuanto a su localización genómica, se encontraron insertadas en genes que codifican TS (IK-38, IK-8, SOL-15, SOL-38) o flanqueadas por genes que codifican FG-GAP y TS (IK-51, Y-14, CL-102), genes de  $\alpha$  y  $\beta$  tubulina y proteína hipotética (BM-1), minicírculo de ADN de kinetoplasto (BM-2) o genes codificantes de canal de cloruro y tirosina-aminotransferasa (Y-15) (**Figura R4**). La secuencia Pr77Rz de L1Tc se incluyó como control positivo en los ensayos.



**Figura R4.** Representación esquemática de la posición genómica de las *single Pr77 hallmarks* seleccionadas.

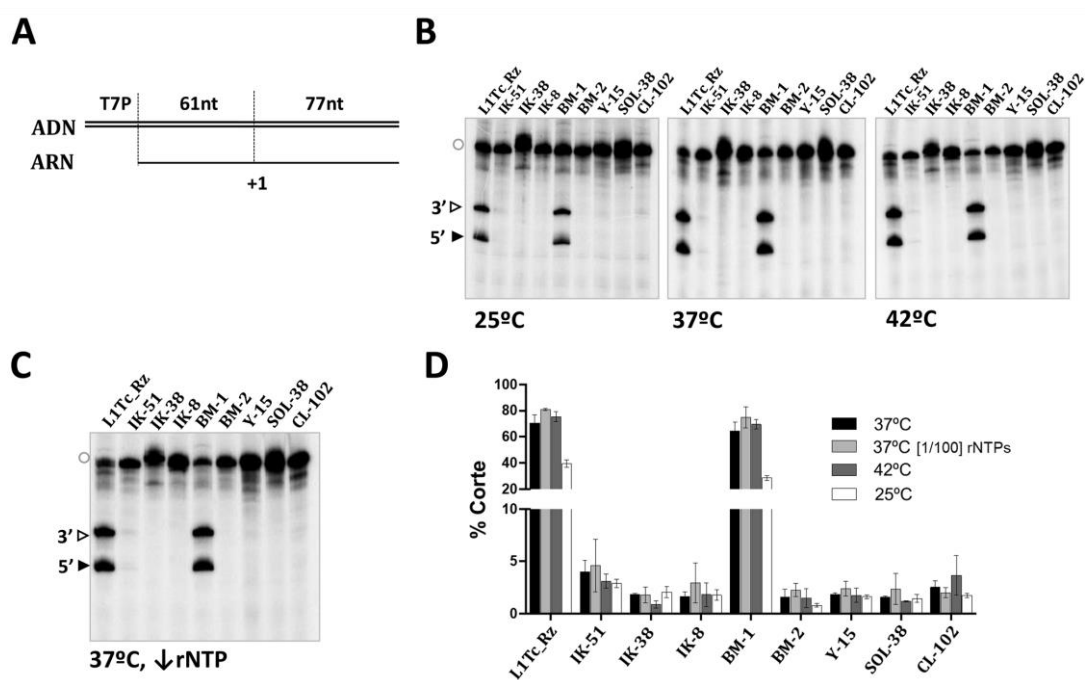
### A.2.5. Determinación de la función ribozima de las Pr77Sh candidatas seleccionadas

#### A.2.5.1. Actividad co-transcripcional *in vitro*

La actividad ribozima tipo HDV de cada Pr77Sh se determinó mediante la transcripción *in vitro* de los moldes de ADN generados tras clonar cada secuencia Pr77Sh en el vector pGEMT-Easy y la subsiguiente amplificación por PCR empleando el oligonucleótido M13Fw(-20) como cebador en 5' y el oligonucleótido antisentido complementario a cada secuencia Pr77Sh como cebador en 3', tal y como se describe en la sección de *Métodos 2.7*. La actividad de corte co-transcripcional se midió por transcripción *in vitro*, empleando la enzima T7 ARN polimerasa, de estos moldes que contenía el promotor de la T7 ARN polimerasa del vector pGEM-T, secuencia del plásmido pGEM-T Easy de 61 nt y la secuencia de cada Pr77Sh, después de resolverse las reacciones en geles de poliacrilamida desnaturalizantes como se describe en *Métodos 2.10*. Esta aproximación permitiría la detección, en el caso de ser funcionales y producirse corte en el nucleótido +1 de las diferentes Pr77Sh (Sánchez-Luque et al., 2011), de los dos productos resultantes del corte: el correspondiente a la región de pGEMT situada corriente arriba de las huellas Pr77Sh (de 61 nt) y la correspondiente a Pr77Sh (**Figura R5A**).

## RESULTADOS

Las reacciones de transcripción *in vitro* se realizaron a 25°C, 37°C y 42°C durante 2 h (**Figura R5B**) en las condiciones óptimas de actividad de L1TcRz. Se analizó también la actividad a concentración 100 veces menor de rNTP (**Figura R5C**). Los resultados obtenidos indicaron que BM-1 poseía una actividad co-transcripcional comparable a la detectada en L1TcRz a 25°C, 37°C y 42°C, así como a baja concentración de rNTP (**Figura R5B y 5C**). También se detectaron productos de corte en las reacciones realizadas con las secuencias IK-51 y CL-102 y en menor proporción en las demás secuencias estudiadas. Se realizó la cuantificación de los productos de corte para detectar las diferencias en las huellas Pr77Sh a estudio. Como se muestra en la **Figura R5D** se detectaron productos de corte en todas las secuencias, con mayor actividad a 37°C y baja concentración de rNTP, que fueron cercanos al 2%, con la excepción de IK-51 (4,6%) y BM-1 (75%) que mostraron mayor actividad, siendo el valor obtenido en L1TcRz del 81%. Cuando las reacciones se realizaron a 25°C se observó una importante reducción en la actividad detectada, especialmente en las ribozimas más activas IK-51 (2,9%), BM-1 (28,5%) y en el control L1TcRz (39,6%) (**Figura R5D**).



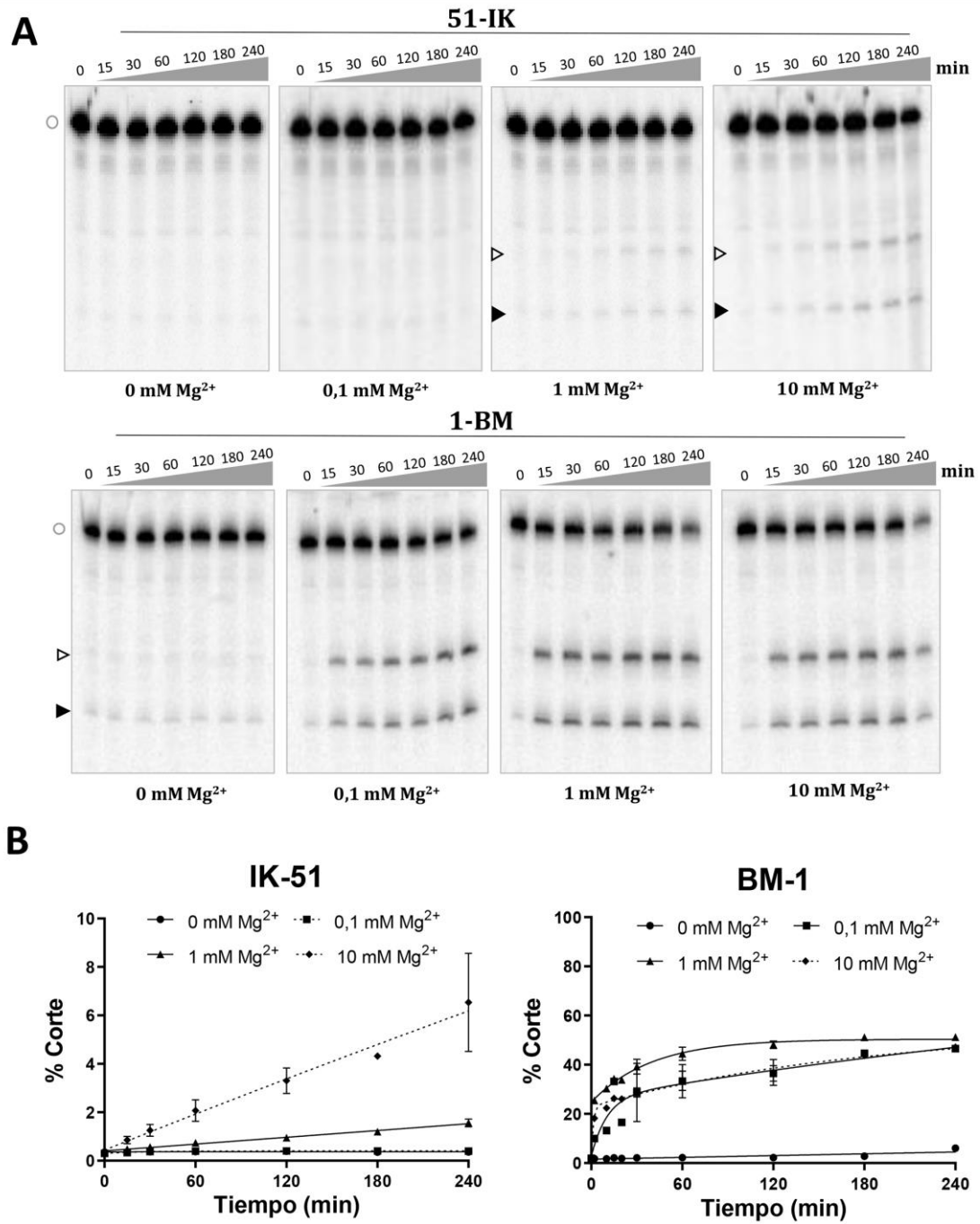
**Figura R5. Corte co-transcripcional de huellas Pr77Sh de *T. cruzi* a diferentes temperaturas.** (A) Las construcciones ensayadas se representan como líneas en el esquema. La línea vertical marcada como +1 representa el punto de corte esperado. (B) Se representan las autorradiografías del análisis electroforético de las reacciones de corte co-transcripcional a diferentes temperaturas y (C) a menor concentración de ribonucleótidos. El fragmento de longitud completa sin escindir se representa con círculos vacíos; el producto de corte 5' por puntas de flechas negras (▶) y el producto de corte 3' por puntas de flecha blancas (▷). (D) Representación de la cuantificación del porcentaje de corte promedio de duplicados independientes de cada reacción co-transcripcional. El eje Y se interrumpe para mejorar la visualización de los valores de corte inferiores al 5%.



### ***A.2.5.2. Actividad post-transcripcional in vitro***

Para analizar las propiedades de corte post-transcripcional de las ribozimas estudiadas, los transcritos que permanecían sin cortar en los ensayos co-transcripcionales de actividad ribozima tipo HDV, se escindieron de los geles y eluyeron en tampón libre de  $Mg^{2+}$  para su renaturalización (ver sección de *Métodos 2.9*). En el análisis se valoró la capacidad de corte post-transcripcional de las Pr77Sh seleccionadas a 37°C, a distintos tiempos (15, 30, 60, 120, 180 y 240 min) en presencia de EDTA y a tres concentraciones diferentes de  $Mg^{2+}$  (0,1, 1 y 10 mM). Las reacciones se resolvieron electroforéticamente en geles desnaturalizantes al 8% y se revelaron y cuantificaron los productos de corte, como se detalla en la sección de *Métodos 2.10*.

IK-51 y BM-1 mostraron tener una importante actividad post-transcripcional ya que se detectaron productos de corte en ambos casos. La actividad mostró dependencia iónica y mejorar con el tiempo. Así, la máxima actividad se detectó tras 240 min de reacción a 10 mM de  $Mg^{2+}$  para IK-51 (6,5%) (**Figura R6A**) y a 1 mM de  $Mg^{2+}$  para BM-1 (52,7%) (**Figura R6B**). En el resto de las huellas analizadas, si bien se detectó actividad post-transcripcional tipo HDV, esta fue baja (datos no mostrados).



**Figura R6. Corte post-transcripcional de las huellas Pr77Sh IK-51 y BM-1 identificadas en el genoma de *T. cruzi*.** (A) Se representan las autorradiografías del análisis electroforético de las reacciones de corte post-transcripcional *in vitro* en ausencia y a diferentes concentraciones de Mg<sup>2+</sup> (0,1 mM, 1 mM y 10 mM). El fragmento de longitud completa sin cortar se representa con círculos vacíos; el producto de corte 5' por puntas de flechas negras (▶) y el producto de corte 3' por puntas de flecha blancas (▷). (B) Cuantificación de los duplicados de cada cinética. Los datos se ajustan a una curva cinética exponencial *two-phase decay*.

### A.3. ESTUDIO GENÓMICO Y DEL CONTENIDO EN RETROELEMENTOS EN LA CEPA SOL DE *T. cruzi* SECUENCIADA POR TECNOLOGÍA PACBIO

La cepa SOL de *T. cruzi*, clasificada como DTU V, fue aislada de un niño nacido en España infectado por transmisión vertical (Murcia et al., 2013). Dada la relevancia de esta cepa, su genoma se secuenció también en el laboratorio empleando tecnología PacBio. PacBio se caracteriza por generar lecturas de mayor tamaño que facilitan el correcto ensamblaje del genoma en *contigs* más grandes permitiendo un análisis más preciso del contenido génico y de retroelementos los cuales tienen un tamaño de hasta cerca de 5 kb.

#### A.3.1. Secuenciación, ensamblaje y anotación del genoma

El ADN genómico de la cepa SOL de *T. cruzi* se secuenció y ensambló en *contigs* por tecnología de secuenciación *Pacific Biosciences* (PacBio) por otros miembros del laboratorio. Posteriormente, este genoma se reensambló mediante el alineamiento de las lecturas resultantes de la secuenciación y, seguidamente, refinamiento con las lecturas cortas procedentes de la secuenciación *Illumina* (sección *Métodos 1.5*), método que se ha descrito aumenta con precisión la calidad del ensamblaje (Chen et al., 2021).

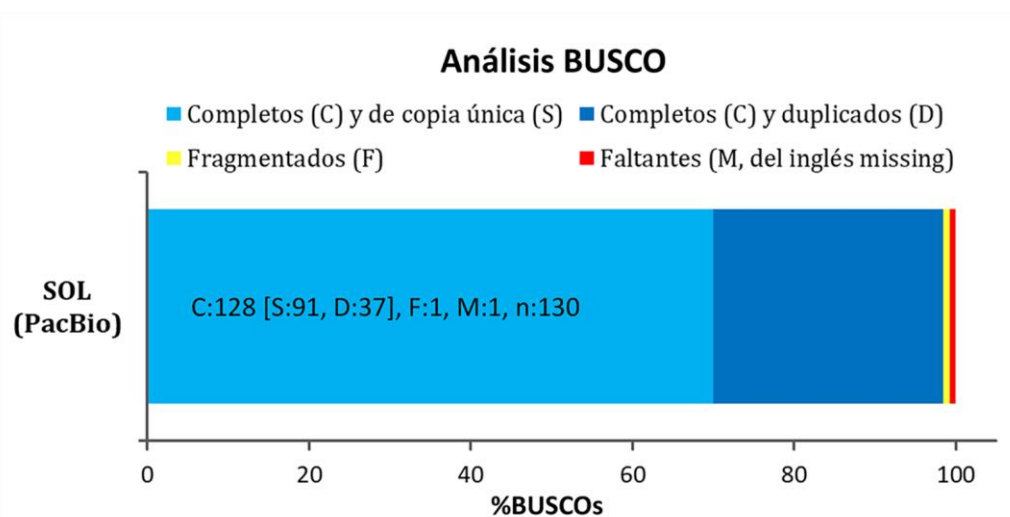
En la **Tabla R11** se muestra el resumen de las métricas obtenidas tras la secuenciación y ensamblaje del genoma. Así, se obtuvieron 692.620 lecturas crudas que se filtraron y corrigieron por calidad empleando el programa Canu v2.0, dando lugar a 1.783 *contigs* que abarcaron los 67.122.901 pares de bases (pb) que constituyen el genoma ensamblado. La longitud promedio de los *contigs* fue de 37.646 pb, siendo el menor tamaño identificado de 1.009 pb, mientras que el mayor *contig* presentó un tamaño de 566.077 pb. Adicionalmente, se detectó un valor de N50, calculado en pb, de 51.086 pb (**Tabla R11**).

**Tabla R11. Resumen de métricas de secuenciación y ensamblaje del genoma de la cepa SOL de *T. cruzi* secuenciado por tecnología PacBio.**

Cepa	Secuenciación Lecturas crudas	Ensamblaje					Bases totales
		<i>contigs</i>	N50 (pb)	Menor tamaño	Promedio <i>contig</i> (pb)	Mayor tamaño	
SOL (PacBio)	692.620	1.783	51.086	1.009	37.646	566.077	67.122.901

## RESULTADOS

A continuación, se evaluó la integridad del ensamblaje resultante mediante el análisis de evaluación comparativa de ortólogos universales de copia única (BUSCO) para el conjunto de datos de Euglenozoa como predeterminado, el cual contiene 130 proteínas BUSCO que se espera estén presentes en el genoma como genes de copia única. Los resultados representados en la **Figura R7**, mostraron que el porcentaje de genes BUSCO completos (C) y de copia única (S) identificados fue del 70% de los 130 grupos de BUSCO. Para BUSCOs completos y duplicados (D) el porcentaje obtenido fue del 28,5%, reduciéndose hasta un 0,8% para los BUSCOs fragmentados (F) y BUSCOs faltantes (M) (**Figura R7**). Los porcentajes obtenidos reflejan muy buena integridad del genoma ensamblado al contener 70% de genes BUSCO completos y de una sola copia.



**Figura R7. Análisis BUSCO de la integridad del ensamblaje del genoma.** Para la evaluación del ensamblaje del genoma se empleó el conjunto de datos BUSCO (*Benchmarking set of Universal Single-Copy Orthologues*) de Euglenozoa odb\_10, que incluye 130 BUSCOs. El gráfico de barras muestra: % de genes BUSCO completos y de copia única (azul claro), completos y duplicados (azul oscuro), fragmentados (amarillo) y faltantes (rojo) en el ensamblaje.

Del mismo modo que para los genomas secuenciados por tecnología Ion Torrent descritos en la sección de *Resultados A.1*, la anotación de los genes presentes en este genoma se llevó a cabo utilizando la herramienta de predicción de genes AUGUSTUS y el flujo de trabajo OrthoMCL, tal y como se describe en la sección de *Métodos 1.7*. Para la predicción de genes mediante AUGUSTUS, se empleó como referencia el genoma de *T. cruzi* CL Brener non-Esmeraldo-like disponible en la base de datos TriTrypDB. La selección de este genoma de referencia se vio motivada por tratarse de uno de los haplotipos de la cepa de referencia CL Brener, la primera

secuenciada de este organismo (El-Sayed, Myler, Bartholomeu, et al., 2005) y ampliamente estudiada. Por otro lado, CL Brener non-Esmeraldo-like fue la cepa seleccionada para la anotación del genoma de SOL secuenciado por tecnología Ion Torrent (por ser la cepa frente a la que presentó mayor número de *best hit*). Además, el genoma de CL Brener non-Esmeraldo-like posee una buena anotación, constituido por secuencias de elevado tamaño al estar ensamblado en cromosomas.

Tras la anotación, se determinaron otros parámetros del genoma que se recopilan en la **Tabla R12**. Este genoma poseía un contenido de G+C del 51,8%. El número de genes identificados y anotados fue de 31.570, lo que representa el 57,7% del total de pares de bases del genoma ensamblado, con un tamaño promedio de 1.227 pb. Del número total de genes anotados, pudo asignarse información a 20.257 genes basada en la identidad de su secuencia con proteínas homólogas descritas en otros organismos. Por otro lado, se identificaron 9.308 genes anotados como genes codificantes de proteínas hipotéticas (**Tabla R12**).

**Tabla R12. Características del genoma ensamblado y genes anotados en la cepa SOL de *T. cruzi* secuenciada por tecnología PacBio.**

Cepa	%GC <sup>a</sup>	Genes	Tamaño medio de los genes <sup>b</sup> (pb)	% genoma genes anotados <sup>c</sup>	Homología con proteínas de función conocida <sup>d</sup>	Proteínas hipotéticas <sup>e</sup>
SOL (PacBio)	51,8	31.570	1.227	57,7	20.257	9.308

a: Contenido en guanina-citosina.

b: Suma de la longitud de todos los genes identificados dividido entre el total de genes identificados.

c: Porcentaje del genoma ocupado por los genes anotados.

d: Número de genes que poseen anotación basado en la identidad con proteínas de función conocida.

e: Número de genes que codifican proteínas hipotéticas.

### A.3.2. Estimación del contenido génico

Para la estimación del contenido génico de este genoma, se determinó el número de copias de genes codificantes de 16 familias de proteínas de función conocida y de proteínas hipotéticas, así como el porcentaje que representaban del total de genes identificados en el genoma. Los genes seleccionados codificaban para trans-sialidasas (TS), mucinas, proteínas de superficie asociadas a mucinas (MASP), proteínas RHS (*Retrotransposon Hot Spot*), la metaloproteasa de zinc GP63 (GP63), DGF-1 (*Dispersed Gene Family protein 1*) y TASV (Tripomastigote Alanina, Serina,

Valina). Estas proteínas se asocian con procesos de virulencia del parásito o con retroelementos. Además, también se analizaron genes de expresión constitutiva como quinesinas, la proteína de membrana de kinetoplastos de 11 kDa (KMP11), helicasas de ARN, proteínas quinasas, glicosiltransferasas, cisteína peptidasas, factor de elongación 1 gamma (EF-1 $\gamma$ ), proteínas de choque térmico de 70 kDa (HSP70) y proteínas paraflagelares (PFR).

Los resultados obtenidos, resumidos en la **Tabla R13**, mostraron que, de los genes analizados, el mayor número de copias correspondía con genes codificantes de TS (1.740) y RHS (1.298). Además, se detectó un elevado número de genes codificantes para mucinas (517), MASP (967), DGF-1 (628) y GP63 (301) en esta cepa. Del mismo modo, también se identificó un elevado número de copias de genes de expresión constitutiva como quinasas (915), cisteína peptidasa (143), glicosiltransferasa (177), quinesina (115) y helicasas de ARN (93). De igual manera, se identificaron 4 copias del gen codificante de la proteína KMP11 lo cual es consistente con el número de copias de este gen identificadas experimentalmente, lo cual soporta la calidad del genoma ensamblado (Bridge et al., 1998; Thomas et al., 2000). Por otro lado, se identificaron 71 copias de genes codificantes de EF-1 $\gamma$ , 54 de TASV, 44 de HSP70 y 47 de genes codificantes de PFR. Los resultados también mostraron la existencia en el genoma de SOL de un elevado número de genes codificantes de proteínas hipotéticas (29,5% del total de genes identificados) (**Tabla R13**).

**Tabla R13. Contenido génico del genoma de SOL de *T. cruzi* (PacBio).** Número de copias de cada gen identificadas en el genoma ensamblado de *T. cruzi* de la cepa SOL y cuánto representan del total de genes identificados (en porcentajes) en cada caso.

Gen	
Trans-sialidasa (TS)	<b>1.740</b> (5,5%)
Quinesina	<b>115</b> (0,36%)
Mucina	<b>517</b> (1,6%)
Helicasa de ARN	<b>93</b> (0,29%)
Proteína de superficie asociada a mucina (MASP)	<b>967</b> (3,1%)
RHS	<b>1.298</b> (4,1%)
Quinasa	<b>915</b> (2,9%)
Glicosiltransferasa	<b>177</b> (0,56%)
Metaloproteasa de zinc GP63	<b>301</b> (0,95%)
DGF-1	<b>628</b> (2%)
Cisteína peptidasa	<b>143</b> (0,45%)
KMP11	<b>4</b> (0,01%)
Factor de elongación 1 gamma (EF-1 $\gamma$ )	<b>71</b> (0,22%)
TASV	<b>54</b> (0,17%)
HSP70	<b>44</b> (0,14%)
Proteínas paraflagelares (PFR)	<b>47</b> (0,15%)
Proteínas hipotéticas	<b>9.308</b> (29,5%)

**RHS:** *Retrotransposon Hot Spot*; **DGF-1:** *Dispersed Gene Family protein 1*; **KMP11:** Proteína de membrana de kinetoplastidos de 11 kDa; **TASV:** Tripomastigote, Alanina, Serina, Valina; **HSP70:** Proteína de choque térmico de 70 kDa.

### A.3.3. Análisis del contenido en retroelementos

Al igual que se estudió en los genomas resultantes de la secuenciación Ion Torrent, en el genoma de esta cepa secuenciada por PacBio se identificaron las huellas Pr77, así como los retrotransposones L1Tc, NARTc y DIRE que la portaban. Con este fin, se realizaron búsquedas independientes de cada secuencia en el genoma ensamblado empleando la herramienta BLASTN. Los resultados obtenidos, resumidos en las **Tablas R14 y R15** muestran el número de secuencias identificadas en los genomas que presentaron un tamaño de alineamiento de, al menos, el 70% del tamaño de la secuencia en cuestión. De este modo, se identificaron 282 secuencias Pr77 en el genoma de SOL, con una identidad de entre un 89% y 100% con la secuencia Pr77 de L1Tc (**Tabla R14**). De igual manera, se identificaron 78 secuencias del retrotransposón L1Tc, todas ellas con identidad superior al 92%. Un 88,5% de estas secuencias (69) presentaron un tamaño superior al 95% del tamaño de L1Tc (**Tabla R14**). Respecto al retroelemento NARTc, se hallaron 93 copias con una identidad superior al 80%. Casi la totalidad de ellas (91, suponiendo un 97,8%)

## RESULTADOS

mostraron un tamaño de, al menos, el 95% del tamaño de la secuencia NARTc (Tabla R14).

**Tabla R14. Secuencias Pr77, L1Tc y NARTc identificadas en el genoma de la cepa SOL de *T. cruzi* (PacBio) mediante análisis BLASTN.** Se indica el número de secuencias identificadas con un tamaño de alineamiento de, al menos, el 70%, 80%, 90% o 95% del tamaño de la secuencia a estudio.

Secuencia Pr77 <sup>a</sup>	L1Tc <sup>b</sup> (tamaño: 4893 nt)		NARTc <sup>c</sup> (tamaño: 263 nt)	
	282	>70% (>3.425 nt)	78	>70% (>184 nt)
	>80% (>3.914 nt)	74	>80% (>210 nt)	93
	>90% (>4.403 nt)	72	>90% (>236 nt)	92
	>95% (>4.648 nt)	69	>95% (>249 nt)	91

**a:** Secuencia Pr77 de L1Tc; **b:** Secuencia L1Tcg55 (L1Tc\_X83098); **c:** Secuencia NARTc AF215898.

Para la identificación en el genoma de SOL de elementos DIREs, se emplearon como sondas las secuencias de 5 DIREs distintos (Bringaud, Ghedin, et al., 2006). Así, se hallaron 107 DIREs con un tamaño de, al menos, el 70% del tamaño de las secuencias DIREs, 88 de ellos de tamaño completo (Tabla R15). La identidad de estos elementos con la secuencia DIRE empleada para su identificación fue del 80,4% al 99,7%.

**Tabla R15. Secuencias DIREs identificadas en el genoma de la cepa SOL de *T. cruzi* (PacBio) mediante análisis BLASTN.** Se indica el número de secuencias identificadas con un tamaño de alineamiento de, al menos, el 70%, 80% y 90% del tamaño de la secuencia a estudio.

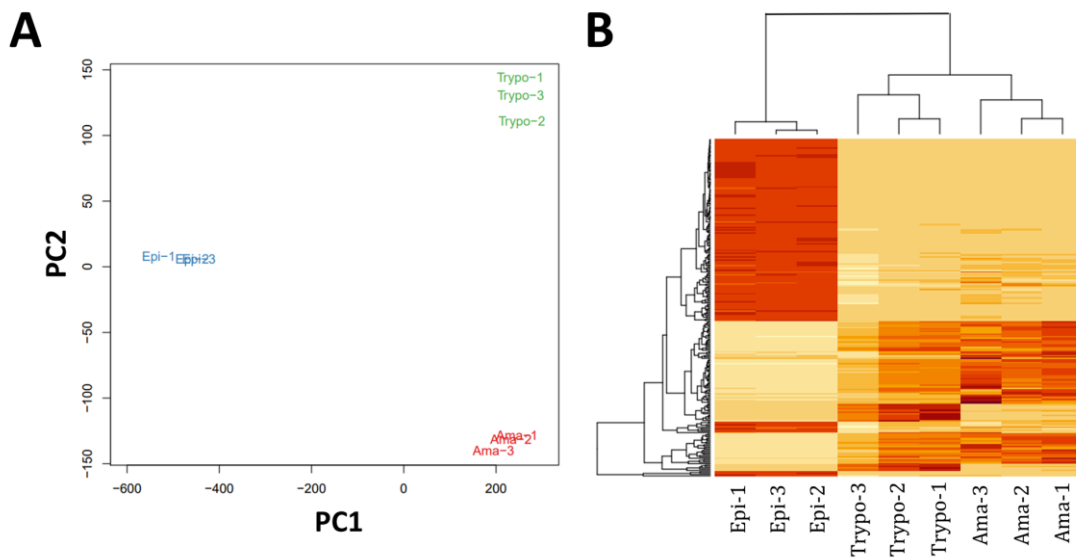
DIREs	Tamaño de alineamiento		
	> 70%	> 80%	> 90%
AAHK01000664.1 (4.072 nt)	31	28	28
AAHK01001018.1 (4.210 nt)	12	11	11
AAHK01001032.1 (3.122 nt)	2	2	2
AAHK01000322.1 (4.106 nt)	51	46	46
AAHK01000335.1 (3.357 nt)	11	6	1
<b>Total</b>	107	93	88



### A.3.I. ANÁLISIS DE LA EXPRESIÓN GÉNICA DIFERENCIAL EN LAS TRES FORMAS DE *T. cruzi* EPIMASTIGOTE, TRIPOMASTIGOTE Y AMASTIGOTE MEDIANTE SECUENCIACION MASIVA DE ARN (ARN-seq)

Uno de los objetivos planteados en la presente Tesis Doctoral, fue el estudio comparativo del transcriptoma y el análisis de expresión génica diferencial en los 3 estadios (epimastigote, tripomastigote y amastigote) de la cepa SOL de *Trypanosoma cruzi*. Para ello, se obtuvieron formas tripomastigotas y amastigotas del parásito y purificó ARN total de cada una de las formas del parásito y realizó secuenciación masiva de ARN (ARN-seq), por triplicado, en cada una de ellas. El alineamiento de cada una de las tres réplicas con el genoma ensamblado y anotado de la cepa SOL de *T. cruzi* (PacBio) permitió conocer la identidad de cada uno de los transcritos y realizar análisis comparativos de abundancia de cada uno de ellos en las tres formas del parásito. Así, el análisis realizado detectó expresión de 22.541 genes, lo que representa un 71,4% de los genes identificados en el genoma ensamblado de esta cepa (31.570 genes).

Para el análisis comparativo, en una primera aproximación, se realizó un análisis de componentes principales (PCA) y un mapa de calor basado en los valores de expresión de las diferentes muestras correspondientes a los tres estadios (**Figura R8**). Como se observa en la **Figura R8A**, cuando se representó la componente principal 1 (PC1) y la componente principal 2 (PC2), las muestras de cada estadio se agruparon de forma independiente. A su vez, se observó que las 3 réplicas de cada estadio poseían puntuaciones similares, ocupando las mismas posiciones en el gráfico lo cual confirmó la idéntica calidad de las réplicas en cuanto a nivel de expresión génica. La representación del nivel de expresión génica en las tres formas del parásito en un mapa de calor que incluye la expresión de los 250 genes que experimentan mayores diferencias de expresión nos permitió observar, con claridad, un patrón de expresión diferente en las formas epimastigotes respecto a las formas tripomastigotes y amastigotes que agrupaban a la forma epimastigote en un clado independiente (**Figura R8B**). Si bien se observó una mayor similitud entre los niveles de expresión génica de la forma tripomastigote y amastigote, también se distinguieron diferencias importantes en la expresión de algunos genes representados, generándose un nuevo clado que agrupó, de forma independiente, las muestras correspondientes a cada estadio (**Figura R8B**).



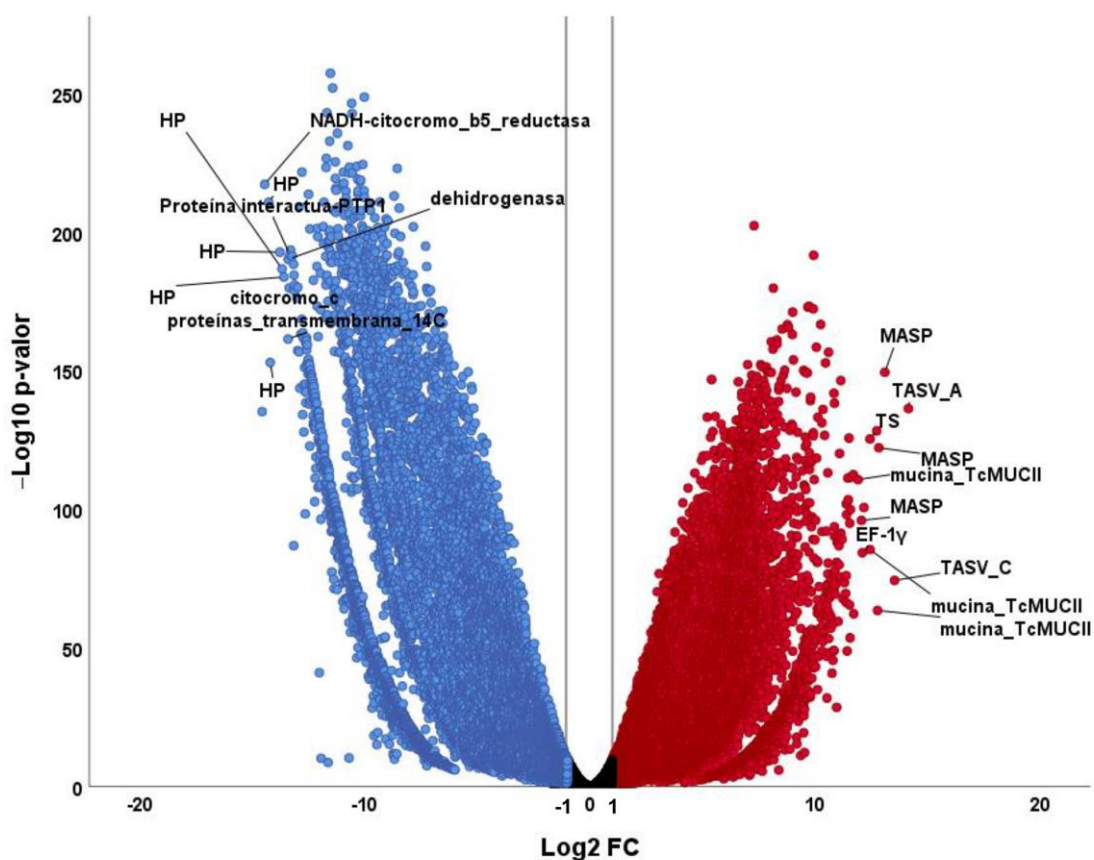
**Figura R8. Análisis de expresión génica diferencial entre los estadios epimastigote (Epi), tripomastigote (Trypo) y amastigote (Ama) de *T. cruzi*.** (A) Análisis de componentes principales sobre los valores de expresión de las muestras analizadas. Las componentes principales 1 (PC1) y 2 (PC2) se representan en el eje X e Y, respectivamente. (B) Mapa de calor de los 250 genes con mayores diferencias de expresión entre estadios de *T. cruzi*. Los valores del nivel de expresión génica de cada gen se representan como color, que van del rojo oscuro al amarillo claro, según el valor de expresión más alto o más bajo de cada muestra para cada gen, respectivamente.

Con objeto de conocer las diferencias de expresión entre estadios, se comparó la expresión de los 22.541 genes identificados entre los diferentes estadios de *T. cruzi*. Para ello, los datos procedentes del análisis comparativo se representaron en gráficos de volcán para visualizar gráficamente la magnitud del cambio y la significancia estadística del mismo.

### A.3.I.1. Expresión génica diferencial en tripomastigote versus epimastigote

El análisis de expresión génica diferencial entre tripomastigote y epimastigote reveló que 18.224 genes se expresaban diferencialmente entre estos dos estadios con significancia estadística ( $FDR < 0,05$ ), siendo las diferencias de expresión de, al menos, el doble o la mitad ( $\text{Log}_2 \text{FC} > 1$  o  $< -1$ , respectivamente). El número de genes sobreexpresados fue de 9.153 y el de infraexpresados 9.071 en tripomastigote respecto a la forma epimastigote con  $p < 0,05$  (Figura R9).

En el gráfico de volcán mostrado en la **Figura R9** se indica la identidad de los 20 genes con mayores diferencias de expresión, los 10 con mayor sobreexpresión y los 10 con mayor infraexpresión en tripomastigote *versus* epimastigote. Entre los genes con mayor sobreexpresión en tripomastigote nos encontramos genes que codifican para TS, EF-1 $\gamma$ , mucinas, proteínas de superficie asociadas a mucinas (MASP) y para proteínas de la familia Tripomastigote, Alanina, Serina, Valina (TASV).



**Figura R9.** Representación en gráfico de volcán de los transcritos expresados diferencialmente entre los estadios tripomastigote y epimastigote de *T. cruzi* determinado por ARN-seq. El eje X representa el  $\log_2$  del *fold change* de expresión entre tripomastigote y epimastigote ( $\log_2$  FC), donde el FC se calcula como la ratio de la expresión (medida en RPKM) entre los dos estadios (tripomastigote/epimastigote). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor *p* ajustado ( $-\log_{10}$  p-valor). Las líneas verticales grises representan el  $\log_2$  de FC de -1 y 1 (correspondiente a FC de 0,5 y 2, respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a los genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre tripomastigote y epimastigote.

## RESULTADOS

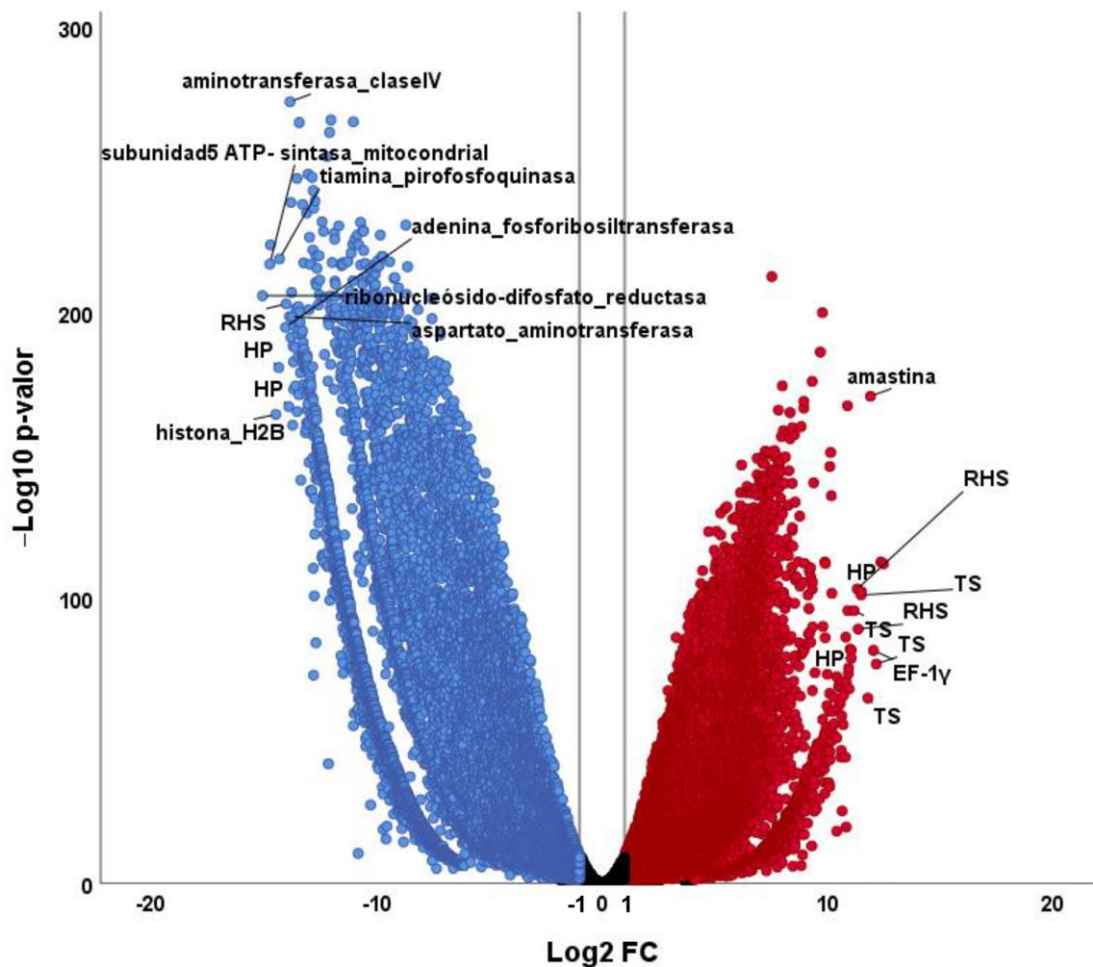
Dado el elevado número de genes expresados diferencialmente, se analizó si estos superaban umbrales de cambio de expresión superiores. Los resultados, recopilados en la **Tabla R16**, mostraron que 6.591 genes se expresaban, al menos, 8 veces más en tripomastigote ( $\text{Log}_2 \text{FC} > 3$ ) y 6.817 genes, al menos, una octava parte ( $\text{Log}_2 \text{FC} < -3$ ) al comparar con la forma epimastigote. De estos genes, 2.994 y 4.003, se expresaron, al menos, 64 veces más ( $\text{Log}_2 \text{FC} > 6$ ) o menos de 1/64 ( $\text{Log}_2 \text{FC} < -6$ ), respectivamente, en tripomastigote *versus* epimastigote. Aun exigiendo, valores de  $\text{log}_2 \text{FC} > 10$  o  $< -10$ , se detectaron 958 genes, 123 sobreexpresados ( $\text{Log}_2 \text{FC} > 10$ ) y 835 infraexpresados ( $\text{Log}_2 \text{FC} < -10$ ) en tripomastigote en comparación con la forma epimastigote del parásito (**Tabla R16**).

**Tabla R16.** Número de genes expresados diferencialmente (GED) en la forma tripomastigote de *T. cruzi* *versus* epimastigote determinado por ARN-seq. FC: fold change.

Tripomastigote <i>versus</i> Epimastigote	GED	Sobre-expresados	Infra-expresados	Expresión
$\text{Log}_2 \text{FC} > 1$ o $< -1$	18.224	9.153	9.071	FC = 2; FC = 1/2
$\text{Log}_2 \text{FC} > 3$ o $< -3$	13.408	6.591	6.817	FC = 8; FC = 1/8
$\text{Log}_2 \text{FC} > 5$ o $< -5$	9.157	4.278	4.879	FC = 32; FC = 1/32
$\text{Log}_2 \text{FC} > 6$ o $< -6$	6.997	2.994	4.003	FC = 64; FC = 1/64
$\text{Log}_2 \text{FC} > 7$ o $< -7$	4.775	1.627	3.148	FC = 128; FC = 1/128
$\text{Log}_2 \text{FC} > 10$ o $< -10$	958	123	835	FC = 1.024; FC = 1/1.024

### A.3.I.2. Expresión génica diferencial en amastigote *versus* epimastigote

Basados en los valores de RPKM obtenidos, los resultados revelaron que 17.981 genes se expresaban de forma diferencial entre el estadio amastigote y epimastigote con significancia estadística ( $\text{FDR} < 0,05$ ), considerando siempre como umbral un *fold change* mayor de 2 o inferior a 1/2 ( $\text{Log}_2 \text{FC} > 1$  o  $< -1$ , respectivamente). Así, 8.982 genes se sobreexpresaban y 8.999 se infraexpresaban en amastigote en comparación con el estadio epimastigote con  $p < 0,05$  (**Figura R10**). En el gráfico de volcán mostrado en la **Figura R10** se indica la identidad de los 20 genes con mayores diferencias de expresión, los 10 con mayor sobreexpresión y los 10 con mayor infraexpresión en amastigote *versus* epimastigote. Entre ellos cabe destacar la presencia entre los más sobreexpresados de genes codificantes para trans-sialidasas (TS), RHS (*Retrotransposon Hot Spot*) y factor de elongación 1 $\gamma$  (EF-1 $\gamma$ ).



**Figura R10.** Representación en gráfico de volcán de los transcritos expresados diferencialmente entre los estadios amastigote y epimastigote de *T. cruzi* determinado por ARN-seq. El eje X representa el  $\log_2$  del *fold change* de expresión entre amastigote y epimastigote ( $\log_2$  FC), donde el FC se calcula como la ratio de la expresión (medida en RPKM) entre los dos estadios (amastigote/epimastigote). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ajustado ( $-\log_{10}$  p-valor). Las líneas verticales grises representan el  $\log_2$  de FC de -1 y 1 (correspondiente a FC de 0,5 y 2, respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a los genes sobreexpresados (puntos rojos) en amastigote. Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre amastigote y epimastigote.

Al igual que en la comparación anteriormente descrita, se analizó el número de genes con expresión génica diferencial estableciendo otros umbrales de cambio de expresión génica. Como se muestra en la **Tabla R17**, aun exigiendo una diferencia de expresión de 8 veces superior o inferior (correspondiente a un  $\log_2$  FC  $> 3$  o  $< -3$ , respectivamente), la cifra de genes que lo cumplían se mantenía elevada, incluyendo a 13.258 genes. Cabe destacar que los genes con mayores diferencias de expresión

## RESULTADOS

se encontraban en su mayoría infraexpresados en el estadio amastigote *versus* epimastigote. Más de 1.100 genes presentaban un valor  $\log_2 FC > 10$  o  $< -10$ , equivalente a 1.024 más expresión o una expresión de 1/1.024, respectivamente (Tabla R17).

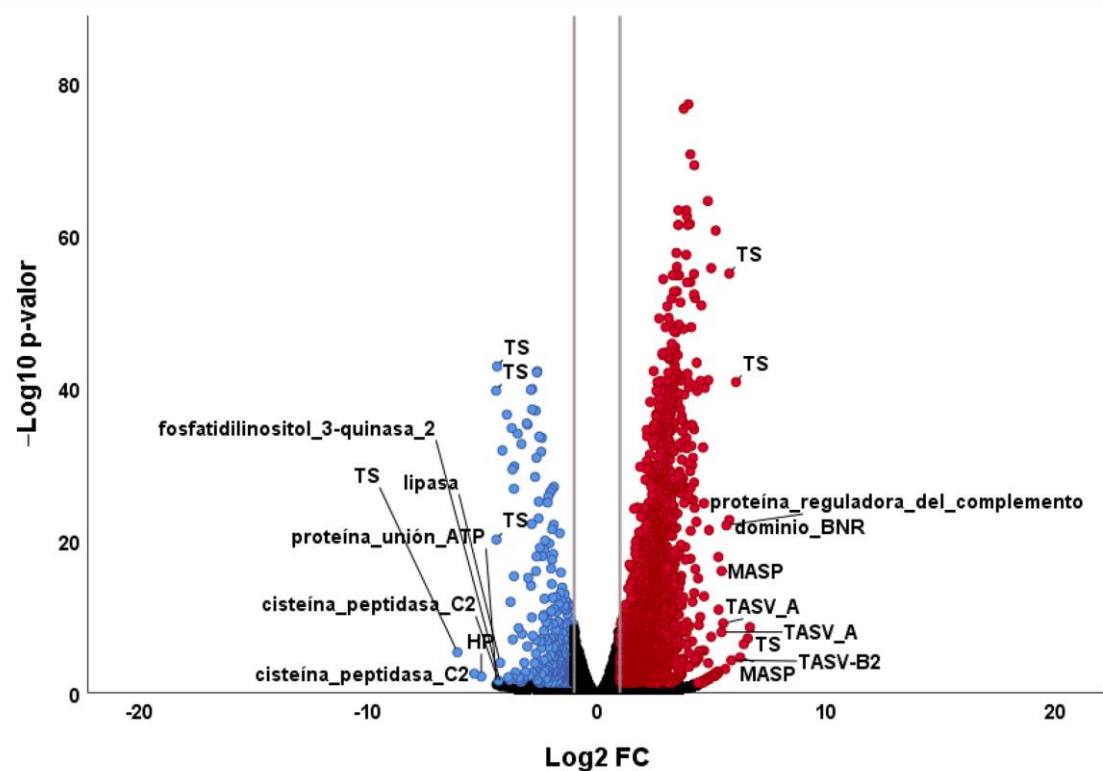
Tabla R17. Número de genes expresados diferencialmente (GED) en la forma amastigote de *T. cruzi* *versus* epimastigote determinado por ARN-seq. FC: fold change.

Amastigote <i>versus</i> Epimastigote	GED	Sobre-expresados	Infra-expresados	Expresión
$\log_2 FC > 1$ o $< -1$	17.981	8.982	8.999	FC = 2; FC = 1/2
$\log_2 FC > 3$ o $< -3$	13.258	6.331	6.927	FC = 8; FC = 1/8
$\log_2 FC > 5$ o $< -5$	8.874	3.849	5.025	FC = 32; FC = 1/32
$\log_2 FC > 6$ o $< -6$	6.779	2.585	4.194	FC = 64; FC = 1/64
$\log_2 FC > 7$ o $< -7$	4.545	1.206	3.339	FC = 128; FC = 1/128
$\log_2 FC > 10$ o $< -10$	1.177	55	1.122	FC = 1.024; FC = 1/1.024

### A.3.I.3. Expresión génica diferencial en tripomastigote *versus* amastigote

Las diferencias en el nivel de expresión génica entre los estadios tripomastigote y amastigote de *T. cruzi* basados en los valores RPKM detectados, mostraron que la expresión de 2.485 genes variaba de forma significativa entre ambas formas del parásito. Como puede observarse en el gráfico de volcán de la **Figura R11**, las diferencias de expresión fueron estadísticamente significativas ( $-\log_{10} p\text{-valor} > 1,3$ , equivalente a  $p < 0,05$ ). En particular, 2.024 de estos genes se sobreexpresaban en la forma tripomastigote *versus* amastigote con, al menos, doble nivel de expresión ( $\log_2 FC > 1$ ) y 461 se infraexpresaban en tripomastigote con, al menos, la mitad de expresión de la detectada en la forma amastigote ( $\log_2 FC < -1$ ).

Entre los genes que presentaron mayores diferencias de nivel de expresión entre los estadios tripomastigote y amastigote se observó que un número significativo de genes sobreexpresados en tripomastigote *versus* amastigote correspondía a genes codificantes de TS, MASP y TASV.



**Figura R11.** Representación en gráfico de volcán de los transcritos expresados diferencialmente entre los estadios tripomastigote y amastigote de *T. cruzi* determinado por ARN-seq. El eje X representa el  $\log_2$  del *fold change* de expresión entre tripomastigote y amastigote ( $\log_2$  FC), donde el FC se calcula como la ratio de la expresión (medida en RPKM) entre los dos estadios (tripomastigote/amastigote). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ajustado ( $-\log_{10}$  p-valor). Las líneas verticales grises representan el  $\log_2$  de FC de -1 y 1 (correspondiente a FC de 0,5 y 2, respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a los genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre tripomastigote y amastigote.

En la **Tabla R18** se muestra el número de genes que superan los diferentes umbrales de diferencias en el nivel de expresión establecidos. Entre tripomastigote y amastigote, el número de genes diferencialmente expresados con un  $\log_2$  FC  $> 3$  o  $< -3$  fue de 456, estando 420 de ellos sobreexpresados y 36 infraexpresados en tripomastigote respecto a amastigote. De estos genes, 27 presentaban valores de  $\log_2$  FC  $> 5$  o  $< -5$ , correspondientes a 32 veces más expresión (24 genes) o menor a 1/32 (3 genes) en la forma tripomastigote en comparación con amastigote. La mayor diferencia de expresión génica observada correspondía a valores de FC de 64 o 1/64, presentes en 6 genes, 5 sobreexpresados y 1 infraexpresado en tripomastigote al comparar con la expresión detectada en la forma amastigote (**Tabla R18**).



Tabla R18. Número de genes expresados diferencialmente (GED) en la forma tripomastigote de *T. cruzi* versus amastigote determinado por ARN-seq. FC: fold change.

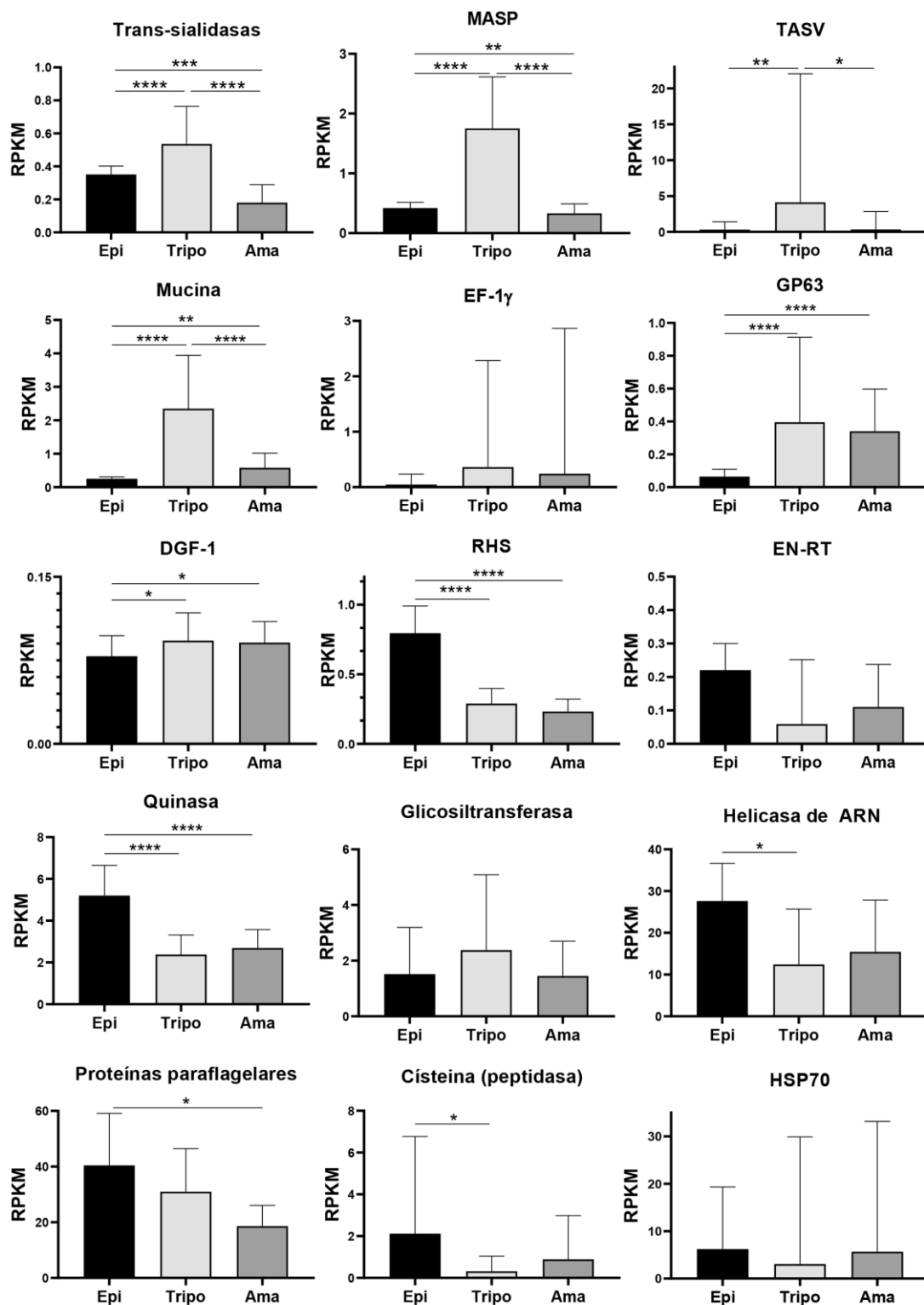
Tripomastigote versus Amastigote	GED	Sobre-expresados	Infra-expresados	Expresión
$\text{Log}_2 \text{FC} > 1 \text{ o } < -1$	2.485	2.024	461	FC = 2; FC = 1/2
$\text{Log}_2 \text{FC} > 3 \text{ o } < -3$	456	420	36	FC = 8; FC = 1/8
$\text{Log}_2 \text{FC} > 5 \text{ o } < -5$	27	24	3	FC = 32; FC = 1/32
$\text{Log}_2 \text{FC} > 6 \text{ o } < -6$	6	5	1	FC = 64; FC = 1/64

#### A.3.I.4. Diferencia en la expresión de genes particulares entre los 3 estadios morfológicos de *T. cruzi*

En un esfuerzo por analizar los resultados más relevantes, decidimos centrar nuestra atención en genes que conocemos son importantes para *T. cruzi*. Así, se analizó en cada estadio la expresión de genes relacionados con procesos de virulencia del parásito, así como de ciertos genes de expresión constitutiva. De esta forma, los valores RPKM, obtenidos del ARN-seq, de cada gen de interés se compararon entre estadios mediante la representación en gráficos de barras (**Figura R12**). Para los genes codificantes de trans-sialidasas (TS), mucinas y MASP, la expresión detectada fue superior en el estadio tripomastigote siendo las diferencias de expresión estadísticamente significativas respecto al nivel de expresión observado en amastigote y epimastigote ( $p < 0,0001$ ). También lo fueron las diferencias de expresión en estos genes entre epimastigotes y amastigotes ( $p < 0,001$  para TS y  $p < 0,01$  para mucinas y MASP). La expresión del gen GP63 fue también mayor en el estadio tripomastigote, siendo las diferencias en el nivel de expresión observadas estadísticamente significativas entre las formas tripomastigote y epimastigote y entre amastigote y epimastigote ( $p < 0,0001$ ). Respecto al gen codificante de TASV, el mayor nivel de expresión se identificó en el estadio tripomastigote con significancia estadística respecto a epimastigote ( $p < 0,01$ ) y amastigote y ( $p < 0,05$ ). Adicionalmente, se observaron diferencias estadísticamente significativas ( $p < 0,05$ ) en el nivel de expresión del gen DGF-1, sobreexpresado en tripomastigote y amastigote respecto al estadio epimastigote. Al igual que para los genes ya mencionados, el mayor nivel de expresión de los genes EF-1 $\gamma$  y glicosiltransferasa, fue detectado en el estadio tripomastigote, aunque las diferencias de expresión entre estadios no poseían significancia estadística. De forma contraria, el nivel de expresión de los genes codificantes para RHS,



endonucleasa-retrotranscriptasa (EN-RT), quinasa, helicasa de ARN, proteínas paraflagelares, cisteína peptidasa y HSP70 fue mayor en el estadio epimastigote. En el caso de EN-RT y HSP70, las diferencias de expresión no fueron estadísticamente significativas mientras que sí lo fueron las detectadas entre las formas epimastigote y tripomastigote para los genes codificantes de helicasas de ARN y cisteína peptidasa ( $p < 0,05$ ) y entre epimastigote y amastigote para los genes codificantes de proteínas paraflagelares ( $p < 0,05$ ). Por último, se observaron importantes diferencias, estadísticamente significativas, en el nivel de expresión de los genes RHS y quinasa entre los estadios epimastigote y tripomastigote ( $p < 0,0001$ ), así como entre epimastigote y amastigote ( $p < 0,0001$ ).



**Figura R12. Análisis comparativo del nivel de expresión de los genes indicados en las formas de *T. cruzi* epimastigote (Epi), tripomastigote (Tripo) y amastigote (Ama).** En los gráficos de barras se representa la mediana de la expresión medida en RPKM y el intervalo de confianza del 95%. Se indican las diferencias estadísticamente significativas determinadas por la prueba de Mann-Whitney (\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$  y \*\*\*\*  $p < 0,0001$ ). *MASP* (proteína de superficie asociada a mucina); *TASV* (Tripomastigote, Alanina, Serina, Valina); *EF-1γ* (Factor de elongación 1γ); *GP63* (metaloproteasa de zinc GP63); *DGF-1* (*Dispersed Gene Family protein 1*); *RHS* (*Retrotransposon Hot Spot*); *EN-RT* (Endonucleasa-retrotranscriptasa); *HSP70* (proteína de choque térmico de 70 kDa).

### A.3.II EXPRESIÓN DIFERENCIAL DE RETROELEMENTOS PORTADORES DE LA HUELLA Pr77 ENTRE LAS DISTINTAS FORMAS DE *T. cruzi* (CEPA SOL)

Para determinar la expresión de los retroelementos portadores de la huella Pr77 y analizar, en su caso, la existencia de expresión diferencial de los mismos asociada a la forma del parásito, se obtuvo el nivel de expresión de las regiones genómicas correspondientes a la huella Pr77 y a los retroelementos L1Tc, NARTc y DIREs obtenido en el ARN-seq de las formas epimastigote, tripomastigote y amastigote y, posteriormente, se realizaron análisis comparativos entre las diferentes formas del parásito.

Las comparaciones realizadas, representadas en los gráficos de volcán de la **Figura R13**, fueron tripomastigote *versus* epimastigote (**Figura R13A**), amastigote *versus* epimastigote (**Figura R13B**) y tripomastigote *versus* amastigote (**Figura R13C**). En el gráfico de volcán se representa el  $\log_2$  del *fold change*, calculado como la ratio entre los valores de expresión (en RPKM) de los dos grupos a comparar. Se consideró que aquellas secuencias con valores de  $\log_2$  FC superior a 1 o inferior a -1 se expresaban diferencialmente, puesto que correspondían a un nivel de expresión de más del doble o inferior a la mitad, respectivamente.

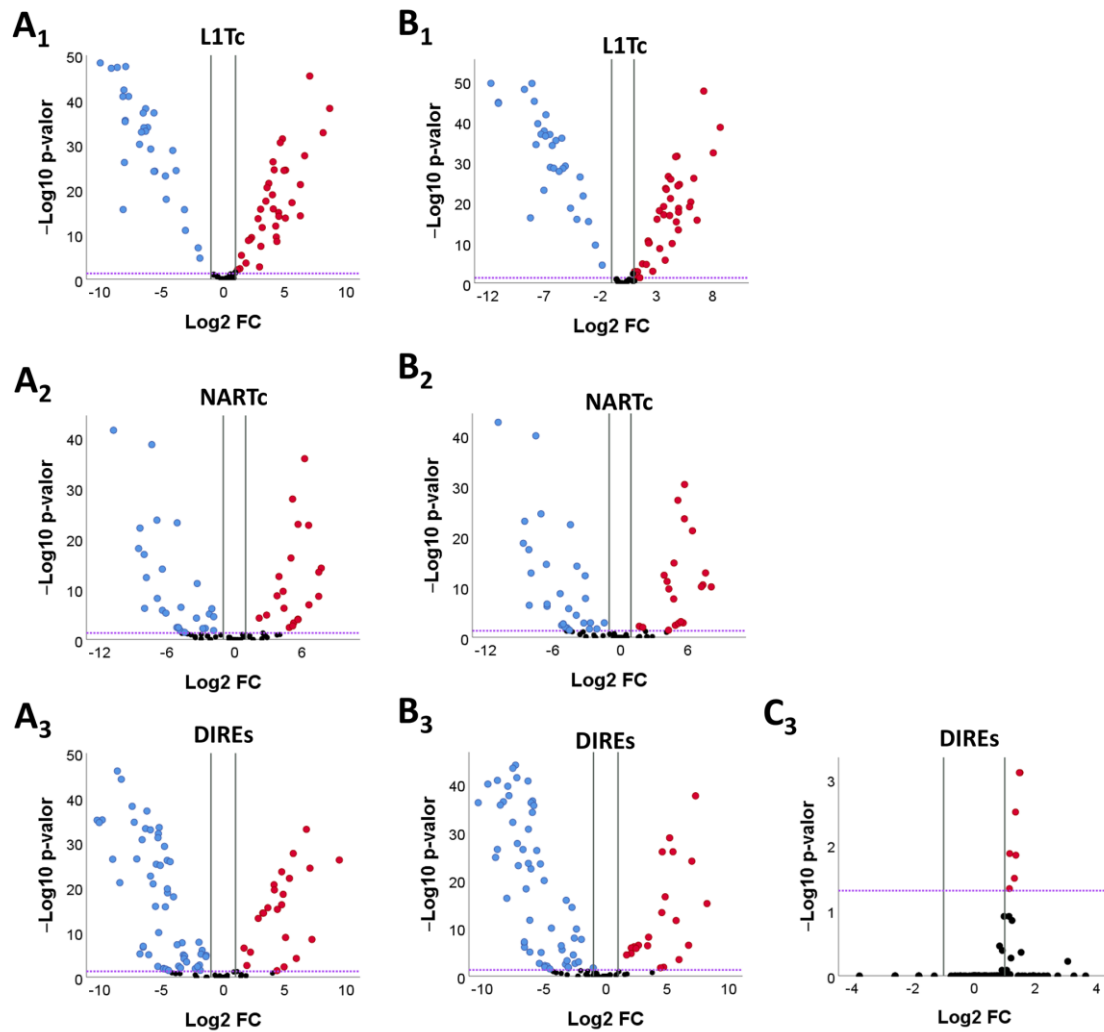
En el caso de los 78 retroelementos L1Tc identificados en el genoma, se observó que 66 de ellos se encontraban diferencialmente expresados con significancia estadística ( $FDR < 0,05$ ) en tripomastigote *versus* epimastigote (36 sobreexpresados y 30 infraexpresados). Estos mismos 66 estaban incluidos en los 68 retroelementos que se expresaban diferencialmente entre los estadios amastigote y epimastigote (38 sobreexpresados y 30 infraexpresados) (**Figura R13A<sub>1</sub> y B<sub>1</sub>**, **Tabla R19**).

Respecto a los retroelementos NARTc, de los 93 identificados en el genoma, se detectaron 47 con expresión diferencial entre los estadios tripomastigote y epimastigote (20 sobreexpresados y 27 infraexpresados) y 49 entre los estadios amastigote y epimastigote (20 sobreexpresados y 29 infraexpresados), siendo 46 de los retroelementos diferencialmente expresados comunes entre ambas comparaciones tripomastigote/amastigote *versus* epimastigote (**Figura R13A<sub>2</sub> y B<sub>2</sub>**, **Tabla R19**).

## RESULTADOS

---

Las diferencias de expresión de los elementos DIREs se representaron en un gráfico de volcán para cada comparación realizada (**Figura R13A<sub>3</sub>, B<sub>3</sub> y C<sub>3</sub>**). Los resultados revelaron que 79 DIREs mostraban expresión génica diferencial entre las formas tripomastigote y epimastigote con significancia estadística ( $FDR < 0,05$ ), encontrándose 23 sobreexpresados y 56 infraexpresados en la forma tripomastigote *versus* epimastigote. Del mismo modo, se hallaron 77 DIREs expresados diferencialmente al comparar la expresión detectada en los estadios amastigote *versus* epimastigote (21 sobreexpresados y 56 infraexpresados). 76 DIREs fueron comunes entre ambos análisis presentando expresión diferencial en ambas formas tripomastigote/amastigote respecto a epimastigote. Cuando se analizaron las diferencias de expresión de los DIREs entre los estadios tripomastigote y amastigote, se observó una sobreexpresión en la forma tripomastigote, con significancia estadística, de 7 DIREs, todos ellos identificados a partir de la sonda AAHK01000322.1 (**Figura R13C<sub>3</sub>, Tabla R19**).



**Figura R13.** Representación en gráfico de volcán de los retroelementos expresados diferencialmente entre los estadios (A) tripomastigote *versus* epimastigote, (B) amastigote *versus* epimastigote y (C) tripomastigote *versus* amastigote determinado por ARN-seq. El eje X representa el  $\log_2$  del *fold change* de expresión ( $\text{Log}_2 \text{FC}$ ), donde el FC se calcula como la ratio de la expresión (medida en RPKM) entre los dos estadios a comparar. El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ajustado ( $-\text{Log}_{10} p\text{-valor}$ ). Las líneas verticales grises representan el  $\log_2$  de FC de -1 y 1 (correspondiente a FC de 0,5 y 2, respectivamente). Los valores negativos corresponden a retroelementos infraexpresados (puntos azules) y los positivos a retroelementos sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan retroelementos cuya expresión no es diferencial entre los estadios comparados. Sondas empleadas para la identificación de los retroelementos en el genoma: L1Tc: L1Tc\_X83098; NARTc: NARTc\_AF215898; DIREs: AAHK01000664.1, AAHK01001018.1; AAHK01001032.1; AAHK01000322.1; AAHK01000335.1.

## RESULTADOS

**Tabla R19. Retroelementos portadores de Pr77 expresados diferencialmente en las tres formas del parásito identificados en el genoma de SOL.** En negrita se indica el número de retroelementos expresados diferencialmente que presentaban un valor de  $\log_2 FC > 1$  (sobreexpresados) o  $< -1$  (infraexpresados) y  $FDR < 0,05$ . Sondas empleadas para la identificación de los retroelementos en el genoma: *L1Tc*: L1Tc\_X83098; *NARTc*: NARTc\_AF215898; *DIREs*: AAHK01000664.1, AAHK01001018.1; AAHK01001032.1; AAHK01000322.1; AAHK01000335.1.

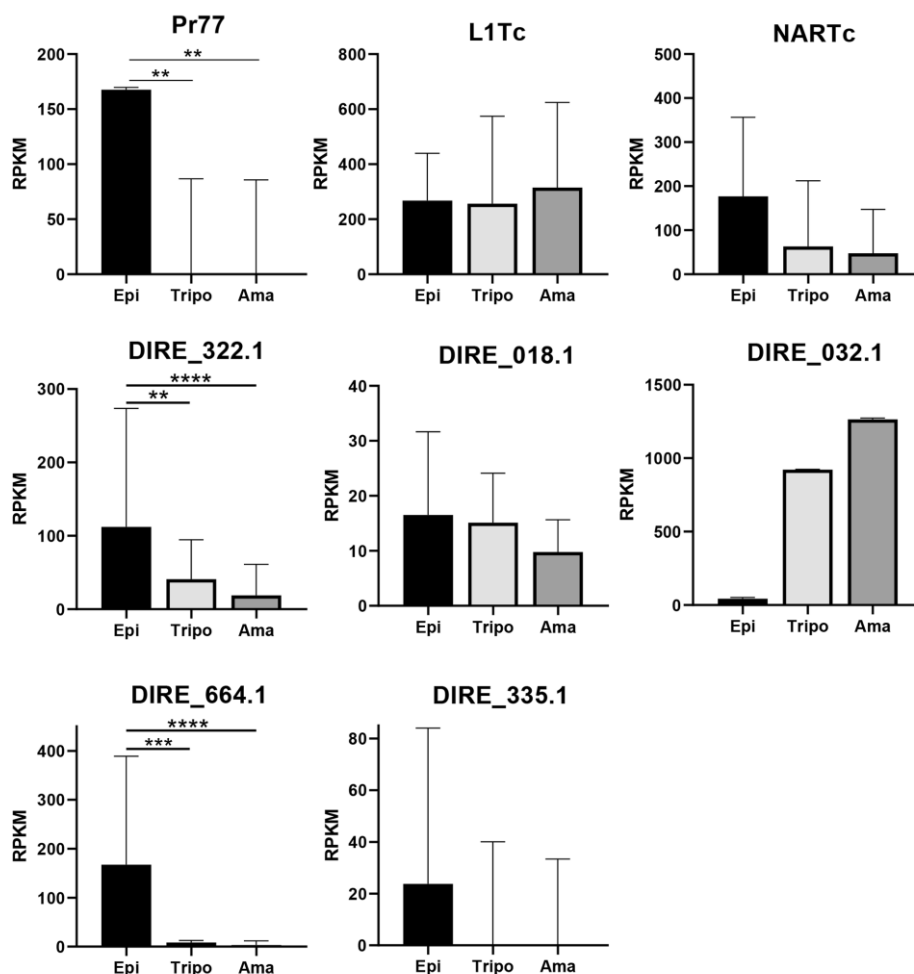
	Identificados en el genoma*	Tripomastigote versus Epimastigote		Amastigote versus Epimastigote		Tripomastigote versus Amastigote	
		Sobre-expresados	Infra-expresados	Sobre-expresados	Infra-expresados	Sobre-expresados	Infra-expresados
<b>L1Tc</b>	78	36	30	38	30		
		<b>66</b>		<b>68</b>			
<b>NARTc</b>	93	20	27	20	29		
		<b>47</b>		<b>49</b>			
<b>DIREs</b>	107	23	56	21	56	7	-
		<b>79</b>		<b>77</b>		<b>7</b>	

\* Con tamaño de alineamiento de, al menos, el 70% del tamaño del retroelemento a estudio.

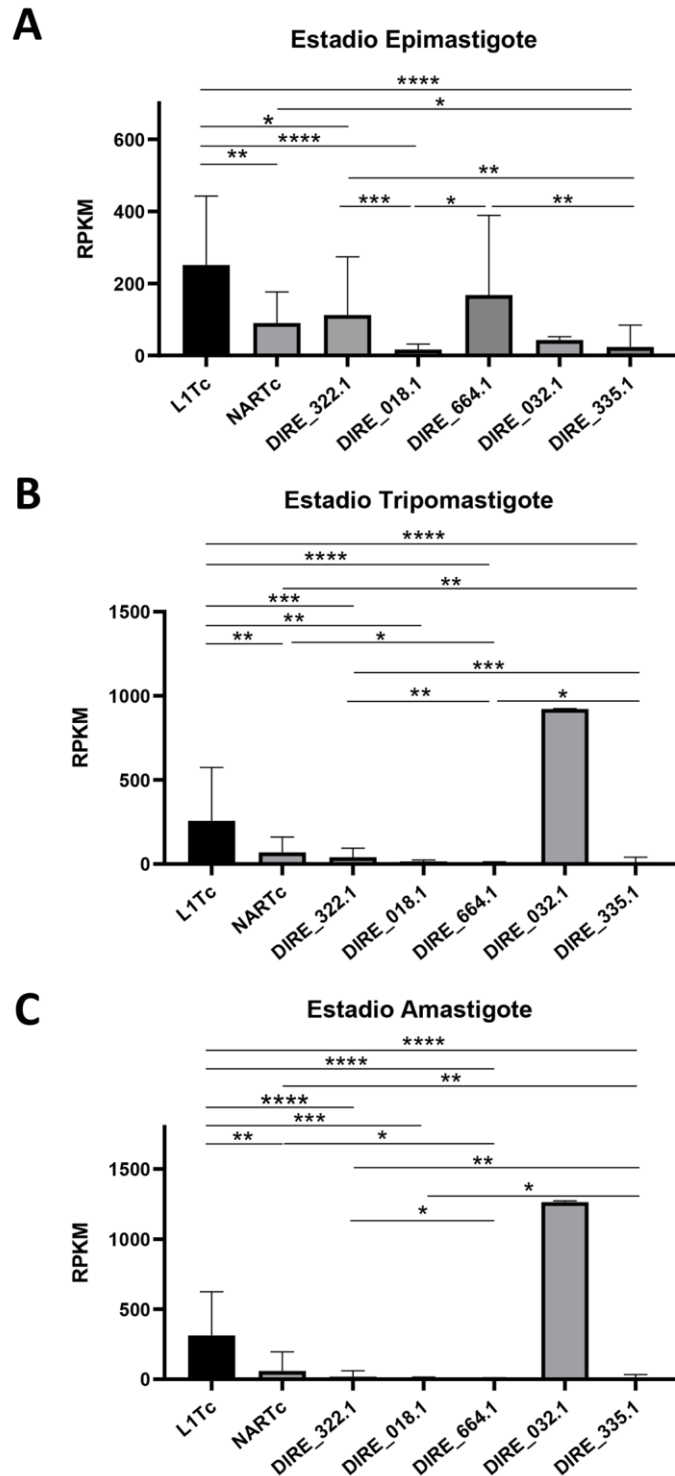
Con objeto de conocer en qué estadio de *T. cruzi* se observaba el mayor nivel de expresión de cada retroelemento, se compararon los valores de expresión (en RPKM) detectados para cada retroelemento en cada estadio tras la secuenciación ARN-seq. Del mismo modo, se analizó la expresión de las secuencias Pr77 identificadas en el genoma. El análisis, representado en la **Figura R14**, reveló que existía una mayor expresión de las secuencias Pr77 en el estadio epimastigote en comparación con los estadios tripomastigote y amastigote, siendo las diferencias estadísticamente significativas ( $p < 0,01$ ). El mismo perfil se observó en el elemento DIRE\_322.1 y DIRE\_664.1, detectándose el mayor nivel de expresión en epimastigote con diferencias estadísticamente significativas respecto a la expresión identificada en los estadios tripomastigote ( $p < 0,01$  para DIRE\_322.1 y  $p < 0,001$  para DIRE\_664.1) y amastigote ( $p < 0,0001$  para ambos). En el caso de los elementos NARTc, L1Tc y el resto de DIREs identificados en el genoma, las diferencias en el nivel de expresión observadas entre las distintas formas del parásito no fueron estadísticamente significativas (**Figura R14**).

Por otra parte, se representó en un mismo gráfico la expresión de todos los retroelementos analizados para cada estadio individualmente. Esto permitió el análisis comparativo de la expresión de todos ellos en cada estadio de *T. cruzi*. Los resultados, representados en la **Figura R15**, revelaron que en la forma epimastigote,

las secuencias con mayor nivel de expresión fueron aquellas que correspondían al retroelemento L1Tc mostrando significancia estadística al comparar con la expresión de la mayoría de retroelementos a estudio (**Figura R15A**). En las formas tripomastigote y amastigote, el mayor nivel de expresión se detectó para el elemento DIRE\_032.1, seguido de L1Tc y, a continuación, NARTc (**Figura R15B y R15C**). La significancia estadística de las distintas comparaciones realizadas se indica en cada gráfico (**Figura R15**).



**Figura R14. Análisis comparativo del nivel de expresión de retroelementos identificados en el genoma de SOL en las 3 formas de *T. cruzi* epimastigote (Epi), tripomastigote (Tripo) y amastigote (Ama).** Sondas empleadas para la identificación de los retroelementos en el genoma: Pr77: Secuencia Pr77 de L1Tc; L1Tc: L1Tc\_X83098; NARTc: NARTc\_AF215898; DIREs: AAHK01000664.1, AAHK01001018.1; AAHK01001032.1; AAHK01000322.1; AAHK01000335.1. Se representa la mediana de la expresión medida en RPKM y el intervalo de confianza del 95%. Se indican las diferencias estadísticamente significativas determinadas por la prueba de Mann-Whitney (\*\*  $p < 0,01$ , \*\*\*  $p < 0,001$  y \*\*\*\*  $p < 0,0001$ ).



**Figura R15. Análisis comparativo del nivel de expresión de retroelementos identificados en el genoma de SOL en el estadio epimastigote (A), tripomastigote (B) y amastigote (C).** Sondas empleadas para la identificación de los retroelementos en el genoma: *Pr77*: Secuencia *Pr77* de *L1Tc*; *L1Tc*: *L1Tc\_X83098*; *NARTc*: *NARTc\_AF215898*; *DIREs*: *AAHK01000664.1*, *AAHK01001018.1*; *AAHK01001032.1*; *AAHK01000322.1*; *AAHK01000335.1*. Se representa la mediana de la expresión medida en RPKM y el intervalo de confianza del 95%. Se indican las diferencias estadísticamente significativas determinadas por la prueba de Mann-Whitney (\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$  y \*\*\*\*  $p < 0,0001$ ).



## B. BIOMARCADORES DE RESPUESTA INMUNOLÓGICA

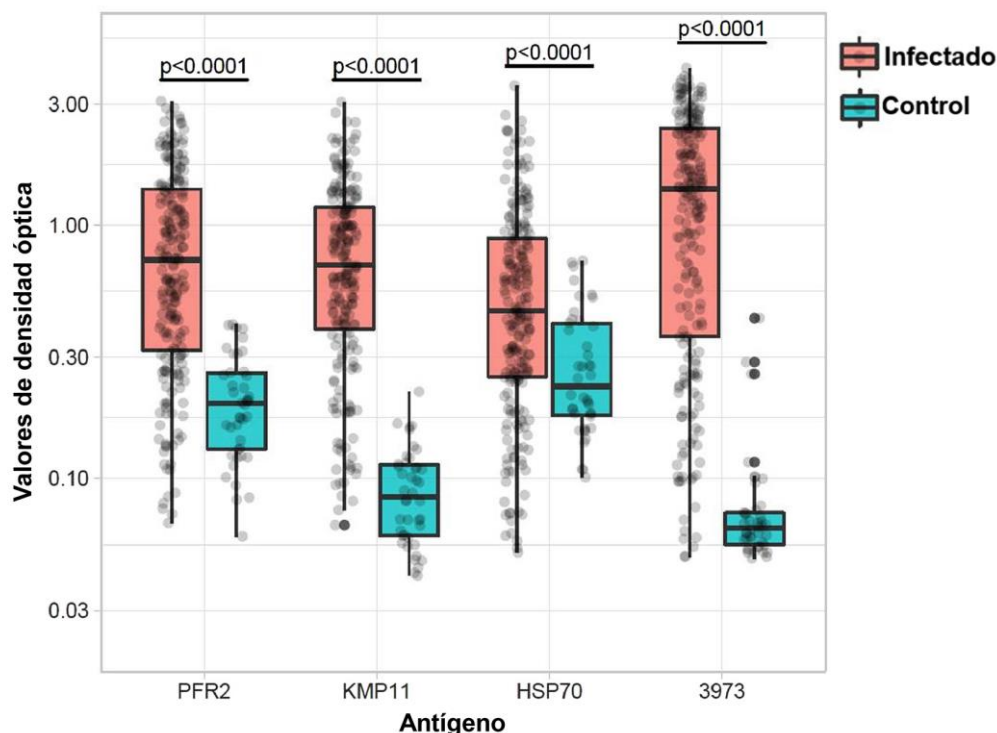
### B.1. RESPUESTA SEROLÓGICA

#### B.1.1. ANÁLISIS DE LA SENSIBILIDAD Y ESPECIFICIDAD DE LOS ANTÍGENOS KMP11, HSP70, PFR2 Y 3973<sup>d</sup> DE *T. cruzi* PARA LA DETECCIÓN DE INFECCIÓN EN PACIENTES CON ENFERMEDAD DE CHAGAS

Formando parte de un estudio a ciegas retrospectivo multicéntrico en el que se evaluaron distintos antígenos de *T. cruzi* (Alonso-Padilla et al., 2021), se determinó la sensibilidad y especificidad de los antígenos KMP11, HSP70, PFR2 y péptido 3973<sup>d</sup> de *T. cruzi* (caracterizados previamente en el laboratorio), midiendo la reactividad presente en el suero de pacientes con enfermedad de Chagas en fase indeterminada (IND).

El nivel de anticuerpos antígeno-específicos se determinó mediante el empleo de la técnica de ELISA en los pacientes con enfermedad de Chagas, así como en los sujetos sanos, tal y como se describe en la sección de *Métodos 4.4*. Los valores de reactividad, basados en la densidad óptica (D.O.), fueron claramente diferentes, presentando los sujetos infectados un nivel de anticuerpos frente a los mencionados antígenos estadísticamente significativo en comparación con el detectado en los donantes sanos (DS), con un valor  $p < 0,0001$  (**Figura R16**).

Como se muestra en la **Tabla R20**, los antígenos fueron reconocidos con una alta sensibilidad y especificidad por el suero de pacientes con enfermedad de Chagas. En concreto, KMP11 obtuvo una eficacia diagnóstica del 92%, con una sensibilidad (Se) superior al 90%. Para este mismo antígeno, la especificidad (Sp) fue del 98%. Respecto a la tasa de falsos positivos o falsos negativos, KMP11 mantuvo los valores por debajo del 10% debido a sus altos parámetros Se y Sp registrados. El péptido 3973<sup>d</sup> fue el segundo biomarcador mejor posicionado, tras KMP11, por su eficacia diagnóstica, parámetro para el que adquirió un valor del 85%. La especificidad (Sp) para HSP70 fue del 83%, y del 93% y 95% para los antígenos 3973<sup>d</sup> y PFR2, respectivamente.



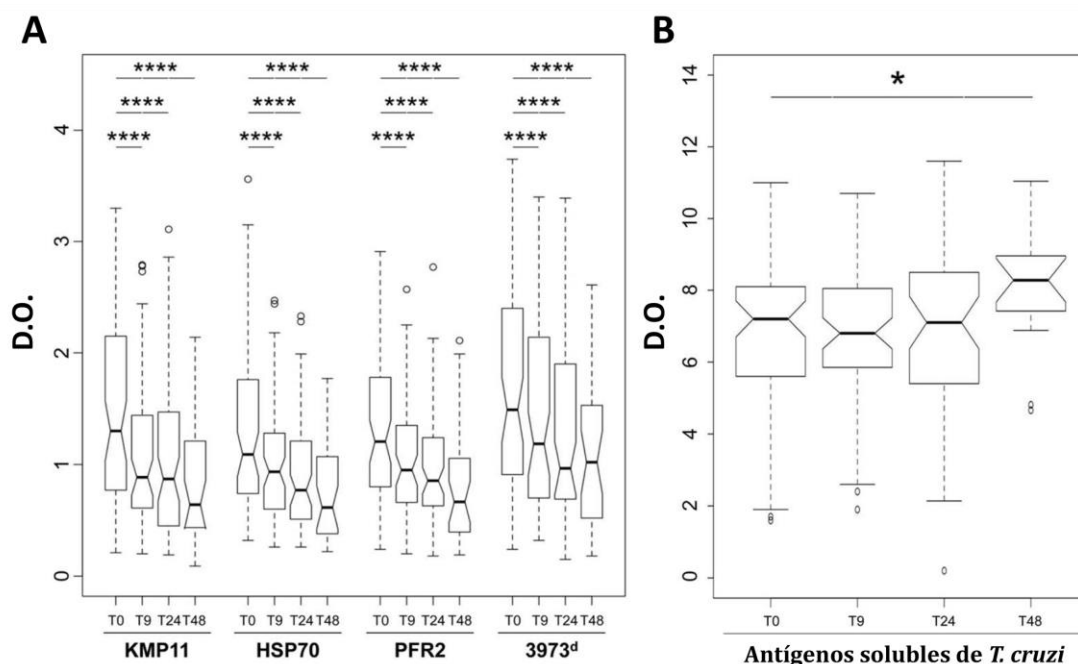
**Figura R16.** Diagramas de cajas que muestran la reactividad de los sueros de sujetos infectados con *T. cruzi* y sujetos sanos (valores de densidad óptica, representados en escala logarítmica) frente a cada uno de los antígenos. Cada punto representa una sola muestra. Las medianas se compararon entre grupos empleando la prueba de suma de rangos de Wilcoxon. El número de muestras fue 42 sujetos control para todos los antígenos y, al menos, 211 sujetos infectados: 217, 211, 217 y 219 para PFR2, KMP11, HSP70 y 3973<sup>d</sup>. Las diferencias se debieron a la disponibilidad de las muestras. *Adaptada de Alonso-Padilla et al., 2021.*

**Tabla R20. Resumen de la clasificación de cada antígeno evaluado.** *DE:* eficacia diagnóstica; *Se:* sensibilidad; *Sp:* especificidad; *FPR:* tasa de falsos positivos; *FNR:* tasa de falsos negativos. Todos los valores se dan en % y los  $\geq 85\%$  están resaltados en negrita. *Adaptada de Alonso-Padilla et al., 2021.*

Parámetro	PFR2	KMP11	HSP70	3973 <sup>d</sup>
N.º Sujetos	217	211	217	219
ED	74	<b>92</b>	55	<b>85</b>
Se	70	<b>91</b>	49	84
Sp	<b>95</b>	<b>98</b>	83	<b>93</b>
FPR	5	2	17	7
FNR	30	9	51	16

### **B.1.2. EVALUACIÓN DEL SET DE ANTÍGENOS KMP11, HSP70, PFR2 Y 3973<sup>d</sup> DE *T. cruzi* COMO BIOMARCADORES DE EFICACIA TERAPÉUTICA EN PACIENTES CON ENFERMEDAD DE CHAGAS. ESTABLECIMIENTO DEL ALGORITMO PARA EVALUACIÓN DE LA EFICACIA TERAPÉUTICA**

Con el fin de evaluar la utilidad de los biomarcadores KMP11, HSP70, PFR2 y 3973<sup>d</sup> como herramienta para determinar la eficacia terapéutica, se analizó la reactividad del suero de 66 pacientes con enfermedad de Chagas en fase crónica indeterminada (IND) frente al mencionado set de antígenos, antes del tratamiento con benznidazol (T0) y a los 9 (T9), 24 (T24) y 48 (T48) meses después del tratamiento. En la **Figura R17** se muestran los títulos de anticuerpos (valores de D.O.) específicos frente a las 4 moléculas detectados en el suero de los pacientes IND mediante ensayo ELISA. Estos fueron elevados para los 4 antígenos en la mayoría de los pacientes a T0 y disminuyeron progresivamente durante el seguimiento post-tratamiento (**Figura R17A**). Esta disminución en la reactividad frente a KMP11, HSP70, PFR2 y 3973<sup>d</sup>, medida como la caída del nivel de anticuerpos frente a cada molécula y en cada paciente, fue estadísticamente significativa a T9 ( $p < 0,0001$ ), T24 ( $p < 0,0001$ ) y T48 ( $p < 0,001$ ) al comparar con el valor obtenido pre-tratamiento (T0) aplicando la prueba de Friedman seguida de la prueba Post Hoc de Dunn para comparaciones múltiples. Por el contrario, el nivel de anticuerpos frente antígenos totales del parásito (STcA) se mantuvo invariable a lo largo del tiempo tras el tratamiento (**Figura R17B**). De hecho, se detectó un aumento estadísticamente significativo de los niveles de anticuerpos frente a STcA a largo plazo tras el tratamiento (T48) respecto a los valores de D.O. detectados a T0 en estos pacientes (**Figura R17B**).



**Figura R17. Reactividad de los sueros de pacientes asintomáticos de la enfermedad de Chagas frente a antígenos de *T. cruzi* antes y tras el tratamiento con benznidazol. (A)** Los niveles de anticuerpos (IgG) frente a KMP11, HSP70, PFR2 y 3973<sup>d</sup> se midieron por ELISA en los sueros de 66 pacientes indeterminados (IND) antes del tratamiento con benznidazol (T0) y a los 9 (T9), 24 (T24) meses tras tratamiento y en 32 pacientes a 48 (T48) meses tras el tratamiento. Los sueros se analizaron siempre por triplicado y a una dilución de 1/100, 1/200, 1/400, 1/800 y 1/1600 para cada una de las cuatro moléculas en T0. Para el seguimiento posterior al tratamiento, el suero de cada paciente se analizó a dos diluciones, que se seleccionaron en función de la reactividad observada en T0. Los sueros de los 66 pacientes se analizaron a diferentes diluciones frente a cada antígeno (1/100 [5 pacientes analizados para KMP11, 1 para HSP70 y 2 para 3973<sup>d</sup>], 1/200 [13 pacientes analizados para KMP11, 4 para HSP70, 5 para PFR2 y 5 para 3973<sup>d</sup>], 1/400 [41 pacientes analizados para KMP11, 13 para HSP70, 4 para PFR2 y 1 para 3973<sup>d</sup>], 1/800 [7 pacientes analizados para KMP11, 48 para HSP70, 55 para PFR2 y 54 para 3973<sup>d</sup>], y 1/1600 [2 pacientes analizados para PFR2 y 4 para 3973<sup>d</sup>]). Se indican las diferencias estadísticamente significativas (\*\*\*,  $p < 0,0001$  y \*\*\*\*,  $p < 0,0001$ ) analizadas empleando la prueba de Friedman y la prueba post hoc de Dunn. **(B)** Detección de los niveles de anticuerpos IgG frente a los antígenos solubles de *T. cruzi* (STcA) en los pacientes IND antes (T0) y a los 9, 24 y 48 meses tras el tratamiento. Se analizó la reactividad de los sueros frente a STcA en 62 pacientes IND a T0, 60 a T9, 45 a T24 y 16 a T48. Los datos de ambos gráficos se expresan como la densidad óptica (D.O.) medida a 492 nm. Las muescas se extienden para 1,58 x rango intercuartílico (IQR)/raíz cuadrada de todos los valores, ya que esto da un intervalo de confianza del 95% para comparar las medianas. En este gráfico, los valores atípicos son las observaciones que se encuentran fuera de 1,58 x IQR, donde IQR es la diferencia entre los cuartiles 75 y 25. Los valores  $p$  se obtuvieron mediante la prueba de Wilcoxon, indicándose las diferencias estadísticamente significativas (\*,  $p < 0,05$ ). Adaptada de Egui et al., 2019.

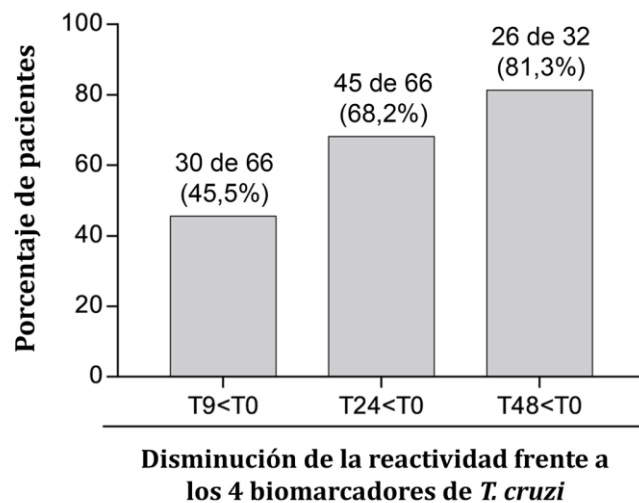
Tal y como se muestra en la **Tabla R21**, el porcentaje de pacientes en los cuales existía una disminución en el nivel de anticuerpos frente a cada uno de los biomarcadores (BMKs) variaba de 76% a 82% a los 9 meses post-tratamiento (T9), de 86% a 89% a los 24 meses (T24) y del 91% al 97% a los 48 meses (T48). Cuando se evaluó la reactividad de los sueros de los pacientes frente a los 4 BMKs conjuntamente, el porcentaje de pacientes en los que la reactividad disminuía

respecto a T0 fue del 45,5% a los 9 meses, 68,2% a los 24 meses y 81,3% a los 48 meses tras el tratamiento (**Figura R18**).

**Tabla R21. Porcentaje de pacientes con enfermedad de Chagas que presentaban una disminución en el nivel de anticuerpo frente a los antígenos de *T. cruzi* tras el tratamiento con benznidazol.** *Adaptada de Egui et al., 2019.*

Intervalos de tiempo (caída de la reactividad)	Biomarcadores de <i>T. cruzi</i> *			
	KMP11	HSP70	3973 <sup>d</sup>	PFR2
<b>T9 &lt; T0</b>	54 de 66 (82%)	53 de 66 (80%)	50 de 66 (76%)	52 de 66 (79%)
<b>T24 &lt; T0</b>	59 de 66 (89%)	57 de 66 (86%)	58 de 66 (88%)	57 de 66 (86%)
<b>T48 &lt; T0</b>	31 de 32 (97%)	29 de 32 (91%)	31 de 32 (97%)	29 de 32 (91%)

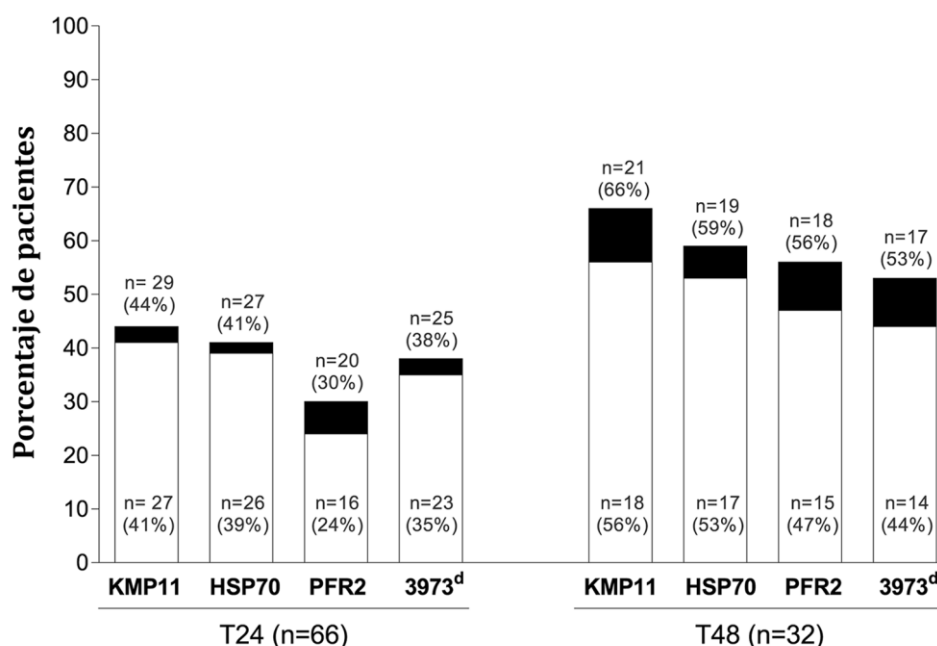
\* La reactividad se midió frente a las moléculas KMP11, HSP70, PFR2 y 3973<sup>d</sup>. Los números entre paréntesis representan el porcentaje de pacientes IND que mostraron un nivel de anticuerpos menor a los 9 (T9), 24 (T24) y 48 (T48) meses después del tratamiento con benznidazol en comparación con el tiempo pre-tratamiento (T0). Los intervalos analizados fueron de T0 a T9, de T0 a 24 y de T0 a T48. El tamaño de la población de estudio fue de 66 IND en T0, T9 y T24; y 32 pacientes IND en T48.



**Figura R18. Frecuencia de pacientes con enfermedad de Chagas que mostraron una disminución de los niveles de anticuerpos frente a los cuatro antígenos de *T. cruzi* evaluados después del tratamiento.** La reactividad se midió frente a las moléculas KMP11, HSP70, PFR2 y 3973<sup>d</sup>. Las barras representan los porcentajes de pacientes IND que mostraron menores niveles de anticuerpos frente a los cuatro biomarcadores a los 9 (T9), 24 (T24) y 48 (T48) meses tras el tratamiento con benznidazol en relación al punto de tiempo previo al tratamiento (T0). Los intervalos analizados fueron T0-T9, T0-24 y T0-T48. El tamaño de la población del estudio fue de 66 pacientes IND en T0, T9 y T24; y 32 pacientes en T48. *Adaptada de Egui et al., 2019.*

**- Establecimiento de criterio de eficacia terapéutica y análisis del comportamiento del set de BMKs serológicos**

Con objeto de identificar patrones serológicos con los que evaluar el impacto del tratamiento o predecir su efecto, se estableció un criterio de eficacia terapéutica estándar (STEC) basado en una caída continua de la reactividad frente a los cuatro BMKs a lo largo del tiempo y que la misma fuese notable (un 40% para KMP11, PFR2 y 3973<sup>d</sup> y un 30% para HSP70), al menos, para dos de los antígenos. Así, se observó que un 41%, 39%, 24% y 35% de los pacientes mostraban una caída en la reactividad frente a KMP11, HSP70, PFR2 y 3973<sup>d</sup>, respectivamente, a los 24 meses tras el tratamiento y un 56%, 53%, 47% y 44% de los pacientes a los 48 meses (Figura R19, barras blancas). Estos porcentajes fueron superiores si se incluían aquellos pacientes que, si bien presentaban una importante disminución a T24 y T48, mostraban un ligero incremento en la reactividad a 9 meses (inferior al 20%), el cual podría estar motivado por la lisis del parásito por efecto del benznidazol y la correspondiente liberación al torrente sanguíneo de antígenos del parásito (Figura R19, porción negra de las barras).



**Figura R19. Número y porcentaje de pacientes con enfermedad de Chagas cuya reactividad frente a cada una de las cuatro moléculas cumplía con el criterio STEC a los 24 y 48 meses tras el tratamiento.** La reactividad se midió frente a las moléculas KMP11, HSP70, PFR2 y 3973<sup>d</sup> antes (T0) y 9 (T9), 24 (T24) y 48 (T48) meses tras el tratamiento. Las barras blancas representan el número y porcentaje de pacientes IND cuyos sueros reaccionan frente a cada uno de los cuatro BMKs ensayados cumpliendo las siguientes condiciones: T9 < T0, T24/T48 < T9 y T24/T48 < T0, al menos, un 40% para KMP11, PFR2 y 3973<sup>d</sup> o un 30% para HSP70. La parte negra de las barras incluye, además, a los pacientes que no presentan cambios o un ligero aumento de la reactividad (por debajo del 20%) a los 9 meses tras el tratamiento (T9 ≥ T0). Adaptada de Egui et al., 2019.

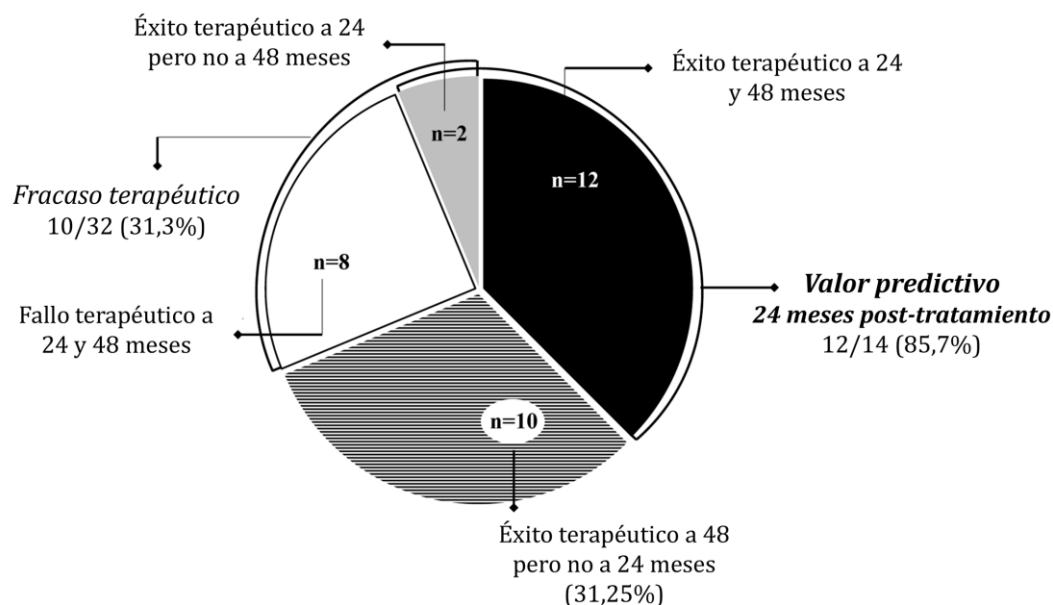
Con el fin de medir el valor predictivo del set de BMKs, además del criterio STEC ya descrito, se estableció una segunda condición de eficacia terapéutica más estricta (RTEC) en la que, además de la caída continua y constante en la reactividad frente a los 4 BMKs, se exigía que la reducción en la reactividad fuese > 40% para KMP11, PFR2 y 3973<sup>d</sup> y > 30% para HSP70) para, al menos, tres de los cuatro antígenos. Estos criterios se aplicaron mediante el empleo de un algoritmo sobre los valores de reactividad (en D.O.) detectados en los sueros de los pacientes frente a los 4 BMKs. Los resultados mostrados en la **Tabla R22** revelaron que un 42,4% y 68,75% de los pacientes IND obedecían el criterio de eficacia terapéutica (STEC) a los 24 y 48 meses tras tratamiento, respectivamente. Cuando se aplicó el criterio más estricto (RTEC), el porcentaje de pacientes que lo cumplieron fue de un 15,2% a 24 meses. A 48 meses tras tratamiento, un 40,6% de los pacientes seguían este patrón, siendo este porcentaje similar al obtenido a los 24 meses con el criterio STEC.

**Tabla R22. Pacientes con enfermedad de Chagas que cumplen con el criterio de eficacia terapéutica establecido.** *Adaptada de Egui et al., 2019.*

Criterio de Eficacia Terapéutica	A 24 meses	A 48 meses
STEC*	28 de 66	22 de 32
<i>(Standard therapeutic efficacy condition)</i>	(42,4%)	(68,75%)
RTEC**	10 de 66	13 de 32
<i>(Restrictive therapeutic efficacy condition)</i>	(15,2%)	(40,6%)

\* Cuatro BMK cumplen  $T_{24}/T_{48} < T_0$  y dos BMK obedecen  $T_{24}/48 < T_0$  (una disminución del 40% para KMP11, PFR2 y 3973<sup>d</sup> y 30% para HSP70);  $T_{24}/48 < T_9$  y  $T_9 < T_0$  o  $T_9 \geq T_0$  (por debajo del 20%).  
 \*\* Cuatro BMK cumplen  $T_{24}/T_{48} < T_0$  y tres de ellos obedecen  $T_{24}/48 < T_0$  (una disminución del 40% para KMP11, PFR2 y 3973<sup>d</sup> y 30% para HSP70);  $T_{24}/48 < T_9$  y  $T_9 < T_0$  o  $T_9 \geq T_0$  (por debajo del 20%).

Se calculó el valor predictivo del criterio STEC a los 24 meses del tratamiento a partir de los datos obtenidos de 32 pacientes IND evaluados, tanto a los 24 como a los 48 meses tras el tratamiento (**Figura R20**). Así, 14 pacientes cumplieron los criterios de eficacia terapéutica a 24 meses, superándolo 12 de ellos también a 48 meses, por lo que el valor predictivo, en términos de eficacia terapéutica, a 24 meses después del tratamiento fue del 85,7% (**Figura R20**). Por otro lado, dos pacientes mostraron éxito terapéutico a los 24 meses, pero no a los 48 meses y ocho pacientes no cumplieron el STEC ni a los 24 ni a los 48 meses, lo que indicó una tasa de potencial fracaso terapéutico del 31,3% (10 de 32 pacientes) (**Figura R20**).

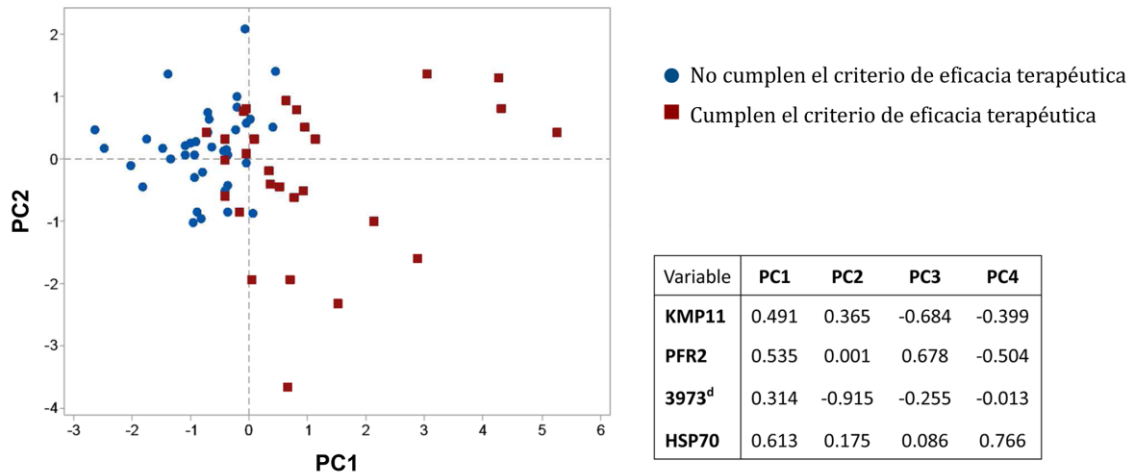


Pacientes en seguimiento 24 y 48 meses tras el tratamiento con benznidazol (n=32)

**Figura R20. Representación en forma de gráfico circular del porcentaje de pacientes con enfermedad de Chagas que cumplieron o no el criterio de eficacia terapéutica establecido.** Se evaluaron 32 pacientes crónicos en fase indeterminada (IND). Se representa a los pacientes que mostraron éxito terapéutico a los 24 (T24) y 48 (T48) meses del tratamiento (porción negra, n=12), pacientes que cumplieron el criterio de eficacia terapéutica STEC a T24, pero no a T48 tras el tratamiento (porción gris, n=2), pacientes con éxito terapéutico a T48, pero no a T24 después del tratamiento (porción rayada, n=10) y pacientes que no cumplieron las condiciones de eficacia terapéutica (porción blanca, n=8). Los arcos negros del gráfico indican el valor predictivo de la eficacia terapéutica a los 24 meses del tratamiento y el fracaso terapéutico. *Adaptada de Egui et al., 2019.*

La contribución de cada BMK a la varianza que se observó en la reactividad de los pacientes tras el tratamiento se evaluó mediante un análisis de componentes principales (PCA) entre los pacientes que cumplieran el criterio STEC y los que no lo hacían (**Figura R21**). Cuando se representó la componente principal 1 (PC1) y la componente principal 2 (PC2), las cuales explican las mayores varianzas, se observaron diferencias entre los pacientes que cumplieran el criterio de eficacia terapéutica estándar (STEC) y los que no lo cumplieran para la PC1 (**Figura R21**). La PC1, que representa el 53% de la varianza de los datos, se asoció significativamente con las cuatro variables (KMP11, PFR2, 3973<sup>d</sup> y HSP70). Por otro lado, PC2, PC3 y PC4, que representaron el 22%, 17% y 7% de la varianza de los datos, respectivamente, se asociaron con tres de las cuatro variables, KMP11, 3973<sup>d</sup> y HSP70.





**Figura R21.** Gráfico de puntuación PCA para las diferencias en la reactividad (varianza) de 66 pacientes en fase indeterminada frente a cada uno de los cuatro biomarcadores (4 variables) entre el tiempo previo al tratamiento y tras 24 meses de tratamiento. Los puntos rojos corresponden a los pacientes que cumplieron el criterio STEC, y los azules a los que no lo cumplieron. La tabla incluida en la figura contiene la carga factorial de cada variable para las cuatro componentes principales (PC1, PC2, PC3 y PC4). *Adaptada de Egui et al., 2019.*

Adicionalmente, se analizó la presencia de *T. cruzi* en sangre periférica de 33 de los 66 pacientes a estudio. Ninguno de los pacientes que cumplían el criterio STEC presentaba PCR positiva a 9, 24 y 48 meses tras el tratamiento. Por el contrario, la presencia del parásito sí fue detectada por PCR en 4 pacientes que no obedecían el criterio de eficacia terapéutica establecido (**Tabla R23**).

**Tabla R23.** Detección de *T. cruzi* por PCR en pacientes en fase indeterminada antes y tras el tratamiento con benznidazol. Los números entre paréntesis representan el porcentaje de pacientes IND que presentaron una PCR positiva. *Adaptada de Egui et al., 2019.*

	Antes del tratamiento	Tras el tratamiento	Tiempo de tratamiento (meses)		
			9	24	48
<b>Pacientes con PCR positiva (%)</b>	18 de 33 (54,5%)	0 de 33 (0%)	2 de 33 (6,1%)	1 de 33 (3%)	1 de 19 (5,3%)

### B.2. RESPUESTA CELULAR

#### B.2.1. EXPRESIÓN DIFERENCIAL DE GENES DE RESPUESTA INMUNE EN PACIENTES ASINTOMÁTICOS CON ENFERMEDAD DE CHAGAS *VERSUS* SUJETOS SANOS

Para mejorar el conocimiento de la respuesta inmune antígeno-específica generada durante la infección en pacientes con enfermedad de Chagas, se ha analizado en estos pacientes el patrón de expresión de genes particulares involucrados en la respuesta inmune generada tras la infección con *Trypanosoma cruzi*. Así, se ha determinado el nivel de expresión de 106 genes relacionados con el sistema inmunitario en respuesta a proteínas del parásito en células PBMC de pacientes con enfermedad de Chagas en la fase indeterminada de la enfermedad (IND, n=39). Así mismo, se han incluido células de donantes sanos procedentes de áreas endémicas (DSe, n=10) y no endémicas (DSc, n=20) de la enfermedad de Chagas.

Las PBMC de ambos grupos se estimularon con antígenos solubles de *T. cruzi*, para seguidamente obtener su ARN total. En todas las muestras incluidas en este estudio el ARN extraído tuvo valores de RIN (*RNA integrity number*) entre 7,4 y 10. A continuación, cada muestra de ARN fue sometida a retrotranscripción y amplificación por RT-qPCR de alto rendimiento, como se detalla en las secciones de *Métodos 5.4* y *5.6*. Los datos procedentes de las RT-qPCR se procesaron y analizaron siguiendo los protocolos descritos en la sección *Métodos 5.7*.

Durante el procesamiento, se llevó a cabo la normalización del conjunto de datos, evaluando la estabilidad de la expresión de los genes incluidos. Los genes *STAT3*, *IL10RA* e *IFNAR* mostraron los valores de expresión más estables con los diferentes algoritmos aplicados (con M-valor en geNorm < 0,5) empleándose como genes de referencia para normalizar y obtener los valores de expresión génica relativa normalizada (NRQ). Estos valores fueron los empleados para llevar a cabo los diferentes análisis comparativos que se detallan en los siguientes apartados.

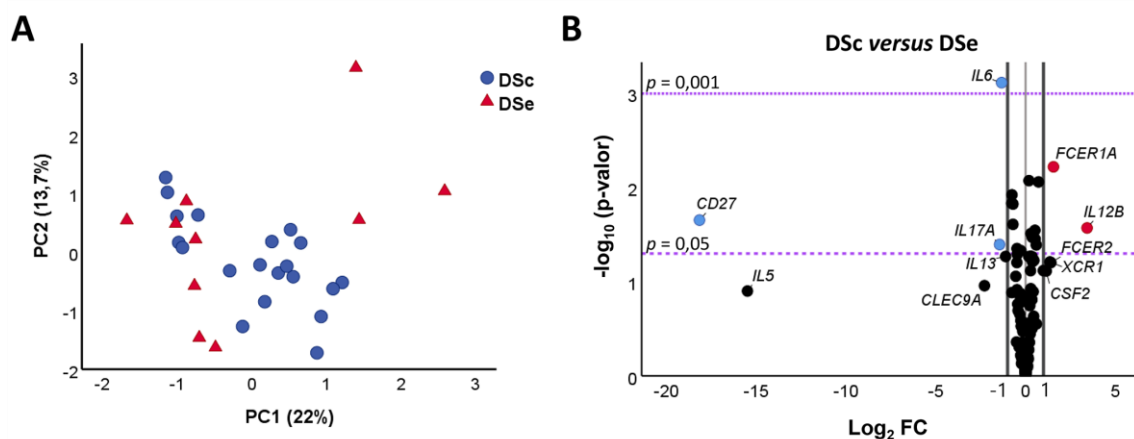
### **B.2.1.1. Análisis comparativo del perfil de expresión génica entre individuos sanos**

Para determinar si existían diferencias en el nivel de expresión génica entre los donantes sanos (DS) relacionadas con su procedencia geográfica, los valores NRQ obtenidos para los 106 genes analizados en sujetos sanos procedentes de áreas endémicas (DSe) y no endémicas (DSc) de la enfermedad de Chagas se analizaron y compararon empleando el software GenEx. Como se observa en el mapa de calor que se muestra en la **Figura R22**, no se observaron diferencias en los valores de expresión génica entre sujetos de países endémicos y no endémicos, ya que los diferentes *clades* generados por el software incluían, indistintamente, sujetos de ambos grupos.

Los valores NRQ de los 106 genes obtenidos en los sujetos DSe y DSc también se emplearon para determinar la estructura del conjunto de datos mediante análisis de componentes principales. Los resultados obtenidos representados en la **Figura R23A** indicaron que la componente principal 1 (PC1) y la componente principal 2 (PC2) representaron el 22% y el 13,7% de la varianza existente entre los individuos, respectivamente. Como se observó, DSe y DSc no exhibieron diferencias en los valores de expresión génica de los genes en estudio (**Figura R23A**) ya que presentaron una distribución muy similar y no fueron separados por las componentes principales. Estos resultados fueron confirmados por la prueba de Mann-Whitney o T de Student, empleado según fuese apropiado, mostrando que no hubo diferencias estadísticamente significativas entre las puntuaciones obtenidas en los dos grupos para cada componente (PC1  $p = 0,65$ , PC2  $p = 0,23$ ).



Para dilucidar si existían diferencias en el nivel de expresión génica de genes individuales entre los donantes sanos procedentes de áreas endémicas y aquellos de regiones no endémicas, se llevó a cabo un tercer análisis en el que se determinó la expresión génica diferencial basado en el *fold change* de la expresión génica (FC) y su significancia estadística. Los resultados obtenidos, representados en un gráfico de volcán (**Figura R23B**) para ilustrar tanto la significancia estadística como la magnitud de los cambios, mostraron que 5 de los 106 genes estudiados se expresaron diferencialmente en el grupo DSc frente a DSe ( $\text{Log}_2 \text{fold change (FC)} > 1$  o  $< -1$ ) con significancia estadística ( $p < 0,05$ ). Dos de los genes identificados (*FCER1A* e *IL12B*) estaban sobreexpresados en DSc en comparación con los sujetos DSe ( $\text{Log}_2 \text{FC} > 1$ , correspondiente a más del doble de expresión). Por el contrario, tres genes (*CD27*, *IL6* e *IL17A*) se encontraban infraexpresados en sujetos DSc frente a DSe ( $\text{Log}_2 \text{FC} < -1$ , correspondiente a, al menos, la mitad de expresión).

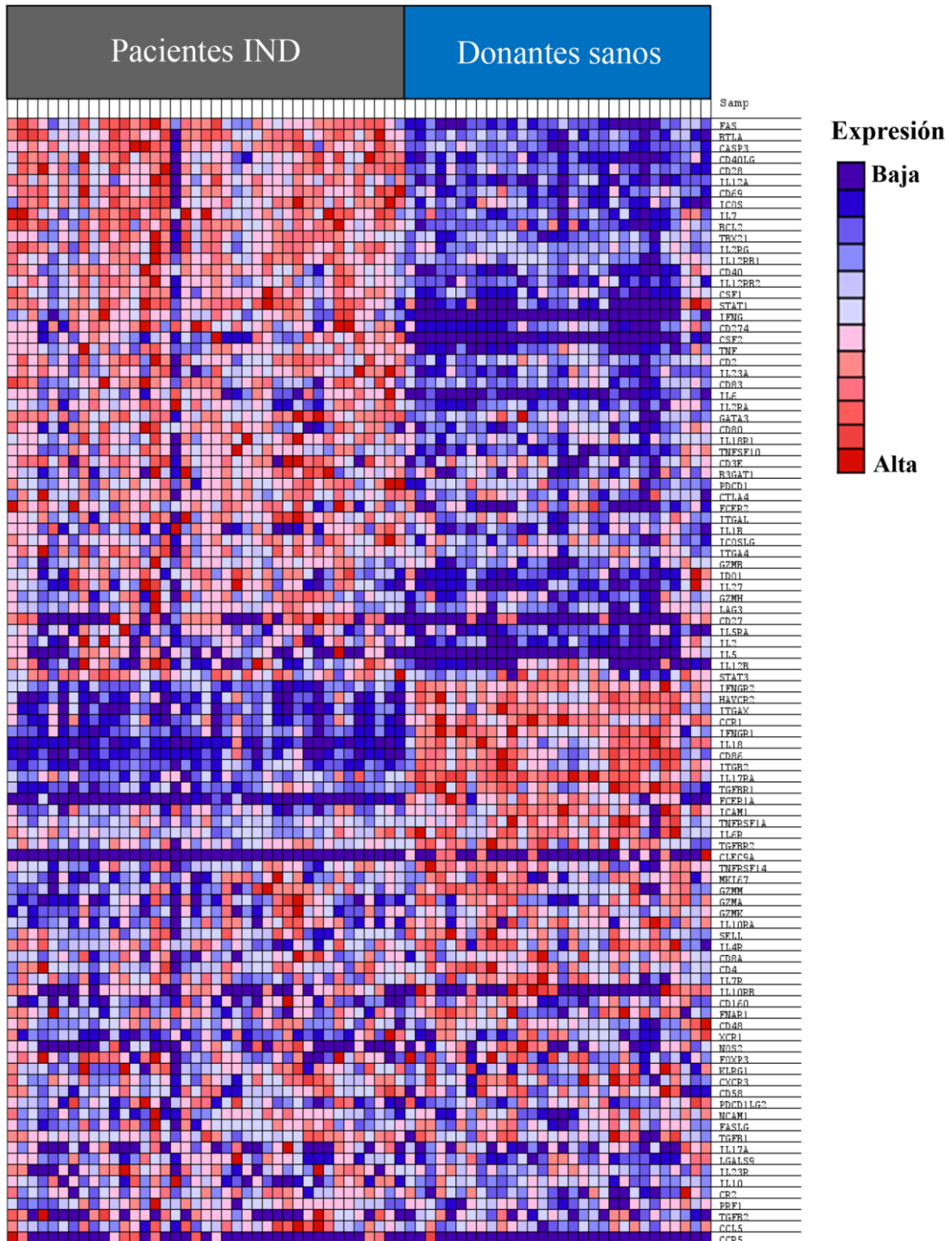


**Figura R23. Análisis comparativo de la expresión génica de 106 genes en donantes sanos procedentes de áreas no endémicas (DSc, n=20) y endémicas (DSe, n=10) de la enfermedad de Chagas. (A)** Análisis de componentes principales (PCA) de los valores NRQ (expresión génica relativa normalizada) de los 106 genes analizados en DSc (círculos azules) y DSe (triángulos rojos). Las componentes principales 1 (PC1) y 2 (PC2) se representan en los ejes X e Y, respectivamente, y la proporción de varianza que explica cada componente se indica en porcentajes. **(B)** Gráfico de volcán de la expresión diferencial de los 106 genes analizados entre las muestras del grupo de sujetos DSc y DSe. El eje X representa el  $\text{log}_2$  del *fold change* entre DSc y DSe ( $\text{Log}_2 \text{FC}$ ), donde el FC se calcula como la ratio entre los dos grupos (DSc / DSe). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ( $-\text{Log}_{10} \text{p-valor}$ ). Las líneas horizontales violetas indican los puntos de corte para los valores de significancia estadística  $p = 0,05$  y  $p = 0,001$ . Las líneas verticales negras representan el  $\text{log}_2$  de FC de  $-1$  y  $1$  (correspondiente a FC de  $0,5$  y  $2$ , respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre DSc y DSe. Adaptada de Gómez, Thomas, et al., 2021.

En conjunto, estos resultados indicaron que no existían diferencias significativas en el nivel de expresión del 95,3% de los genes a estudio entre donantes sanos provenientes de áreas endémicas y aquellos de regiones no endémicas, lo que sugirió que podrían considerarse como un solo grupo de donantes sanos (DS). No obstante, también se analizaron los 5 genes expresados diferencialmente entre DSc y DSe, considerando a ambos como grupos de sujetos independientes, al comparar con la expresión identificada en pacientes IND.

### **B.2.1.2. Identificación de genes con expresión diferencial entre pacientes con Chagas crónico en fase indeterminada (IND) y sujetos sanos (DS)**

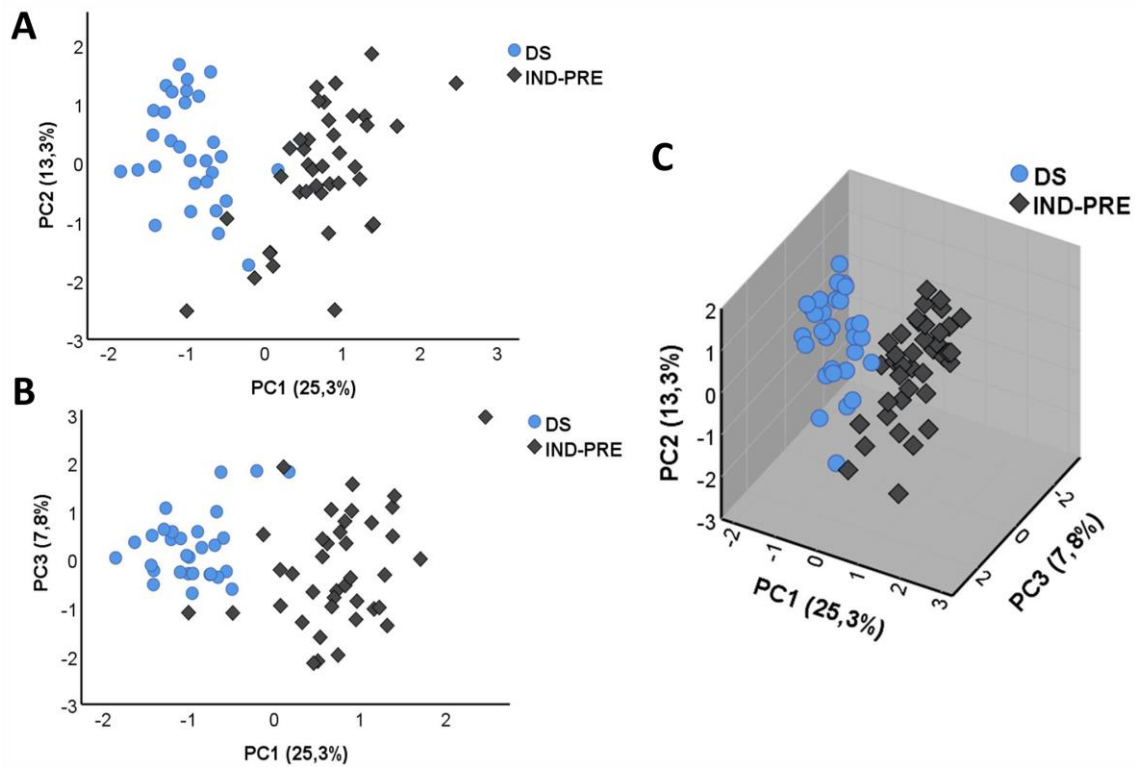
A continuación, se llevó a cabo la comparación del nivel de expresión de los 106 genes entre los pacientes IND y DS. Como primera aproximación, se realizó un análisis de enriquecimiento de conjuntos de genes (GSEA, *Gene Set Enrichment Analysis*) empleando el método computacional GSEA v4.1.0 (Mootha et al., 2003; Subramanian et al., 2005) usando los valores NRQ de los genes en los grupos IND y DS. La **Figura R24** muestra un mapa de calor con los 100 genes principales en los grupos IND y DS, el cual reveló una clara diferencia en el patrón de expresión génica entre los dos grupos. Así, más de la mitad de los genes (al menos 49 genes) se sobreexpresaban en la mayoría de los pacientes IND (**Figura R24**). En cambio, el nivel de expresión de 11 genes se redujo significativamente en la mayoría de los pacientes IND, además de otros 11 genes que también se encontraron infraexpresados en este grupo en comparación con el grupo de donantes sanos (**Figura R24**). Específicamente, el nivel de expresión de los genes *IL18*, *CD86* y *FCER1A* fue extremadamente inferior en prácticamente todos los pacientes del grupo IND en comparación con los sujetos sanos. La expresión de los genes *CLEC9A* y *CCR5* no se detectó en prácticamente ningún individuo incluido en el estudio.



**Figura R24. Mapa de calor de los 100 genes principales determinado por análisis GSEA en pacientes en fase indeterminada (IND) y sujetos sanos (DS).** Los valores del nivel de expresión génica de cada gen se representan como color, que van del rojo oscuro al azul oscuro, según el valor de cantidades relativas normalizadas (NRQ) más alto o más bajo de cada gen, respectivamente. Los genes representados en orden vertical de arriba a abajo son: *FAS, BTLA, CASP3, CD40LG, CD28, IL12A, CD69, ICOS, IL7, BCL2, TBX21, IL2RG, IL12RB1, CD40, IL12RB2, CSF1, STAT1, IFNG, CD274, CSF2, TNF, CD2, IL23A, CD83, IL6, IL2RA, GATA3, CD80, IL18R1, TNFSF10, CD3E, B3GAT1, PDCD1, CTLA4, FCER2, ITGAL, IL1B, ICOSLG, ITGA4, GZMB, IDO1, IL27, GZMH, LAG3, CD27, IL5RA, IL2, IL5, IL12B, STAT3, IFNGR2, HAVCR2, ITGAX, CCR1, IFNGR1, IL18, CD86, ITGB2, IL17RA, TGFBR1, FCER1A, ICAM1, TNFRSF1A, IL6R, TGFB2, CLEC9A, TNFRSF14, MKI67, GZMM, GZMA, GZMK, IL10RA, SELL, IL4R, CD8A, CD4, IL7R, IL10RB, CD160, FNAR1, CD48, XCR1, NOS2, FOXP3, KLRG1, CXCR3, CD58, PDCD1LG2, NCAM1, FASLG, TGFB1, IL17A, LGALS9, IL23R, IL10, CR2, PRF1, TGFB2, CCL5 y CCR5.* Adaptada de Gómez, Thomas, et al., 2021.

Para determinar la estructura del conjunto de datos y examinar la variación entre los sujetos IND y DS, se realizó un análisis de componentes principales. Como se muestra en la **Figura R25A**, la componente principal 1 (PC1) y la componente principal 2 (PC2) acumulan los mayores porcentajes de varianza total alcanzando un 25,3% y 13,3%, respectivamente. Además, la componente principal 3 (PC3) explica el 7,8% de la varianza, representada junto con PC1 en la **Figura R25B**. Estos resultados, junto con la representación gráfica en 3D que se muestra en la **Figura R25C**, indicaron que el nivel de expresión de los genes a estudio era claramente diferente entre los sujetos IND y DS los cuales se ubicaban en grupos separados, principalmente basados en la PC1. Las diferencias observadas en las puntuaciones obtenidas para la PC1 entre ambos grupos de individuos se confirmaron mediante una prueba T de Student que puso de manifiesto la existencia de un perfil de expresión diferencial estadísticamente significativo en IND frente a DS ( $p < 0,0001$ ). A su vez, se aplicó una prueba T de Student a las puntuaciones obtenidas para la PC2 y PC3 que confirmó que las componentes principales 2 y 3 no participaron de manera importante en las diferencias observadas entre los dos grupos de sujetos (PC2  $p = 0,134$ , PC3  $p = 0,061$ ).





**Figura R25. Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes en fase indeterminada (IND, rombos negros) y donantes sanos (DS, círculos azules).** (A) Gráfico PCA de puntuación de las componentes principales 1 (PC1) y 2 (PC2) en los ejes X e Y, respectivamente. (B) Gráfico PCA de puntuación de las componentes principales 1 (PC1) y 3 (PC3) en los ejes X e Y, respectivamente. (C) Gráfico 3D que representa las componentes principales 1, 2 y 3 (PC1, PC2 y PC3). La proporción de la varianza que explica cada componente principal se indica como un porcentaje en el eje junto a la componente principal correspondiente. *Adaptada de Gómez, Thomas, et al., 2021.*

Como se muestra en la **Tabla R24**, la PC1 depende de la expresión de 32 genes con una alta carga factorial. Específicamente, 23 genes estaban correlacionados positivamente con la PC1: *FAS*, *IL12RB1*, *BTLA*, *TBX21*, *BCL2*, *CD40LG*, *IL2RG*, *IL12A*, *CD2*, *IL12RB2*, *CD69*, *CASP3*, *IL7*, *STAT1*, *CSF1*, *ICOS*, *CD28*, *CD40*, *TNF*, *IL18R1*, *GATA3*, *IFNG* y *CD83*), mientras que 9 genes presentaban una correlación negativa con PC1: *ITGB2*, *CCR1*, *IL18*, *HAVCR2*, *CD86*, *IFNGR1*, *IL17RA*, *IFNGR2* y *ITGAX*.

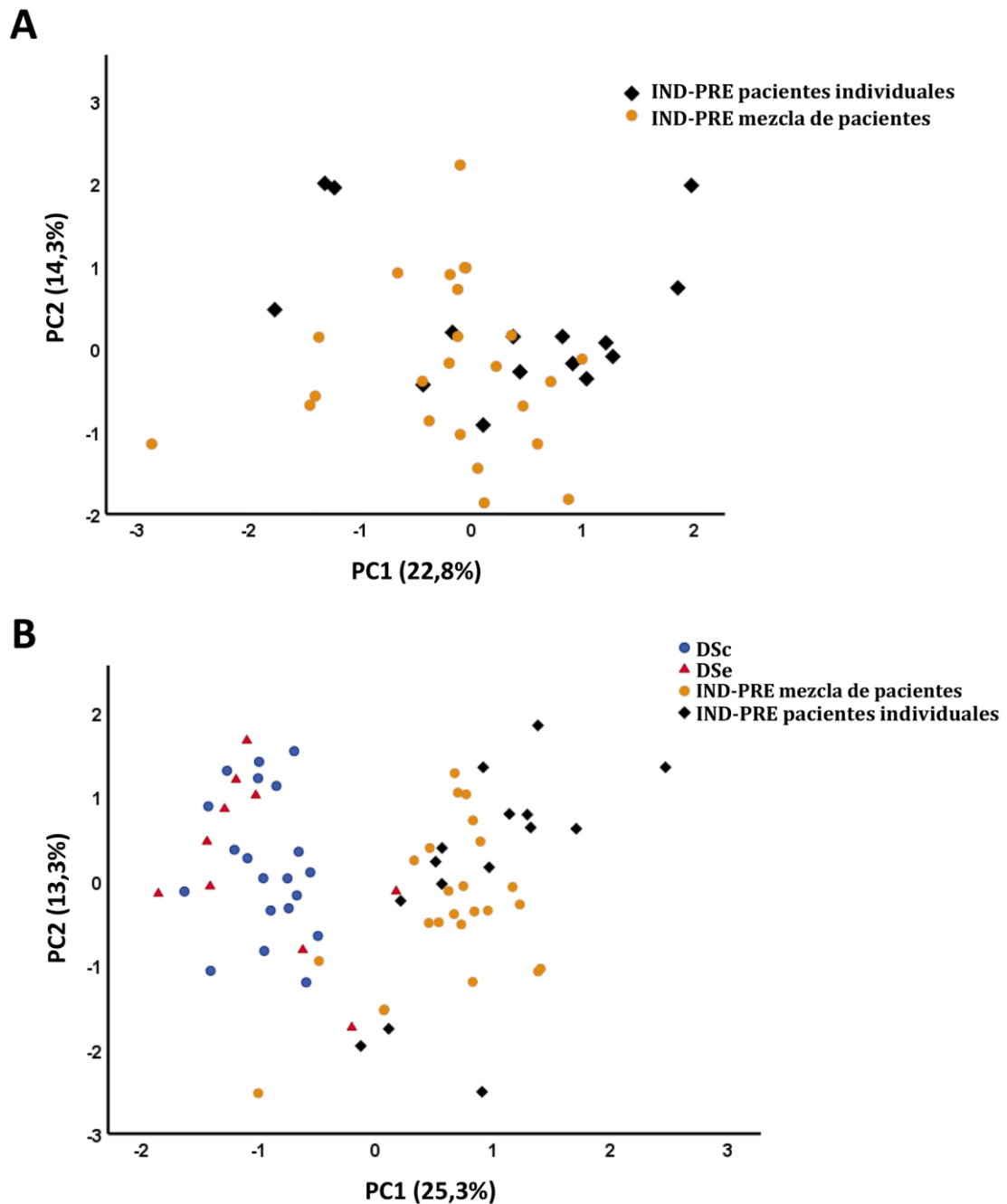
Tabla R24. Genes con carga factorial para la componente principal 1 (PC1) superior a 0,6 o inferior a -0,6 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes en fase indeterminada y donantes sanos. *Adaptada de Gómez, Thomas, et al., 2021.*

Gen	Carga factorial para PC1
FAS	0,887
IL12RB1	0,812
BTLA	0,808
TBX21	0,808
BCL2	0,793
CD40LG	0,789
IL2RG	0,78
IL12A	0,776
CD2	0,771
IL12RB2	0,755
CD69	0,747
CASP3	0,74
IL7	0,736
STAT1	0,734
CSF1	0,732
ICOS	0,728
CD28	0,715
CD40	0,705
TNF	0,69
IL18R1	0,642
GATA3	0,619
IFNG	0,609
CD83	0,608
ITGB2	-0,64
CCR1	-0,643
IL18	-0,691
HAVCR2	-0,697
CD86	-0,698
IFNGR1	-0,736
IL17RA	-0,741
IFNGR2	-0,759
ITGAX	-0,763

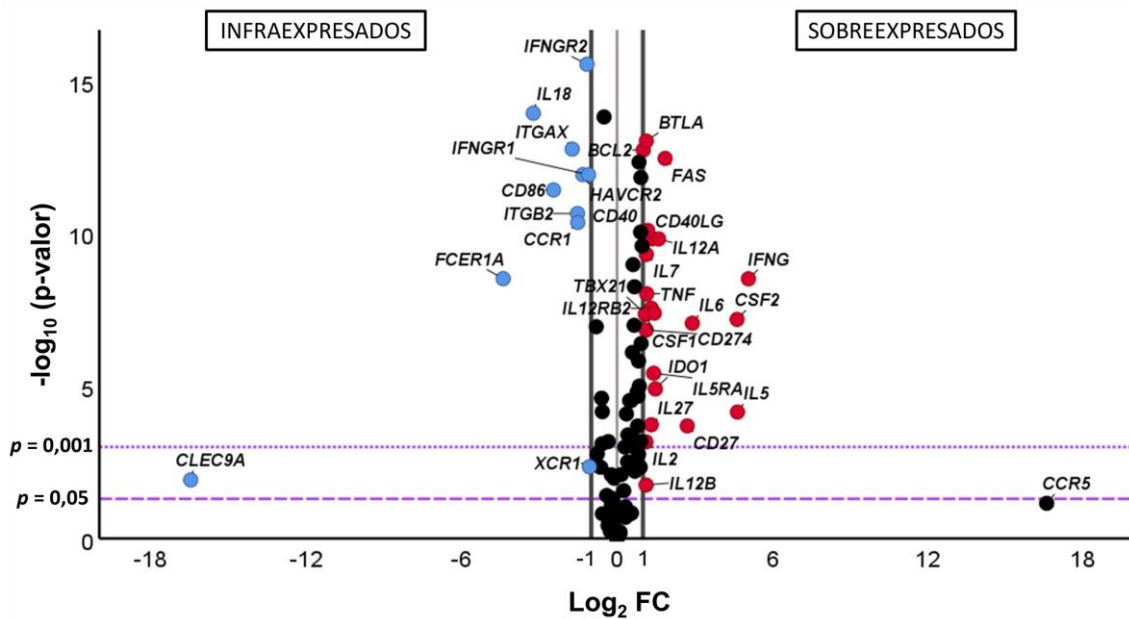
A pesar de que el nivel de expresión de los genes a estudio fue claramente diferente entre el grupo de sujetos IND y el grupo de DS, se analizó la no influencia del agrupamiento de muestras de algunos pacientes IND. Con este propósito, se realizó de nuevo un análisis de componentes principales considerando dos grupos independientes las muestras de pacientes IND que provenían de sujetos independientes y las que contenían el material genético procedente de más de un paciente. Los resultados, que se muestran en la **Figura R26A**, revelaron que todas las muestras de IND mantuvieron la misma distribución independientemente de si provenían de pacientes individuales (38,5% de las muestras) o de varios pacientes. La no existencia de diferencias en el nivel de expresión génica entre los pacientes

IND también fue apoyada por el análisis estadístico aplicado a las puntuaciones obtenidas en los dos grupos para cada componente principal (**Figura R26**). Así se empleó una prueba de Mann-Whitney o prueba T de Student, según correspondiese, que mostró que no hubo diferencias estadísticamente significativas entre las puntuaciones obtenidas en los dos grupos para cada componente (PC1  $p = 0,07$ , PC2  $p = 0,09$ , PC3  $p = 0,23$ , PC4  $p = 0,63$ ). Además, como era de esperar, cuando se compararon estos datos con los de DS (**Figura R26B**), el nivel de expresión de los genes a estudio mostró ser claramente diferente al comparar ambos grupos de IND (pacientes individuales y mezcla de pacientes) con DS, como se había observado previamente cuando el análisis se llevó a cabo con todas las muestras IND *versus* DS (**Figura R25**).

El nivel de expresión génica diferencial en pacientes IND *versus* DS de los genes a estudio se cuantificó usando el *fold change* de la expresión génica (FC), como se describe en la sección *Métodos 5.9*, y junto con la significancia estadística se representaron en un gráfico de volcán (**Figura R27**). Los resultados del análisis comparativo mostrados en la **Figura R27** revelaron que 34 de los 106 genes a estudio estaban diferencialmente expresados entre IND y DS con significancia estadística ( $-\log_{10} p\text{-valor} > 1,3$ , equivalente a  $p < 0,05$ ). De estos genes, 23 se sobreexpresaban considerablemente en los sujetos IND *versus* DS presentando una expresión de, al menos, el doble ( $\text{Log}_2 \text{FC} > 1$ , puntos rojos). Los genes fueron: *BCL2*, *BTLA*, *CD27*, *CD274*, *CD40*, *CD40LG*, *CSF1*, *CSF2*, *FAS*, *IDO1*, *IFNG*, *IL12A*, *IL12B*, *IL12RB2*, *IL2*, *IL27*, *IL5*, *IL5RA*, *IL6*, *IL7*, *STAT1*, *TBX21* y *TNF*. Las diferencias de expresión observadas fueron estadísticamente significativas para todos los genes con un valor  $p < 0,001$  para 22 de los 23 genes y un  $p < 0,05$  para el gen *IL12B*. De forma contraria, 11 genes (*CCR1*, *CD86*, *CLEC9A*, *FCER1A*, *HAVCR2*, *IFNGR1*, *IFNGR2*, *IL18*, *ITGAX*, *ITGB2* y *XCR1*) estaban infraexpresados en pacientes IND cuando su expresión se comparó con la de donantes sanos ( $\text{Log}_2 \text{FC} < -1$ , correspondiente a menos de la mitad de expresión (puntos azules)) con significancia estadística en todos los casos ( $p < 0,05$  para los genes *CLEC9A* y *XCR1* y  $p < 0,001$  para los 9 genes restantes).



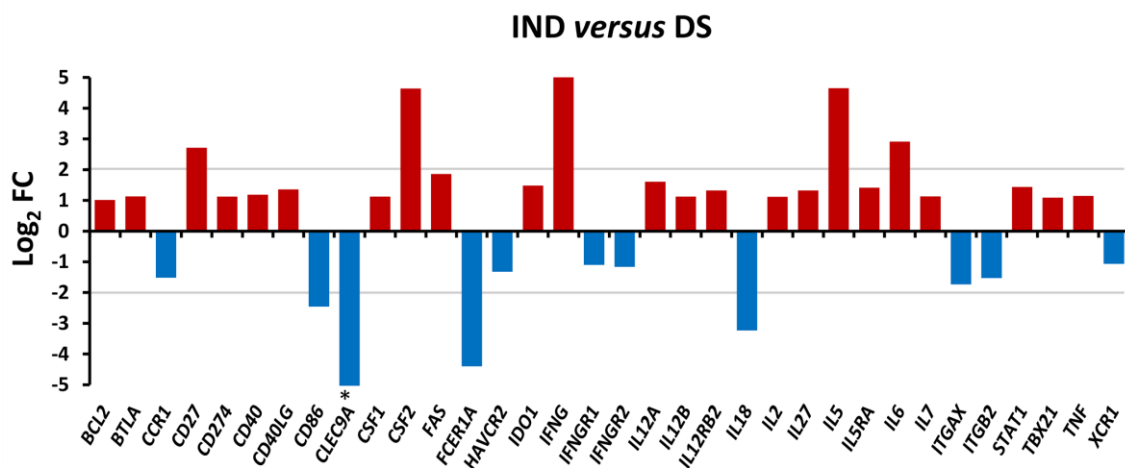
**Figura R26.** Análisis de componentes principales sobre los valores de cantidades relativas normalizadas de expresión génica de 106 genes analizados en muestras correspondientes a (A) mezclas de pacientes IND (círculos naranjas) y pacientes IND individuales (rombos negros) y en (B) DSc (círculos azules), DSe (triángulos rojos), mezclas de pacientes IND (círculos naranjas) y pacientes IND individuales (rombos negros). PC1 y PC2 (componente principal 1 y 2) se trazan en los ejes X e Y, respectivamente, y la proporción de varianza obtenida para ambas componentes se da como porcentaje. Los resultados que se muestran en (A) fueron confirmados por la prueba de Mann-Whitney o la prueba T de Student, según correspondiese. *Adaptada de Gómez, Thomas, et al., 2021.*



**Figura R27. Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes con enfermedad de Chagas en fase crónica indeterminada (IND, n=39) y donantes sanos (DS, n=30) representado en gráfico de volcán.** El eje X representa el  $\log_2$  del *fold change* de expresión entre IND y DS ( $\log_2$  FC), donde el FC se calcula como la ratio entre los dos grupos (IND/DS). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ( $-\log_{10}$  p-valor). Las líneas horizontales violetas indican los puntos de corte para los valores de significancia estadística  $p = 0,05$  y  $p = 0,001$ . Las líneas verticales negras representan el  $\log_2$  de FC de -1 y 1 (correspondiente a FC de 0,5 y 2, respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a los genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre IND y DS. *Adaptada de* Gómez, Thomas, et al., 2021.

Con objeto de visualizar la magnitud del cambio, se llevó a cabo la cuantificación de las diferencias observadas en la expresión génica de los 34 genes expresados diferencialmente entre IND y DS para lo que se emplearon los valores de  $\log_2$  del *fold change* obtenidos para cada gen. Como se muestra en el gráfico de barras de la **Figura R28**, los 34 genes se encontraban sobreexpresados con, al menos, el doble de expresión (barras rojas) o infraexpresados con, al menos, la mitad de expresión (barras azules) en el grupo de IND *versus* DS. Los genes *CD27*, *CSF2*, *IFNG*, *IL5* y *IL6* mostraron las mayores diferencias en el nivel de expresión, estando expresados más de 4 veces ( $FC > 4$ ) en pacientes IND *versus* DS ( $\log_2 FC > 2$ ). Asimismo, los genes *FAS* y *IL12A* se encontraban sobreexpresados, al menos, 3 veces más ( $FC > 3$ ) en IND que en donantes sanos ( $\log_2 FC > 1,5$ ). Por otro lado, el nivel de expresión de los genes *CCR1*, *ITGAX* y *ITGB2* fue, aproximadamente, tres veces menor en estos pacientes ( $\log_2 FC < -1,5$ , correspondiente a  $FC < 0,35$ ) y, por

último, los genes *CD86*, *CLEC9A*, *FCER1A* y *IL18* cuya expresión se redujo más de un cuarto ( $\text{Log}_2 \text{FC} < -2$ , asumiendo un FC menor de 0,25) en pacientes IND en comparación con los donantes sanos (**Figura R28**).



**Figura R28. Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes indeterminados (IND) y donantes sanos (DS).** Cada barra del gráfico corresponde al gen al que se hace referencia en la parte inferior del gráfico (eje X). El eje Y representa el  $\log_2$  del *fold change* ( $\text{Log}_2 \text{FC}$ ) para cada gen. FC es el cociente de la expresión génica media en el grupo de pacientes IND entre la obtenida en el grupo de DS. Los valores positivos (barras rojas) indican genes sobreexpresados y los valores negativos (barras azules) indican genes infraexpresados en IND versus DS. Las barras grises horizontales señalan  $\log_2 \text{FC} = 2$  y  $-2$ , correspondientes a FC de 4 y 0,25, respectivamente. \*Dado que el gen *CLEC9A* no mostró ningún valor de expresión en IND, para aplicar fórmulas, el valor 0 se reemplazó por  $10^{-6}$ , obteniéndose un valor de  $\log_2 \text{FC} = -16,47$ . Adaptada de Gómez, Thomas, et al., 2021.

### B.2.1.3. Análisis de los genes expresados diferencialmente entre sujetos sanos provenientes de áreas endémicas y no endémicas de la enfermedad de Chagas

Dado que se detectaron diferencias estadísticamente significativas en el nivel de expresión de 5 genes (*FCER1A*, *IL12B*, *IL6*, *IL17A* y *CD27*) entre los donantes sanos de áreas endémicas (DSe) y no endémicas (DSc), se analizó en detalle cómo era el nivel de expresión de estos genes en pacientes IND y se determinó si existía alguna relación con lo observado previamente en sujetos sanos. Por consiguiente, se llevaron a cabo análisis comparativos del nivel de expresión génica para cada gen particular entre sujetos IND y los grupos de donantes sanos DSc y DSe de forma independiente. Como se observa en la **Figura R29**, los resultados mostraron diferencias estadísticamente significativas en el nivel de expresión de estos genes entre los grupos de sujetos. Así, el alto nivel de expresión para el gen *FCER1A* detectado en sujetos DSc (NRQ promedio = 13,3) disminuyó en el grupo DSe (NRQ

= 4,5) con  $p < 0,01$ , y se redujo considerablemente en pacientes IND (NRQ = 0,05) con  $p < 0,0001$  al comparar con los valores de expresión de los grupos DSc y DSe. El gen *IL12B* se encontró sobreexpresado en DSc (NRQ = 0,8) frente a DSe (NRQ = 0,1) con  $p < 0,05$  y, a su vez, sobreexpresado en pacientes IND (NRQ = 1,2) con significancia estadística al comparar con la expresión detectada en DSe ( $p < 0,0001$ ). Adicionalmente, se observaron importantes diferencias, estadísticamente significativas ( $p < 0,0001$ ), en el nivel de expresión del gen *IL6*, sobreexpresado en pacientes IND (NRQ = 1,3) en comparación con DSc (NRQ = 0,1) y DSe (NRQ = 0,3). Por otro lado, las diferencias en el nivel de expresión del gen *IL17A* entre IND (NRQ = 0,5) y DSc (NRQ = 0,4) o entre IND y DSe (NRQ = 1,1) no poseían significancia estadística. Respecto al gen *CD27*, la ausencia de expresión del mismo en DSc, el nivel de expresión detectado en DSe (NRQ = 0,3) y, particularmente, su sobreexpresión en pacientes IND (NRQ = 0,7) condujeron a que las diferencias en el nivel de expresión entre IND y DSc tuvieran significancia estadística ( $p < 0,001$ ) (Figura R29). En conjunto, estos resultados sugieren que solo se detectaron diferencias en el nivel de expresión de los genes *FCER1A* e *IL6* en IND frente a DSc y DSe.

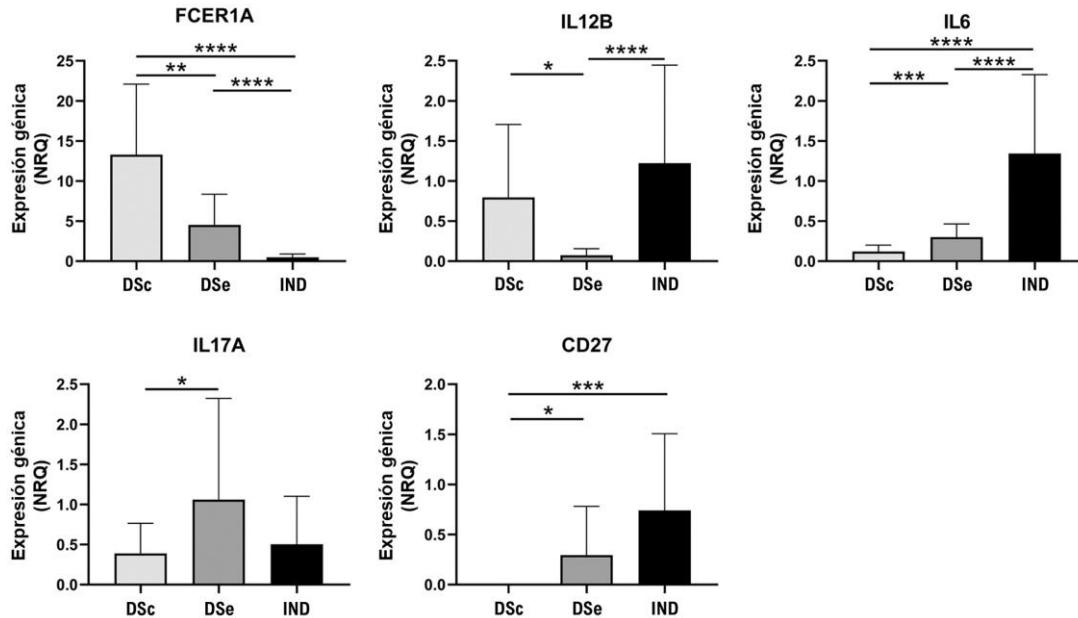
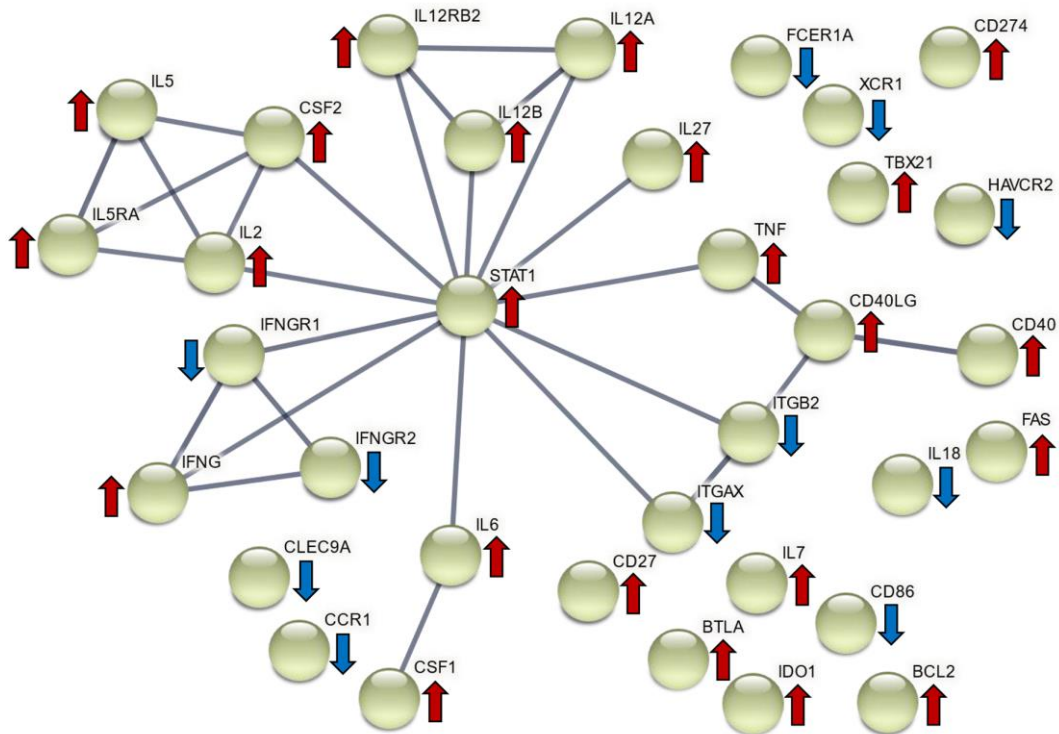


Figura R29. Análisis comparativo del nivel de expresión de los genes *FCER1A*, *IL12B*, *IL6*, *IL17A* y *CD27* medidos como cantidades relativas normalizadas (NRQ) promedio en pacientes con enfermedad de Chagas crónica en fase indeterminada (IND) y donantes sanos de áreas endémicas (DSe) y no endémicas (DSc) de la enfermedad de Chagas. Se indican las diferencias estadísticamente significativas determinadas por la prueba de Mann-Whitney o la prueba T de Student, según corresponda (\*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$  y \*\*\*\*  $p < 0,0001$ ). Adaptada de Gómez, Thomas, et al., 2021.

### **B.2.1.4. Identificación de las vías inmunológicas implicadas en el establecimiento de la infección por *T. cruzi* en pacientes con enfermedad de Chagas crónico en fase indeterminada (IND)**

Con objeto de determinar las rutas inmunológicas implicadas en la infección crónica por *T. cruzi* en pacientes con enfermedad de Chagas en fase crónica indeterminada (IND), se analizó las interacciones conocidas que tenían lugar entre las proteínas que codificaban los genes cuya expresión fue diferencial entre los pacientes IND y los donantes sanos. Los genes que cumplían los umbrales de significancia estadística y biológica ( $p < 0,05$  y  $\text{Log}_2 \text{FC} > 1$  o  $> -1$ , respectivamente) se consideraron biológicamente relevantes y se utilizaron para llevar a cabo esta interpretación. En consecuencia, se construyó una red de interacción proteína-proteína (IPP) utilizando la plataforma STRING, exigiendo el máximo valor de confianza a las interacciones predichas. Los resultados obtenidos, representados en la **Figura R30**, mostraron que las 34 proteínas codificadas por el conjunto de genes expresados diferencialmente en sujetos IND *versus* DS tenían un alto grado de interacción alcanzando hasta 29 interacciones con un valor  $p$  de enriquecimiento de la interacción proteína-proteína  $< 1,0\text{e-}16$ . Sin embargo, solo se esperarían 3 interacciones de un conjunto de genes aleatorio del mismo tamaño por acción del azar.





**Figura R30. Análisis STRING de interacción proteína-proteína (IPP).** Análisis STRING de la relación entre los 34 genes expresados diferencialmente, con significancia estadística, entre pacientes IND y donantes sanos. Los nodos de la red representan las proteínas codificadas por los genes expresados diferencialmente y el enlace de cada nodo representa las interacciones conocidas para estas moléculas en diferentes procesos inmunológicos; Valor  $p$  de enriquecimiento de la interacción proteína-proteína  $< 1,0e-16$ . Las líneas de unión de la red representan la confianza (indicando el grosor de la línea la solidez de los datos). Para la construcción de las redes IPP se fijó el umbral de confianza en el máximo (0,9) y se aplicaron fuentes de interacción activa, incluyendo experimentos publicados, bases de datos, coocurrencia, fusión de genes, vecindad y coexpresión, así como se limitó la especie a *Homo sapiens*. Las flechas indican si los genes que codifican estas proteínas se encontraban sobreexpresados (rojo) o infraexpresados (azul) en los pacientes IND respecto a DS. Adaptada de Gómez, Thomas, et al., 2021.

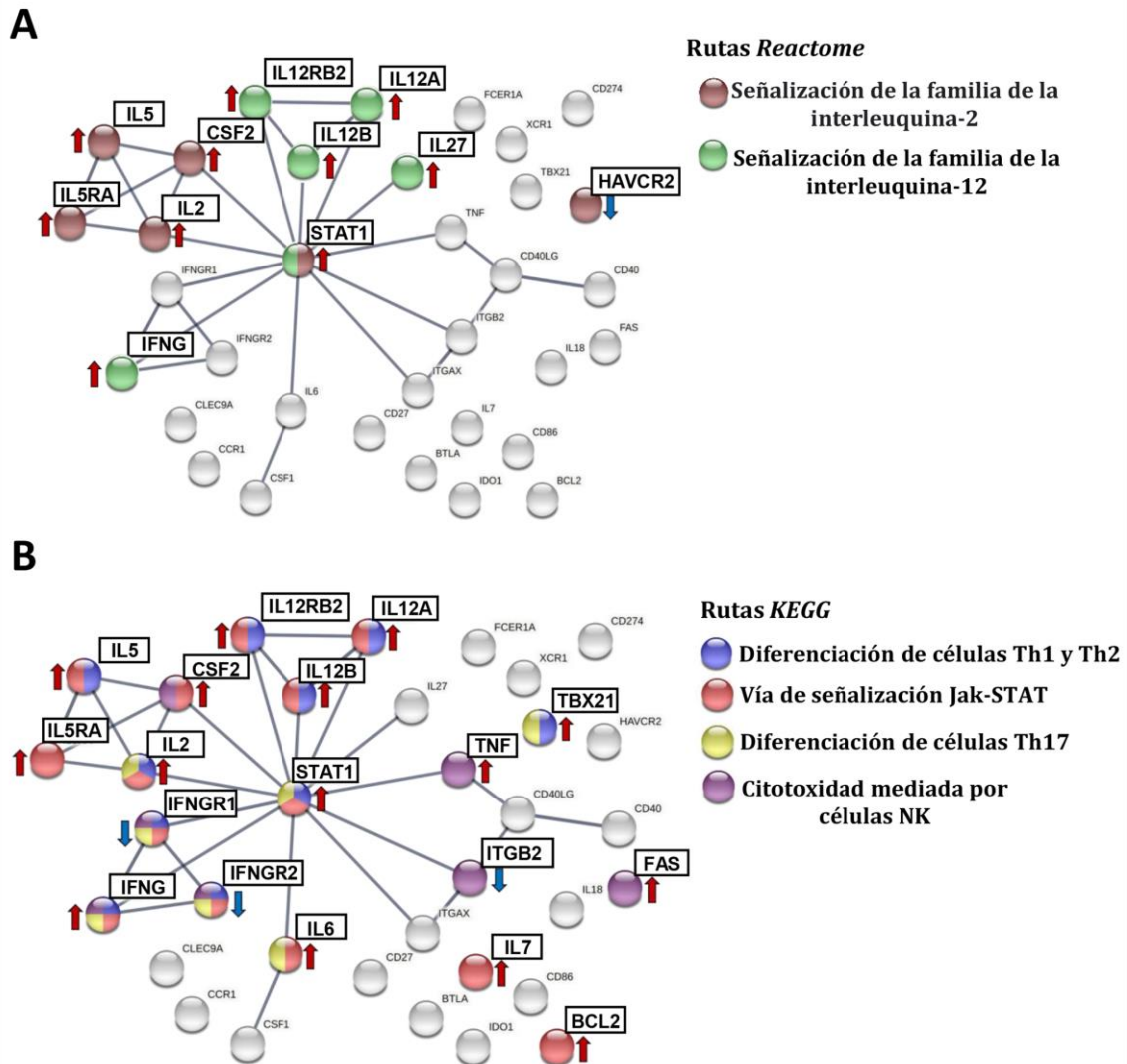
Además, algunas de estas proteínas se agruparon según las rutas biológicas más relevantes en las que están involucradas, pertenecientes a las bases de datos *KEGG* y *Reactome*. Como se resume en la **Tabla R25**, los resultados de STRING mostraron que 6 de estas proteínas participaban en la ruta de señalización de la familia de Interleuquina-2 (HSA-451927) con un FDR (*False Discovery Rate*) de  $5,37e-09$ , 10 proteínas en la diferenciación de células Th1 y Th2 (HSA-04658, FDR  $1,3e-14$ ), 6 proteínas en la señalización de la familia de Interleuquina-12 (HSA-447115, FDR  $2,2e-08$ ), 14 proteínas en la vía de señalización Jak-STAT (HSA-04630, FDR  $4,77e-19$ ), 7 proteínas en la diferenciación de células Th17 (HSA-04659, FDR  $3,02e-09$ ) y 7 proteínas en la ruta de citotoxicidad mediada por células citotóxicas naturales o *natural killer* (NK) (HSA-04650, FDR  $9,75e-09$ ).

## RESULTADOS

**Tabla R25. Principales rutas biológicas en las que participan las proteínas de la red de STRING codificadas por los genes expresados diferencialmente (GED) entre pacientes en fase indeterminada y donantes sanos. FDR: false discovery rate. Adaptada de Gómez, Thomas, et al., 2021.**

Base de datos	Rutas	Descripción	Número de GED	FDR
Rutas <i>Reactome</i>	HSA-451927	Señalización de la familia de Interleuquina-2	6	5,37e-09
Rutas <i>KEGG</i>	HSA-04658	Diferenciación de células Th1 y Th2	10	1,3e-14
Rutas <i>Reactome</i>	HSA-447115	Señalización de la familia de Interleuquina-12	6	2,2e-08
Rutas <i>KEGG</i>	HSA-04630	Ruta de señalización Jak-STAT	14	4,77e-19
Rutas <i>KEGG</i>	HSA-04659	Diferenciación de células Th17	7	3,02e-09
Rutas <i>KEGG</i>	HSA-04650	Citotoxicidad mediada por células NK	7	9,75e-09

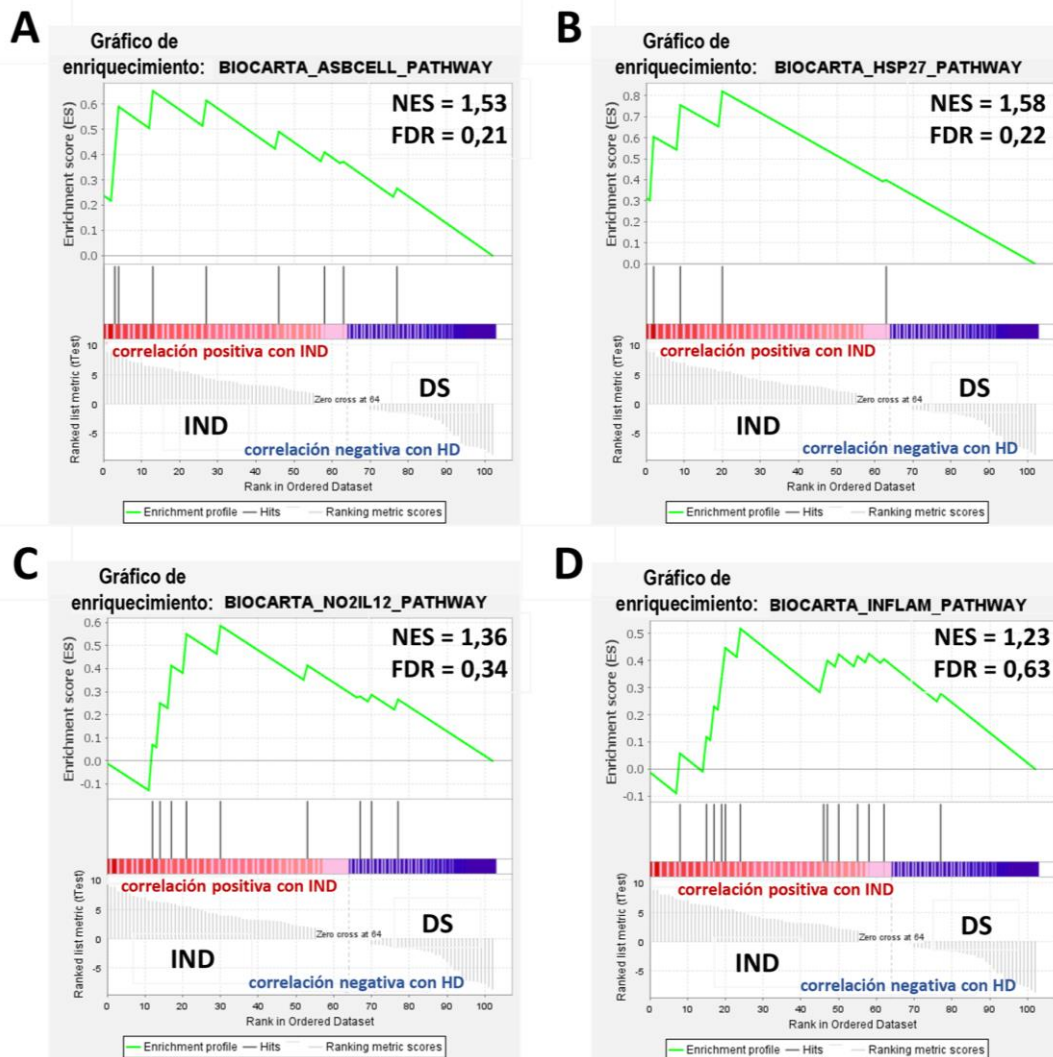
A fin de resaltar los genes sobreexpresados o infraexpresados en pacientes IND que codifican las proteínas involucradas en cada una de las mencionadas rutas inmunológicas, se realizó una representación gráfica de las mismas en STRING como subredes (**Figura R31**). Como se muestra en la **Figura R31A**, las proteínas CSF2, HAVCR2, IL2, IL5, IL5RA y STAT1 estaban involucradas en la ruta de señalización de la Interleuquina-2 (**Figura R31A**, nodos marrones); IFNG, IL12A, IL12B, IL12RB2, IL27 y STAT1 en la ruta de señalización de la Interleuquina-12 (**Figura R31A**, nodos verdes). Por otro lado, las proteínas IFNG, IFNGR1, IFNGR2, IL12A, IL12B, IL12RB2, IL2, IL5, STAT1 y TBX21 participaban en la ruta de diferenciación de células Th1 y Th2 (**Figura R31B**, nodos azules); BCL2, CSF2, IFNG, IFNGR1, IFNGR2, IL12A, IL12B, IL12RB2, IL2, IL5, IL5RA, IL6, IL7 y STAT1 en la ruta de señalización Jak-STAT (**Figura R31B**, nodos rojos) e IFNG, IFNGR1, IFNGR2, IL2, IL6, STAT1 y TBX21 en la ruta de diferenciación de células Th17 (**Figura R31B**, nodos amarillos). Por último, las proteínas CSF2, FAS, IFNG, IFNGR1, IFNGR2, ITGB2 y TNF formaban parte de la ruta de citotoxicidad mediada por células NK (**Figura R31B**, nodos violetas). En todos los casos, los genes que codifican las referidas proteínas se encontraban sobreexpresados en los pacientes IND *versus* DS, con la excepción de los genes *ITGB2*, *HAVCR2*, *IFNGR1* e *IFNGR2* los cuales se infraexpresaban en estos pacientes.



**Figura R31. Subredes STRING de interacción proteína-proteína (IPP).** Representación de las proteínas codificadas por genes expresados diferencialmente entre IND y DS, los cuales se ha descrito están involucrados en rutas biológicas específicas. Para la construcción de las redes IPP se fijó el umbral de confianza en el máximo (0,9) y se aplicaron fuentes de interacción activa, incluyendo experimentos publicados, bases de datos, coocurrencia, fusión de genes, vecindad y coexpresión, así como se limitó la especie a *Homo sapiens*. Los nodos coloreados corresponden a proteínas involucradas en cada ruta biológica: **(A)** señalización de la familia de la Interleuquina-2 (HSA-451927) (marrón) y señalización de la familia de la Interleuquina-12 (HSA-447115) (verde); incluidas en la base de datos *Reactome*. **(B)** Diferenciación de células Th1 y Th2 (HSA-04658) (azul), ruta de señalización Jak-STAT (HSA-04630) (rojo), diferenciación de células Th17 (HSA-04659) (amarillo) y ruta de citotoxicidad mediada por células NK (HSA-04650) (morado); incluidas en la base de datos *KEGG*. Las flechas indican si se encontró que los genes que codifican estas proteínas estaban sobreexpresados (rojo) o infraexpresados (azul) en los pacientes IND *versus* DS. *Adaptada de* Gómez, Thomas, et al., 2021.

### B.2.1.5. Análisis de Enriquecimiento de Conjuntos de Genes (GSEA)

Para identificar las rutas inmunológicas asociadas a los genes con expresión génica diferencial en IND *versus* DS, se realizó un análisis de enriquecimiento de conjuntos de genes (GSEA) empleando los valores NRQ de estos genes y la colección de conjuntos de genes BioCarta incluida en la base de datos MSigDB (*Molecular Signatures Database*) (Liberzon et al., 2011; Subramanian et al., 2005). Se obtuvieron gráficos de enriquecimiento de las rutas seleccionadas para ilustrar la correlación positiva o negativa entre el conjunto de genes específico sobreexpresados o infraexpresados en IND y DS para cada ruta. Como se muestra en la **Figura R32**, el análisis GSEA reveló una correlación positiva en el fenotipo IND para varias rutas inmunológicas, ya que muchos genes implicados en estas vías se sobreexpresaron en pacientes IND frente a DS. Así, algunas de las rutas enriquecidas diferencialmente en IND fueron la activación de células B dependiente de antígeno (BIOCARTA\_ASBCCELL\_PATHWAY, **Figura R32A**), inducción de la regulación de las HSP por el estrés (BIOCARTA\_HSP27\_PATHWAY, **Figura R32B**), la vía IL12 dependiente de NO2 en las células NK (BIOCARTA\_NO2IL12\_PATHWAY, **Figura R32C**) y la vía citoquinas y respuesta inflamatoria (BIOCARTA\_INFLAM\_PATHWAY, **Figura R32D**), con una puntuación de enriquecimiento normalizada (NES, *Normalized Enrichment Score*) de 1,53, 1,58, 1,36 y 1,23 y un q-valor de FDR de 0,21, 0,22, 0,34 y 0,63, respectivamente.



**Figura R32. Gráficos de enriquecimiento procedentes del análisis de enriquecimiento de conjuntos de genes (GSEA).** Se muestra la correlación con pacientes con enfermedad de Chagas en fase indeterminada (IND) y donantes sanos (DS) de genes incluidos en las rutas **(A)** activación de células B dependiente de antígeno (BIOCARTA\_ASBCCELL\_PATHWAY), **(B)** inducción de la regulación de las HSP por el estrés (BIOCARTA\_HSP27\_PATHWAY), **(C)** vía IL12 dependiente de NO2 en las células NK (BIOCARTA\_NO2IL12\_PATHWAY) y **(D)** vía citoquinas y respuesta inflamatoria (BIOCARTA\_INFLAM\_PATHWAY). En color verde se indica la curva de puntuación de enriquecimiento (ES). Los parámetros establecidos para GSEA fueron: colección de conjunto de genes BioCarta (MSigDB), permutaciones: 100.000, tipo de permutación: fenotipo ( $n^{\circ}$  muestras > 7), estadística de enriquecimiento: ponderado, métrica de clasificación: Prueba T, tamaño máximo: 500, tamaño mínimo: 3. NES: puntuación de enriquecimiento normalizada, FDR: false discovery rate q-value. Adaptada de Gómez, Thomas, et al., 2021.

Seguidamente, se analizó en los pacientes IND el número de genes que se encontraban sobreexpresados del total de genes incluidos en cada una de las rutas GSEA. Para ampliar el análisis, también se tuvo en consideración la identidad de los genes sobreexpresados que presentaban un valor de  $\log_2$  FC mayor a 1 y los que se correlacionaban consistentemente con la componente principal 1. Como se puede

observar, la **Tabla R26** integra los resultados obtenidos en GSEA, PCA y gráfico de volcán de los genes expresados diferencialmente en pacientes IND *versus* DS para cada una de las cuatro vías de BioCarta (con carga factorial con PC1 > 0,6 o Log<sub>2</sub> FC > 1). En este estudio se analizaron 10 de los 15 genes incluidos en la ruta de BioCarta "activación de células B dependientes de antígeno". Cinco de ellos (*CD28*, *CD40*, *CD40LG*, *FAS* e *IL2*) sobreexpresados en los pacientes IND en comparación con DS. De un total de 14 genes que estaban incluidos en la ruta de BioCarta "inducción de la regulación de las HSP por el estrés", encontramos 4 genes (*FAS*, *CASP3*, *BCL2* y *TNF*) de los cinco ensayados que estaban sobreexpresados en IND *versus* DS. En cuanto a la ruta "Citoquinas y respuesta inflamatoria", también se detectó un aumento de expresión en IND de 8 (*IL2*, *TNF*, *CSF1*, *CSF2*, *IFNG*, *IL5*, *IL6* e *IL7*) de los 14 genes incluidos en el array de expresión de un total de 29 genes que componen esta ruta. Además, se encontró que 4 genes (*IFNG*, *CD2*, *IL12RB1* e *IL12RB2*) de los 9 genes analizados estaban sobreexpresados en IND *versus* DS, de los 15 genes incluidos en la ruta "IL-12 dependiente de NO<sub>2</sub> en células NK". Cabe destacar que del total de 17 genes que se sobreexpresaron en IND *versus* DS y participaban en una o más de las vías seleccionadas, 4 de ellos (*CSF2*, *IL2*, *IL5* e *IL6*) alcanzaron un log<sub>2</sub> de FC > 1 y *p* < 0,05, en IND *versus* DS, 4 genes (*CASP3*, *CD2*, *CD28* e *IL12RB1*) se correlacionaban con PC1 con una carga factorial superior a 0,6 (componente principal responsable de las diferencias entre ambos grupos) y 9 genes (*BCL2*, *CD40*, *CD40LG*, *CSF1*, *FAS*, *IFNG*, *IL12RB2*, *IL7* y *TNF*) cumplieron ambos criterios.

Tabla R26. Lista de genes expresados diferencialmente entre pacientes en fase indeterminada (IND) y donantes sanos, enriquecidos en el fenotipo IND según el análisis GSEA, que participan en las rutas de BioCarta: activación de células B dependientes de antígeno (BIOCARTA\_ASBCCELL\_PATHWAY), inducción de la regulación de las HSP por el estrés (BIOCARTA\_HSP27\_PATHWAY), citoquinas y respuesta inflamatoria (BIOCARTA\_INFLAM\_PATHWAY) e IL-12 dependiente de NO2 en células NK (BIOCARTA\_NO2IL12\_PATHWAY). GED: genes expresados diferencialmente, FC: fold change, CF: carga factorial, PC1: componente principal 1. Adaptada de Gómez, Thomas, et al., 2021.

Gen	Activación de células B dependientes de antígeno (n=15)	Inducción de la regulación de las HSP por el estrés (n=14)	Citoquinas y respuesta inflamatoria (n=29)	IL-12 dependiente de NO2 en células NK (n=15)	Log <sub>2</sub> FC > 1	CF para PC1 > 0,6
	5 GED / 10 analizados	4 GED / 5 analizados	8 GED / 14 analizados	4 GED / 9 analizados		
<i>CD28</i>	↑					×
<i>CD40</i>	↑				×	×
<i>CD40LG</i>	↑				×	×
<i>FAS</i>	↑	↑			×	×
<i>IL2</i>	↑		↑		×	
<i>CASP3</i>		↑				×
<i>BCL2</i>		↑			×	×
<i>TNF</i>		↑	↑		×	×
<i>CSF1</i>			↑		×	×
<i>CSF2</i>			↑		×	
<i>IFNG</i>			↑	↑	×	×
<i>IL5</i>			↑		×	
<i>IL6</i>			↑		×	
<i>IL7</i>			↑		×	×
<i>CD2</i>				↑		×
<i>IL12RB1</i>				↑		×
<i>IL12RB2</i>				↑	×	×

También se constató la sobreexpresión de estos 17 genes en IND *versus* DS cuando se realizó el mencionado análisis considerando solo las muestras IND que provenían de pacientes independientes. Como era de esperar, se obtuvieron valores de log<sub>2</sub> FC similares a los detectados cuando se analizaban todas las muestras de pacientes IND *versus* DS (Tabla R27).



**Tabla R27. Comparación de los valores de expresión génica diferencial obtenidos para los genes implicados en las vías enriquecidas predichas por GSEA entre todas las muestras IND y sólo las muestras IND de pacientes individuales frente a donantes sanos (DS). Adaptada de Gómez, Thomas, et al., 2021.**

	<b>Log<sub>2</sub> FC IND (todas las muestras) versus DS</b>	<b>Log<sub>2</sub> FC IND (muestras de pacientes individuales) versus DS</b>
<i>CD28</i>	0,967	1,060
<i>CD40</i>	1,185	1,235
<i>CD40LG</i>	1,355	1,328
<i>FAS</i>	1,856	1,898
<i>IL2</i>	1,115	1,254
<i>CASP3</i>	0,846	0,930
<i>BCL2</i>	1,014	1,057
<i>TNF</i>	1,148	1,227
<i>CSF1</i>	1,123	1,250
<i>CSF2</i>	4,638	3,696
<i>IFNG</i>	5,076	5,510
<i>IL5</i>	4,649	4,580
<i>IL6</i>	2,9099	2,041
<i>IL7</i>	1,131	1,061
<i>CD2</i>	0,593	0,668
<i>IL12RB1</i>	0,660	0,725
<i>IL12RB2</i>	1,321	1,484

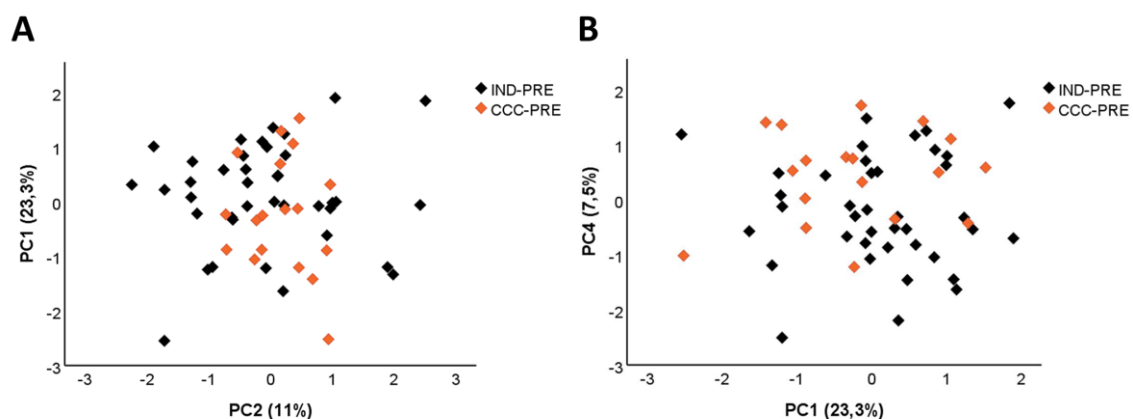
**B.2.2. EXPRESIÓN DIFERENCIAL DE GENES DE RESPUESTA INMUNE EN PACIENTES CON ENFERMEDAD DE CHAGAS EN FASE CARDIACA VERSUS PACIENTES ASINTOMÁTICOS**

Con el objetivo de conocer cómo afecta la progresión de la enfermedad a la respuesta inmune específica generada frente a *T. cruzi* en los pacientes con enfermedad crónica de Chagas, se estudió el patrón de expresión de los 106 genes seleccionados, involucrados en respuesta inmune, en pacientes con enfermedad de Chagas en fase cardiaca (CCC, n=18) en comparación con pacientes en fase indeterminada (IND, n=39). Para llevar a cabo los distintos análisis comparativos, se emplearon los valores de expresión génica relativa normalizada (NRQ) que presentaban las células PBMC de estos pacientes, estimuladas *in vitro* con antígenos del parásito.

En primer lugar, se aplicó un análisis de componentes principales al conjunto de valores NRQ de los 106 genes a estudio de ambos grupos para determinar su estructura y las similitudes de expresión entre los sujetos IND y CCC. Como se muestra en la **Figura R33**, la componente principal 1 (PC1) y la componente



principal 2 (PC2) acumulan los mayores porcentajes de varianza total con valores de 23,3% y 11%, respectivamente (**Figura R33A**). Los resultados mostraron que no existían llamativas diferencias de expresión entre ambos grupos puesto que las puntuaciones adquiridas para las componentes principales representadas eran similares entre ambos grupos de pacientes. Este hecho fue confirmado mediante una prueba T de Student, que indicó que no existían diferencias de expresión estadísticamente significativas en los genes que se correlacionaban con PC1 y PC2 entre los dos grupos (PC1  $p = 0,263$ , PC2  $p = 0,387$ ). Cabe destacar que para la PC2 se observaron puntuaciones más homogéneas en los pacientes CCC, indicativo de la mayor heterogeneidad en la expresión de los genes a estudio en los pacientes IND. La PC2 depende de la expresión de 11 genes con alta carga factorial, todos ellos correlacionados positivamente con dicha componente (**Tabla R28**). En la **Figura R33B**, junto a la PC1, se representa la componente principal 4 (PC4) que explica el 7,5% de la varianza. Si bien las muestras de ambos grupos se distribuían por todo el eje X e Y indistintamente, se podía observar una tendencia a puntuaciones positivas para la PC4 en pacientes CCC y negativas en IND. Esta diferencia en las puntuaciones para PC4 entre ambos grupos de individuos se confirmaron mediante una prueba de Mann-Whitney que puso de manifiesto la existencia de un perfil de expresión diferencial, estadísticamente significativo, en pacientes CCC frente a IND ( $p = 0,008$ ) en los 3 genes que se correlacionaban con alta carga factorial con PC4 (**Tabla R28**).

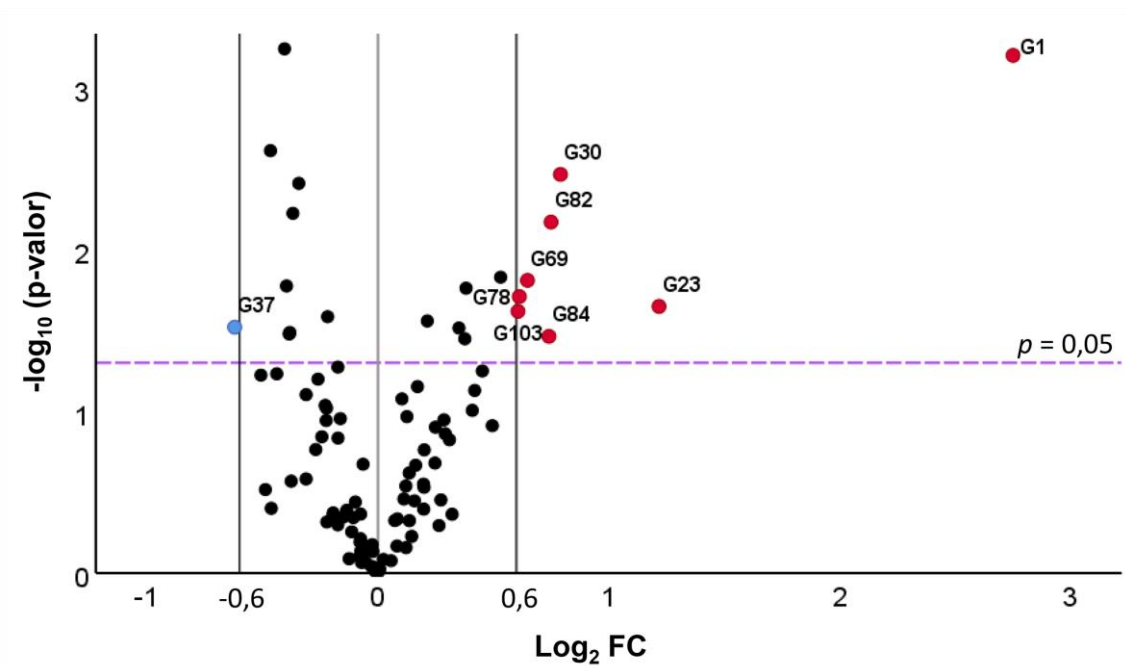


**Figura R33. Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes en fase indeterminada (IND, rombos negros) y pacientes en fase cardiaca (CCC, rombos naranjas).** (A) Gráfico PCA de puntuación de las componentes principales 2 (PC2) y 1 (PC1) en los ejes X e Y, respectivamente. (B) Gráfico PCA de puntuación de las componentes principales 1 (PC1) y 4 (PC4) en los ejes X e Y, respectivamente. La proporción de la varianza que explica cada componente principal se indica como un porcentaje en el eje junto a la componente principal correspondiente.

Tabla R28. Genes con carga factorial para la componente principal 2 (PC2) y componente principal 4 (PC4) superior a 0,54 o inferior a -0,54 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes en fase indeterminada y cardiaca.

Gen	Carga factorial para PC2	Gen	Carga factorial para PC4
G26	0,542	G7	0,541
G32	0,748	G92	-0,582
G36	0,578	G103	0,558
G57	0,544		
G58	0,692		
G64	0,760		
G66	0,717		
G81	0,629		
G89	0,598		
G93	0,643		
G101	0,565		

Una vez realizada la primera aproximación para tener una visión global de los datos, se procedió a analizar el nivel de expresión génica de los 106 genes a estudio en ambos grupos, al objeto de identificar aquellos genes cuya expresión fue diferencial con la progresión de la enfermedad. Así, se realizó un análisis comparativo basado en el *fold change* de la expresión génica (FC) y la significancia estadística, representados ambos parámetros en un gráfico de volcán (**Figura R34**). Los resultados indicaron que 9 de los 106 genes a estudio se encontraban diferencialmente expresados entre pacientes CCC e IND con significancia estadística. Ocho de estos genes se sobreexpresaban en los pacientes CCC *versus* IND ( $\text{Log}_2 \text{FC} > 0,6$  correspondiente a  $\text{FC} > 1,5$ ; puntos rojos), mientras que el gen restante se encontró infraexpresado en pacientes CCC cuando su expresión se comparó con la de pacientes en fase indeterminada ( $\text{Log}_2 \text{FC} < -0,6$ , correspondiente a  $\text{FC} < 0,66$ ; puntos azules). Las diferencias de expresión observadas fueron estadísticamente significativas para los 9 genes con un  $-\log_{10} p\text{-valor} > 1,3$  equivalente a un valor  $p < 0,05$  (**Figura R34**).



**Figura R34. Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes con enfermedad de Chagas en fase cardiaca (CCC, n=18) y en fase indeterminada (IND, n=39) representado en gráfico de volcán.** El eje X representa el  $\log_2$  del *fold change* de expresión entre CCC y IND ( $\log_2$  FC), donde el FC se calcula como la ratio entre los dos grupos (CCC/IND). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ( $-\log_{10}$  p-valor). La línea horizontal violeta indica el punto de corte para el valor de significancia estadística  $p = 0,05$ . Las líneas verticales negras representan el  $\log_2$  de FC de  $-0,6$  y  $0,6$  (correspondiente a FC de  $0,66$  y  $1,5$ , respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre pacientes CCC e IND.

La magnitud del cambio se visualizó mediante la cuantificación de las diferencias observadas en la expresión génica de los 9 genes con expresión diferencial entre pacientes CCC e IND para lo que se emplearon los valores de  $\log_2$  del *fold change* obtenidos para cada gen. La representación empleando un gráfico de barras (**Figura R35**) reveló que, si bien todos presentaban  $\log_2$  FC  $> 0,6$  (barras rojas) o inferior a  $-0,6$  (barra azul), para dos de los genes las diferencias de expresión eran mayores con un valor de  $\log_2$  FC  $> 1$  (G23) y  $\log_2$  FC  $> 2,5$  (G1), lo que supone una sobreexpresión de estos genes en pacientes CCC respecto a IND de más del doble y alrededor de seis veces más, respectivamente.

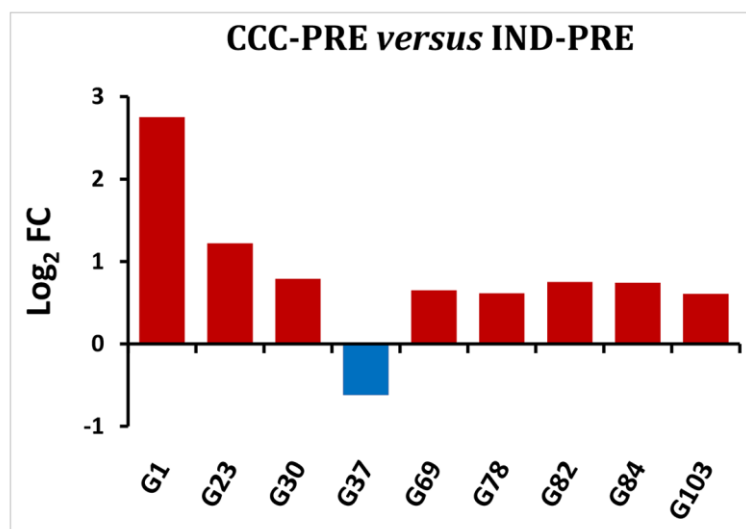


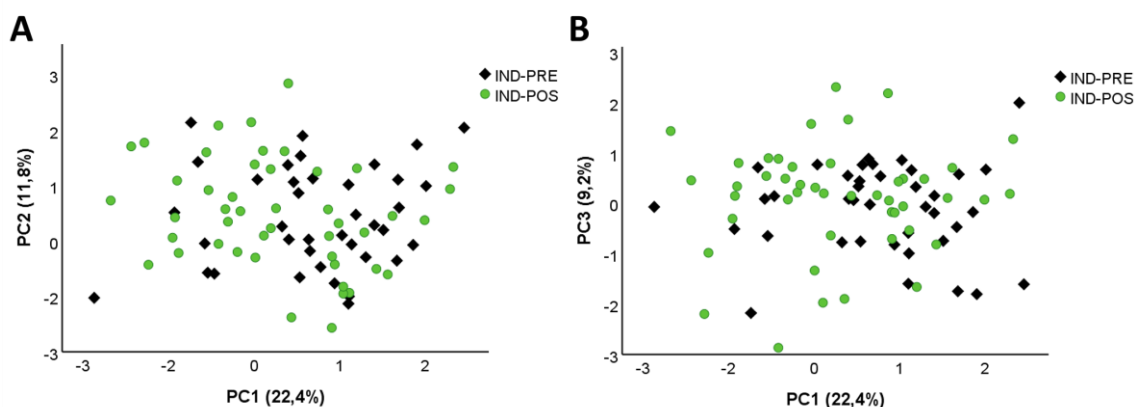
Figura R35. Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes con enfermedad de Chagas en fase cardiaca (CCC) y en fase indeterminada (IND). Cada barra del gráfico corresponde al gen al que se hace referencia en la parte inferior del gráfico (eje X). El eje Y representa el log<sub>2</sub> del *fold change* (Log<sub>2</sub> FC) para cada gen. FC es el cociente de la expresión génica media en el grupo de pacientes CCC entre la obtenida en los pacientes IND. Los valores positivos (barras rojas) indican genes sobreexpresados y los valores negativos (barras azules) indican genes infraexpresados en CCC en comparación con el grupo de pacientes IND.

### B.2.3. EXPRESIÓN DIFERENCIAL DE GENES DE RESPUESTA INMUNE EN PACIENTES CON ENFERMEDAD DE CHAGAS POR EFECTO DEL TRATAMIENTO CON BENZNIDAZOL

Con el fin de identificar los cambios de expresión en genes implicados en la respuesta inmune por efecto del tratamiento con benznidazol, se analizó la expresión de los 106 genes a estudio de forma comparativa en pacientes con enfermedad de Chagas pre y post-tratamiento. Dado que la eficacia del tratamiento varía según el estadio de la enfermedad en que se encuentre el paciente en el momento de su administración, el estudio se llevó a cabo tanto en pacientes en fase indeterminada (IND) como en pacientes con sintomatología cardiaca (CCC) pre y post-tratamiento. Tal y como se ha descrito en apartados anteriores, se emplearon los valores de expresión génica relativa normalizada (NRQ) de las células PBMC de estos pacientes, estimuladas con antígenos de *T. cruzi*, para llevar a cabo los análisis comparativos.

### B.2.3.1. Identificación de genes con expresión diferencial entre pacientes crónicos en fase indeterminada (IND) pre y post-tratamiento

Como primera aproximación se realizó un análisis de componentes principales para conocer la distribución del conjunto de datos en base a sus valores de expresión. Así, se emplearon los valores NRQ del grupo de pacientes IND pre-tratamiento (IND-PRE, n=39) y del grupo IND post-tratamiento (IND-POS, n=46) para obtener los gráficos que se muestran en la **Figura R36**. La componente principal 1 (PC1) y la componente principal 2 (PC2) acumulan los mayores porcentajes de varianza total con valores de un 22,4% y 11,8%, respectivamente (**Figura R36A**). La componente principal 3 (PC3) que explica el 9,2% de la varianza, también se representó junto a PC1 (**Figura R36B**).



**Figura R36.** Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes en fase indeterminada pre-tratamiento (IND-PRE, rombos negros) y pacientes en fase indeterminada post-tratamiento (IND-POS, círculos verdes). **(A)** Gráfico PCA de puntuación de las componentes principales 1 (PC1) y 2 (PC2) en los ejes X e Y, respectivamente. **(B)** Gráfico PCA de puntuación de las componentes principales 1 (PC1) y 3 (PC3) en los ejes X e Y, respectivamente. La proporción de la varianza que explica cada componente principal se indica como un porcentaje en el eje junto a la componente principal correspondiente.

Los resultados mostraron que las puntuaciones obtenidas para la PC1 en el grupo de pacientes pre-tratamiento eran en su gran mayoría positivas, mientras que en el grupo post-tratamiento se distribuían a lo largo de todo el eje X, donde se representaba esta componente. Estas diferencias observadas en las puntuaciones adquiridas para la PC1 entre ambos grupos de individuos se confirmaron mediante una prueba T de Student que indicó que existía un perfil de expresión diferencial estadísticamente significativo entre los pacientes IND pre y post-tratamiento ( $p = 0,025$ ). Adicionalmente, esta misma prueba se aplicó para corroborar que los genes que se correlacionaban con la PC2 y PC3 no presentaba diferencias de expresión

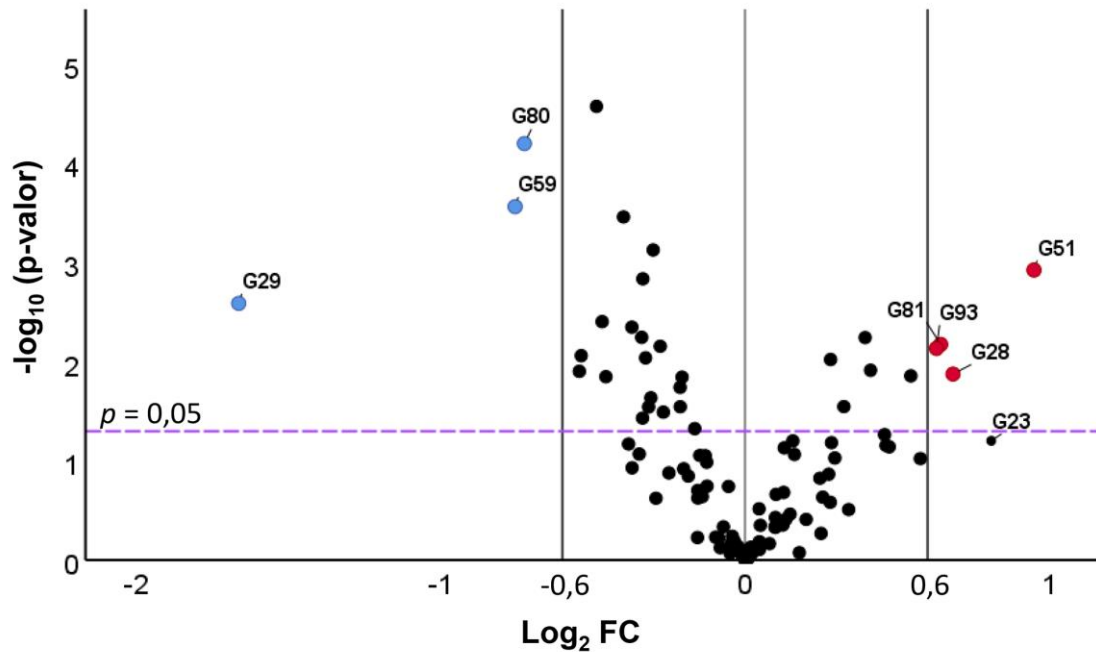
## RESULTADOS

estadísticamente significativas entre ambos grupos (PC2  $p = 0,945$ , PC3  $p = 0,354$ ). La PC1 depende de la expresión de 18 genes, todos ellos correlacionados positivamente con alta carga factorial con PC1 (**Tabla R29**).

**Tabla R29. Genes con carga factorial para la componente principal 1 (PC1) superior a 0,54 o inferior a -0,54 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes crónicos en fase indeterminada pre y post-tratamiento.**

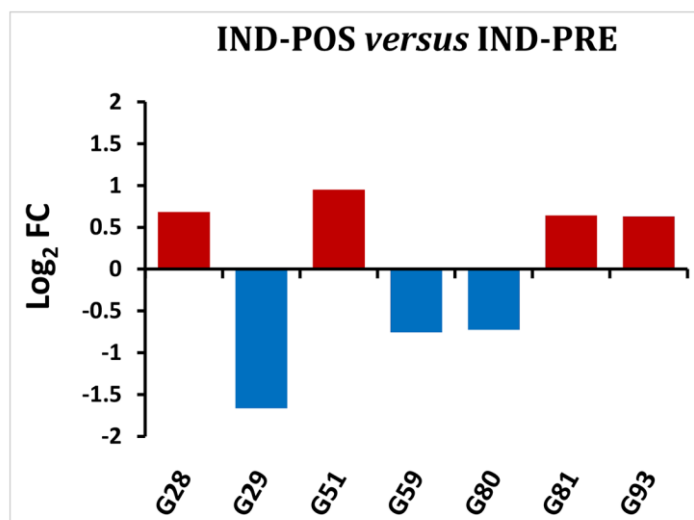
Gen	Carga factorial para PC1	Gen	Carga factorial para PC1
G8	0,638	G52	0,819
G9	0,843	G54	0,684
G11	0,719	G59	0,556
G12	-0,705	G62	0,802
G16	0,866	G70	0,598
G18	0,583	G71	0,541
G25	-0,708	G75	0,682
G26	-0,570	G77	0,809
G31	0,728	G80	0,544
G33	0,681	G81	-0,646
G37	0,564	G87	0,769
G38	-0,640	G92	0,575
G39	0,622	G93	-0,684
G44	0,844	G94	0,715
G46	0,893	G95	0,669
G47	0,693	G101	-0,629

A continuación, para determinar la influencia del tratamiento, se analizó de forma comparativa la expresión génica de los 106 genes en los pacientes IND tratados y no tratados con benznidazol. El gráfico de volcán (**Figura R37**) construido a partir de los valores de *fold change* de la expresión génica (FC) y la significancia estadística para cada gen, reveló que 7 de los 106 genes analizados se encontraban diferencialmente expresados entre los dos grupos de pacientes con significancia estadística ( $-\text{Log}_{10} p\text{-valor} > 1,3$  equivalente a  $p < 0,05$ ). En concreto, 4 genes se sobreexpresaban en los pacientes IND tratados (IND-POS) *versus* pacientes IND no tratados (IND-PRE) ( $\text{Log}_2 \text{FC} > 0,6$  correspondiente a  $\text{FC} > 1,5$ ; puntos rojos). Por el contrario, 3 de los genes estaban infraexpresados en IND-POS en comparación con aquellos que no habían recibido tratamiento (IND-PRE) ( $\text{Log}_2 \text{FC} < -0,6$ , correspondiente a  $\text{FC} < 0,66$ ; puntos azules).



**Figura R37. Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes con enfermedad de Chagas en fase indeterminada pre-tratamiento (IND-PRE, n=39) y pacientes en fase indeterminada post-tratamiento (IND-POS, n=46) representado en gráfico de volcán.** El eje X representa el  $\log_2$  del *fold change* de expresión entre IND-POS e IND-PRE ( $\log_2$  FC), donde el FC se calcula como la ratio entre los dos grupos (IND-POS/IND-PRE). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ( $-\log_{10}$  p-valor). La línea horizontal violeta indica el punto de corte para el valor de significancia estadística  $p = 0,05$ . Las líneas verticales negras representan el  $\log_2$  de FC de  $-0,6$  y  $0,6$  (correspondiente a FC de  $0,66$  y  $1,5$ , respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre pacientes IND-PRE e IND-POS.

Una vez identificados los genes cuya expresión se vio modificada por el tratamiento, se cuantificaron las diferencias de expresión génica detectadas, representando los valores de  $\log_2$  del *fold change* obtenidos para cada gen en un gráfico de barras (**Figura R38**). Los resultados indicaban que aquellos genes sobreexpresados ( $\log_2$  FC  $> 0,6$  (barras rojas)) poseían valores de  $\log_2$  FC bastante similares, experimentando todos, por tanto, el mismo nivel de cambio de expresión. De manera opuesta, entre los genes infraexpresados ( $\log_2$  FC  $< -0,6$  (barras azules)), se identificó que uno de ellos mostraba un menor valor de  $\log_2$  FC, inferior a  $-1,5$ , lo que equivale, al menos, a 3 veces menor expresión de dicho gen en los pacientes IND post-tratamiento respecto al grupo de pacientes IND pre-tratamiento.



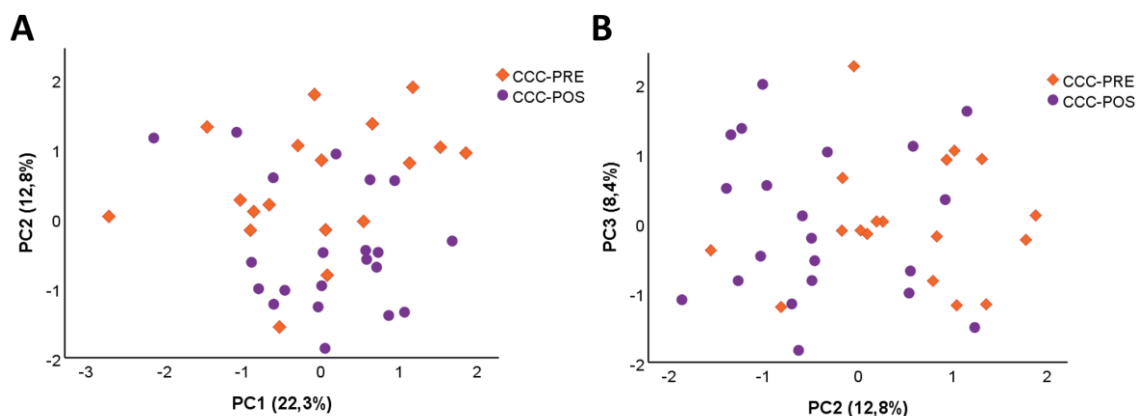
**Figura R38. Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes con enfermedad de Chagas en fase indeterminada pre-tratamiento (IND-PRE) y pacientes en fase indeterminada post-tratamiento (IND-POS).** Cada barra del gráfico corresponde al gen al que se hace referencia en la parte inferior del gráfico (eje X). El eje Y representa el log<sub>2</sub> del *fold change* (Log<sub>2</sub> FC) para cada gen. FC es el cociente de la expresión génica media en el grupo de pacientes IND post-tratamiento entre la obtenida en el grupo IND pre-tratamiento. Los valores positivos (barras rojas) indican genes sobreexpresados y los valores negativos (barras azules) indican genes infraexpresados en IND-POS en comparación con el grupo de pacientes IND-PRE.

### B.2.3.2. Identificación de genes con expresión diferencial entre pacientes crónicos en fase cardíaca (CCC) pre y post-tratamiento

Los cambios en la expresión de los 106 genes a estudio tras el tratamiento con benznidazol también se estudiaron en pacientes crónicos con enfermedad de Chagas en fase cardíaca. Al igual que en las comparaciones anteriores, se realizó un análisis PCA como primera aproximación para conocer la estructura del conjunto de datos. Este análisis se aplicó sobre los valores NRQ de los pacientes con enfermedad de Chagas en fase cardíaca no tratados (CCC-PRE, n=18), así como en el grupo de pacientes CCC que habían recibido el tratamiento (CCC-POS, n=21). En la **Figura R39** se muestran las puntuaciones adquiridas en cada muestra para cada componente principal, explicando la componente principal 1 (PC1) y la componente principal 2 (PC2) un 22,3% y un 12,8% de la varianza total, respectivamente (**Figura R39A**). La componente principal 3 (PC3) que explica el 8,4% de la varianza, también se representó junto a PC1 (**Figura R39B**). Los resultados revelaron que si bien los datos de cada grupo no se distribuían de forma diferencial cuando se representaba la PC1 o la PC3, sí que tendían a agruparse la mayoría de las muestras de cada grupo a lo largo del eje donde se representaba la PC2. Estas diferencias observadas en las



puntuaciones adquiridas para la PC2 entre los pacientes cardiacos pre y post-tratamiento se confirmaron mediante una prueba T de Student que indicó que existía un perfil de expresión diferencial estadísticamente significativo entre ambos grupos de pacientes para los genes correlacionados con la PC2 ( $p = 0,03$ ). Este mismo test estadístico se aplicó sobre las puntuaciones adquiridas para la PC1 y PC3, confirmando que no participaban en las diferencias observadas (PC1  $p = 0,641$ , PC3  $p = 0,914$ ). La expresión de 15 genes de los 106 analizados se correlacionaba con alta carga factorial con la PC2 (**Tabla R30**). En concreto, la correlación con PC2 detectada para 14 de estos genes fue positiva (CF > 0,54) mientras que para el gen restante fue negativa (CF < -0,54).



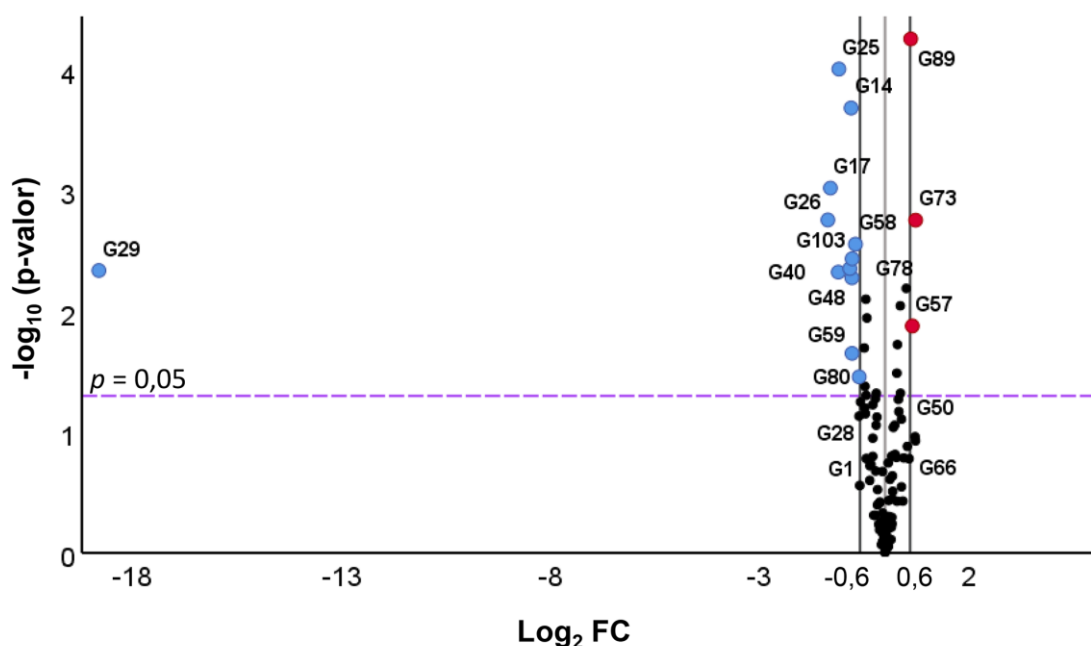
**Figura R39. Análisis de componentes principales (PCA) aplicado sobre valores NRQ de los 106 genes analizados de pacientes crónicos en fase cardiaca pre-tratamiento (CCC-PRE, rombos naranjas) y pacientes en fase cardiaca post-tratamiento (CCC-POS, círculos violetas).** (A) Gráfico PCA de puntuación de las componentes principales 1 (PC1) y 2 (PC2) en los ejes X e Y, respectivamente. (B) Gráfico PCA de puntuación de las componentes principales 2 (PC2) y 3 (PC3) en los ejes X e Y, respectivamente. La proporción de la varianza que explica cada componente principal se indica como un porcentaje en el eje junto a la componente principal correspondiente.

**Tabla R30. Genes con carga factorial para la componente principal 2 (PC2) superior a 0,54 o inferior a -0,54 obtenido en análisis de componentes principales sobre los valores NRQ de pacientes crónicos en fase cardiaca pre y post-tratamiento.**

Gen	Carga factorial para PC2	Gen	Carga factorial para PC2
G5	0,633	G43	0,747
G10	0,623	G58	0,741
G17	0,729	G59	0,708
G18	0,568	G70	0,642
G34	0,879	G76	0,625
G35	0,625	G80	0,787
G40	0,680	G89	-0,680
G42	0,593		

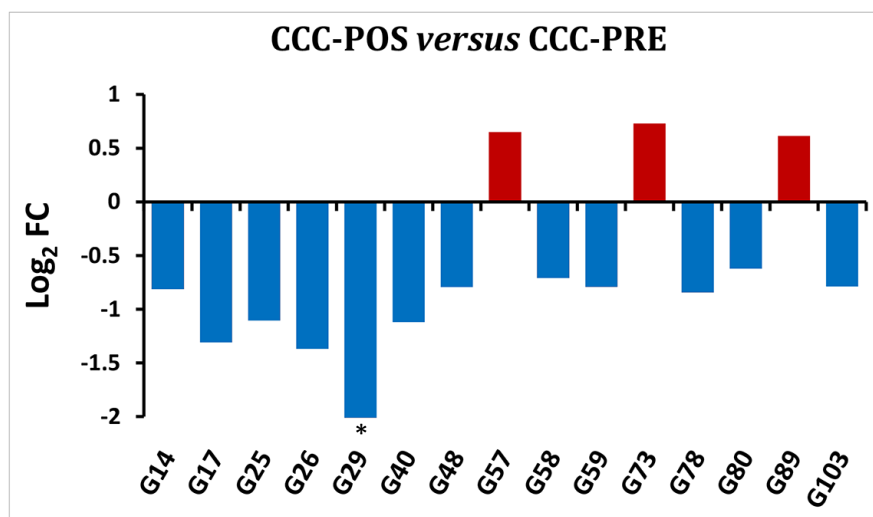
## RESULTADOS

Cuando se determinó la expresión génica diferencial de los 106 genes a estudio en el grupo de pacientes CCC pre-tratamiento y post-tratamiento, se identificaron 15 genes cuyo nivel de expresión fue diferencial entre ambos grupos con significancia estadística (**Figura R40**). El  $\log_2$  del FC de cada gen se representó en un gráfico de volcán junto con la significancia estadística que reveló que, de estos 15 genes expresados diferencialmente, 12 estaban infraexpresados ( $\log_2 \text{FC} < 0,6$ ) en el grupo de pacientes CCC post-tratamiento en comparación con el grupo CCC pre-tratamiento. Por otro lado, los 3 genes restantes se encontraban sobreexpresados en estos pacientes post-tratamiento ( $\log_2 \text{FC} > 0,6$ ). Las diferencias de expresión observadas en cada uno de ellos poseían un valor de  $-\log_{10}$  p-valor superior a 1,3, equivalente a un  $p < 0,05$ .



**Figura R40. Análisis comparativo del nivel de expresión génica diferencial de los 106 genes analizados en pacientes crónicos con enfermedad de Chagas en fase cardiaca pre-tratamiento (CCC-PRE, n=18) y pacientes en fase cardiaca post-tratamiento (CCC-POS, n=21) representado en gráfico de volcán.** El eje X representa el  $\log_2$  del *fold change* de expresión entre CCC-POS y CCC-PRE ( $\log_2 \text{FC}$ ), donde el FC se calcula como la ratio entre los dos grupos (CCC-POS/CCC-PRE). El eje Y corresponde a la significancia estadística, expresada como el logaritmo negativo del valor  $p$  ( $-\log_{10}$  p-valor). La línea horizontal violeta indica el punto de corte para el valor de significancia estadística  $p = 0,05$ . Las líneas verticales negras representan el  $\log_2$  de FC de  $-0,6$  y  $0,6$  (correspondiente a FC de  $0,66$  y  $1,5$ , respectivamente) los cuales se usaron como umbrales biológicos para identificar genes expresados diferencialmente. Los valores negativos corresponden a genes infraexpresados (puntos azules) y los valores positivos a los genes sobreexpresados (puntos rojos). Los puntos negros comprendidos entre los umbrales establecidos representan genes cuya expresión no es diferencial entre pacientes CCC-PRE y CCC-POS.

Finalmente, se representó en un gráfico de barras los valores de  $\log_2$  FC de aquellos genes expresados diferencialmente en pacientes CCC post-tratamiento *versus* pacientes CCC pre-tratamiento (**Figura R41**). Si bien todos superaban el umbral biológico establecido ( $\log_2$  FC > 0,6 o < -0,6) con significancia estadística, en el caso de ciertos genes las diferencias en el nivel de expresión fueron superiores al resto. Concretamente, fueron 5 genes infraexpresados en CCC post-tratamiento en los que se detectó un valor de  $\log_2$  FC inferior a -1, correspondiente a un *fold change* de, al menos, 0,5. De esta forma, la expresión de estos 5 genes se reducía, al menos, a la mitad en el grupo de pacientes CCC post-tratamiento *versus* CCC pre-tratamiento. Por último, cabe destacar que para uno de los genes (G29) no se detectó ningún valor de expresión en el grupo de pacientes CCC post-tratamiento.



**Figura R41. Genes diferencialmente expresados y con diferencias estadísticamente significativas entre pacientes crónicos con enfermedad de Chagas en fase cardiaca pre-tratamiento (CCC-PRE) y pacientes en fase cardiaca post-tratamiento (CCC-POS).** Cada barra del gráfico corresponde al gen al que se hace referencia en la parte inferior del gráfico (eje X). El eje Y representa el  $\log_2$  del *fold change* ( $\log_2$  FC) para cada gen. FC es el cociente de la expresión génica media en el grupo de pacientes CCC post-tratamiento entre la obtenida en el grupo CCC pre-tratamiento. Los valores positivos (barras rojas) indican genes sobreexpresados y los valores negativos (barras azules) indican genes infraexpresados en CCC-POS en comparación con el grupo de CCC-PRE. \*Dado que el gen G29 no mostró ningún valor de expresión en el grupo CCC-POS, para aplicar fórmulas, el valor 0 se reemplazó por  $10^{-6}$ , obteniéndose un valor de  $\log_2$  FC = -18,8.



## **DISCUSIÓN**

---



## 1. VARIACIONES EN EL CONTENIDO GÉNICO Y CONTENIDO DE RETROELEMENTOS PORTADORES DE LA HUELLA Pr77. ANÁLISIS COMPARATIVO ENTRE 6 CEPAS DE *T. cruzi* DE DIFERENTE DTU

Las diferentes cepas de *T. cruzi* causantes de la enfermedad de Chagas presentan variaciones fenotípicas, así como diferentes comportamientos en términos de tropismo tisular, virulencia y fisiopatología, induciendo, consecuentemente, respuestas inmunológicas heterogéneas en los hospedadores (Franzen et al., 2011; Rassi et al., 2012; Revollo et al., 1998; Rodriguez et al., 2014). Se ha descrito que cepas de *T. cruzi* de diferentes DTU, así como de la misma DTU, muestran diferencias en el número de cromosomas, tamaño del genoma y contenido de ADN (Lewis et al., 2009; Vargas et al., 2004). Las diferencias en el contenido génico y el número de copias de genes diana empleados para la detección del parásito y la enfermedad (Duffy et al., 2009; Ramírez et al., 2015), junto con las diferencias en la susceptibilidad a los fármacos (Teston et al., 2013), dificultan enormemente el seguimiento diagnóstico de los pacientes y el desarrollo de nuevos fármacos y vacunas. Así, la caracterización genética de las diferentes cepas, facilitará el manejo de los pacientes y el control de la enfermedad. Con esta premisa y para contribuir al conocimiento sobre la variabilidad genética de *T. cruzi*, se secuenciaron y analizaron los genomas de distintas cepas de *T. cruzi* pertenecientes a diferentes DTUs.

Los retrotransposones constituyen gran parte del genoma de eucariotas, pero representan una limitación para el ensamblaje dificultando este debido a su gran longitud y alto número de copias. Es por ello que, en muchos casos, los elementos transponibles no están bien anotados. Con el objetivo de mejorar el conocimiento de los elementos móviles en *T. cruzi*, se analizó el contenido de los retroelementos L1Tc y NARTc, que portan la huella Pr77, en los genomas secuenciados de *T. cruzi* B. M. López (TcIa), Dm28 (TcId), Y (TcII), Ikiakarora (TcIII), SOL (TcV) y CL Brener (TcVI) por haber sido estas cepas aisladas en diferentes zonas geográficas, endémicas y no endémicas, de Venezuela, España, Colombia y Brasil, a partir de diferentes mamíferos huéspedes (*Homo sapiens*, *Didelphis marsupialis*) e insectos vectores (*Triatoma infestans*, *Rhodnius prolixus*) (Hamuy et al., 2013; Higuera et al., 2013; Murcia et al., 2013; Rodríguez et al., 1998).

Tras la secuenciación y ensamblaje, el análisis de los parámetros obtenidos reveló características similares respecto a las lecturas crudas, lo que permitió realizar estudios comparativos entre las seis cepas. Así, las lecturas filtradas por calidad se ensamblaron generando desde 5.923 (cepa B. M. López) hasta 11.946 (cepa SOL) *contigs*. El tamaño de estos osciló entre un tamaño mínimo de 500 pb para todas las cepas, que fue el umbral establecido *a priori* en el proceso de ensamblaje, y un tamaño máximo de 45.876 pb obtenido para la cepa B. M. López.

Las secuencias repetidas en tándem constituyen más del 50% del genoma de *T. cruzi*, cifra muy superior a la que estas representan en otros genomas de tripanosomátidos como *Leishmania* (20%) y *T. brucei* (25%) (Berriman et al., 2005; El-Sayed, Myler, Bartholomeu, et al., 2005; Ivens et al., 2005), debiéndose, principalmente, a grandes familias multigénicas de proteínas de superficie, retrotransposones y repeticiones subteloméricas. Este contenido multigénico repetitivo, junto con la existencia de cepas híbridas y aneuploidías, hace de *T. cruzi* un organismo con una arquitectura genómica compleja que dificulta la integridad de los ensamblajes genómicos (Reis-Cunha et al., 2015; Reis-Cunha and Bartholomeu, 2019; Vargas et al., 2004). A pesar de la limitación que supone el elevado número de *contigs* generados derivados de la naturaleza de los genomas y la secuenciación mediante Ion Torrent PGM™ de las 6 cepas de *T. cruzi* a estudio, el análisis BUSCO de la integridad de los genomas ensamblados reveló un alto porcentaje de genes BUSCO identificados completos y de una sola copia en todas las cepas, oscilando del 76,2% (SOL) al 96,9% (cepas B. M. López y Dm28). Este hecho, junto con el pequeño número de BUSCOs duplicados, sugirió que los haplotipos fueron correctamente ensamblados y con una buena calidad. Para la anotación de los genomas, se llevó a cabo la predicción de genes para lo que se empleó un genoma de referencia para cada cepa. Las cepas de referencia seleccionadas fueron aquellas con las que tenían el mayor número de *best hits*, excepto en el caso de B. M. López, en el que se eligió Dm28c, con 3.402 *best hits*, como genoma de referencia en lugar de Sylvio X10 (4.266 *best hits*) dada la pobre calidad del genoma disponible públicamente en la base de datos de tripanosomátidos TritrypDB.



El contenido de G+C de cada genoma fue cercano al 50% en todas las cepas y el número de genes predichos en cada genoma osciló entre 7.661 (cepa B. M. López) y 12.829 (cepa SOL), lo que representa entre el 62,9% y el 77,7% del total de pb de cada genoma ensamblado. Cabe destacar que el mayor número de genes predichos se identificó en las cepas CL Brener (11.909) y SOL (12.829), las dos cepas híbridas con el mayor tamaño estimado (55 y 59,5 Mb, respectivamente). La longitud media de los genes identificados osciló entre 1.191 (Ikiakarora) y 1.518 pb (B. M. López). Según la anotación asignada, una gran proporción de los genes identificados en los seis genomas tenían identidad con proteínas de función conocida previamente identificadas en otros genomas. Adicionalmente, aproximadamente un tercio de los genes predichos identificados en cada genoma codificaban proteínas hipotéticas.

El análisis del contenido en retrotransposones portadores de Pr77 en los genomas mostró que estos estaban altamente representados en todas las cepas y eran particularmente abundantes en las cepas Ikiakarora y CL Brener. Además, la proporción de L1Tc era, al menos, 4 veces mayor que la de los elementos NARTc en todas las cepas. Esta proporción podría estar relacionada con el carácter potencialmente autónomo del retrotransposón L1Tc, que podría ser utilizado por el NARTc no autónomo para la transmovilización. Además, la elevada proporción de elementos L1Tc podría relacionarse con la necesidad de las propiedades de reparación de ADN asociadas a NL1Tc, proteína codificada por el ORF1 de L1Tc, que ha demostrado estar dotada de actividad endonucleasa AP, así como 3'-fosfatasa y 3'-fosfodiesterasa (Olivares et al., 1999). Además de las implicaciones que estas actividades tienen en el primer paso del proceso de integración de estos retrotransposones, se ha demostrado *in vivo* que estas funciones enzimáticas conducen a una reducción de aproximadamente el 60% de los daños en el ADN causados por el tratamiento de los parásitos con daunorubicina, proporcionando protección contra los efectos negativos de este fármaco y la radiación gamma (6 o 9 Gy) sobre la supervivencia y la tasa de crecimiento del parásito (Olivares et al., 2003).

Aunque la huella Pr77 fue una de las dos sondas utilizadas para identificar los elementos L1Tc y NARTc en los seis genomas, los análisis posteriores indicaron que no se encontraron retrotransposones que carecieran de Pr77. En cambio, en los

seis genomas de *T. cruzi* se identificaron huellas Pr77 no asociadas a retrotransposones, que se denominaron *single Pr77 hallmarks* (Pr77Sh). La secuencia Pr77 se describió originalmente en el extremo 5' de L1Tc y se demostró que era capaz de activar la transcripción de genes que generaban transcritos abundantes vía ARN polimerasa II (Heras et al., 2007). Además, Pr77 constituye un sistema dual promotor/ribozima, que funciona a nivel de ADN/ARN como promotor y ribozima tipo HDV, respectivamente (Sánchez-Luque et al., 2012). Así, se sugirió que los retrotransposones portadores de la huella Pr77 pueden ser responsables de la expansión de estas funciones a través de los genomas de tripanosomátidos (Sánchez-Luque et al., 2014). En la presente Tesis también describimos que una proporción importante de elementos L1Tc y NARTc se asocian frecuentemente a TS, RHS y MASP. Además, este estudio mostró diferencias entre las cepas analizadas siendo la cepa SOL aquella con mayor número de retroelementos asociados a TS. Llamativamente, la mayor parte de ellos eran Pr77Sh. La presencia de secuencias activas sugiere que la asociación de las huellas Pr77 (como parte de retrotransposones o como Pr77Sh) y estas secuencias particulares puede tener un sentido funcional, probablemente relacionado con un papel regulador en los patrones de expresión génica que ha permitido que estas asociaciones coevolucionen juntas y se mantengan durante la evolución. El grado relativamente alto de dispersión que exhiben los elementos L1Tc y NARTc no codificantes, pero transcripcionalmente activos, refuerza la idea de que estos retrotransposones no LTR desempeñan un papel regulador en la expresión génica y son responsables de la transcripción de genes adyacentes y policistrones, tal y como se ha sugerido previamente (Heras et al., 2007). Estudios anteriores basados en la distribución genómica de los retrotransposones L1Tc y NARTc en el genoma de *T. cruzi*, revelaron la existencia de concatémeros de estos elementos insertados en secuencias no codificantes, así como asociados a secuencias codificantes de genes, particularmente el que codifica trans-sialidasa en la cepa CL Brener (Olivares et al., 2000). Esta asociación se consideró un indicio más de la domesticación de los retrotransposones por parte de los genomas del hospedador que, por su alto contenido en elementos móviles, están lejos de ser sistemas de información estáticos. Una vez adquiridos, estos retrotransposones se hicieron lo suficientemente necesarios para persistir y formar parte de los genomas.

Las familias de multigenes más abundantes codifican proteínas de superficie, como las mucinas, TS, MASP y RHS (El-Sayed, Myler, Bartholomeu, et al., 2005). En los seis genomas estudiados se analizó el número de copias de genes que codifican estas proteínas junto con otros con un papel biológico relevante en tripanosomátidos, así como genes de expresión constitutiva. Los resultados obtenidos mostraron diferencias en el número de copias de genes asociados a procesos de virulencia, pero no en el contenido de genes constitutivos, entre los seis genomas de *T. cruzi*. El número de copias de los genes que codifican MASP, TS y TASV varió mucho entre las cepas siendo más abundantes en las cepas Y (TcII), SOL (TcV) y CL Brener (TcVI) respecto a las demás analizadas. Estudios de estas tres familias de proteínas sugieren que están relacionadas con la interacción hospedador-parásito, dado que todas ellas se han identificado en el estadio tripomastigote en el que el parásito interactúa con el hospedador mamífero (Bartholomeu et al., 2009; Bernabó et al., 2013; Caeiro et al., 2018; de Pablos et al., 2011; El-Sayed, Myler, Bartholomeu, et al., 2005; Frasc, 2000). Las TS constituyen uno de los principales antígenos presentes en el parásito (Frasc, 2000) y la familia MASP fue descrita por primera vez en el genoma de *T. cruzi* CL Brener (El-Sayed, Myler, Bartholomeu, et al., 2005). Varios estudios sugieren que las MASP pueden desempeñar un papel importante en procesos de evasión inmunitaria y de interacción huésped-parásito, presentando una alta variabilidad entre las cepas de *T. cruzi* (Bartholomeu et al., 2009; de Pablos et al., 2011; dos Santos et al., 2012; Seco-Hidalgo et al., 2015). CL Brener e Y fueron también las cepas con mayor contenido de RHS, familia génica asociada a la inserción de elementos móviles (Bringaud, Biteau, et al., 2002). Las cepas Dm28 y B. M. López, ambas pertenecientes a DTU I, contenían la menor frecuencia de mucinas, con grandes diferencias respecto al resto de cepas analizadas. Estas cepas también fueron aquellas con el menor contenido de genes codificantes para TS, MASP, TASV, RHS, cisteína peptidasas y quinesinas. De manera interesante, un estudio en *T. rangeli*, parásito estrechamente relacionado con *T. cruzi* pero que no tiene capacidad de infectar células, reveló que su genoma tiene menos copias de la familia multigénica MASP, TS y mucinas en comparación con *T. cruzi*. *T. rangeli* no resulta patógeno para el huésped mamífero, lo cual puede ser el resultado de la reducción de estas familias multigénicas, destacando el importante

papel que desempeñan en los procesos de invasión celular (Reis-Cunha and Bartholomeu, 2019; Stoco et al., 2014).

La secuenciación de estos genomas contribuye al actual repositorio de *T. cruzi*, revelando los hallazgos descritos la variabilidad entre las diferentes cepas de este complejo parásito en el que Pr77 parece tener un papel funcional relevante.

### **2. CONTEXTO GENÓMICO Y PREDICCIÓN DE ESTRUCTURA SECUNDARIA DE TIPO RIBOZIMA HDV EN SECUENCIAS Pr77Sh. ESTUDIO DE SU FUNCIÓN CATALÍTICA**

El análisis del contenido de retroelementos en el genoma de las 6 cepas de *T. cruzi* a estudio permitió la identificación de huellas Pr77 que no estaban asociadas a retroelementos. Dado que asociada a Pr77 de L1Tc estaba descrita una capacidad promotora, a nivel de ADN, y actividad ribozima de tipo HDV, a nivel de ARN, la primera cuestión a abordar era si estas huellas Pr77 no asociadas a retroelementos, denominadas Pr77Sh, eran activas y conservaban su función dual o no lo eran. En caso de ser funcionales podrían desempeñar un papel funcional importante en la regulación de la expresión génica del parásito. Es por ello que nuestro interés se centró en analizar los sitios de inserción de las huellas Pr77Sh en los genomas que las hospedaban, determinar su estructura secundaria y si esta es compatible con función ribozima HDV como la asociada a L1TcRz y, en su caso, la función asociada a dicha estructura.

La secuencia nucleotídica de las diferentes Pr77Sh identificadas respecto a la de la secuencia Pr77 de L1Tc/NARTc varió entre cepas. Así, se detectaron Pr77Sh con una identidad de hasta el 85,7% en CL Brener, 98,7% en B. M. López y 100% en la cepa SOL. Por otro lado, la secuencia Pr77 de L1Tc contiene el motivo DPE, descrito como motivo importante implicado en la actividad promotora de Pr77 en L1Tc (Macías et al., 2016). A excepción de Dm28, en todas las cepas analizadas se encontraron Pr77Sh con el motivo DPE conservado en términos de composición de la secuencia y localización respecto al sitio de inicio de la transcripción, con la mayor representación en la cepa Y.

La identificación de las secuencias flanqueantes a las Pr77Sh reveló sitios de inserción comunes y otros específicos de alguna/s cepas. De este modo, en todas las

cepas se identificaron Pr77Sh insertadas en genes codificantes de trans-sialidasas (TS), siendo SOL (42,4%) y CL Brener (37,9%) las cepas donde existían mayor número de Pr77Sh insertadas en este lugar. La familia de genes codificantes de las TS está presente en gran número de copias en el genoma de *T. cruzi* y se ha descrito como un importante factor de virulencia del parásito (Chain et al., 2019; El-Sayed, Myler, Bartholomeu, et al., 2005; San Francisco et al., 2017). La abundancia detectada de las huellas Pr77Sh insertadas en TS es consistente con otros hallazgos incluidos en esta Tesis, como la alta representación de TS en los genomas de las 6 cepas de *T. cruzi* estudiadas, así como la asociación de TS con L1Tc, NARTc y huellas Pr77Sh. La inserción de Pr77Sh en el gen ARNr 5S fue también frecuente, representando en las cepas Ikiakarora y CL Brener hasta el 44,2% y el 43,5% de las Pr77Sh identificadas en estas cepas, respectivamente. Inserciones de Pr77Sh en ARNr 5S también se detectaron en las cepas SOL e Y, aunque con menos frecuencia, pero, en cambio, no se detectaron inserciones en este gen ni en la cepa B. M. López ni en Dm28. El genoma de *T. cruzi* contiene unas 1.600 copias de la familia génica codificante de ARNr 5S por núcleo diploide, a diferencia de otros protozoos parásitos como *Leishmania major*, que sólo contiene 11 copias de ARNr 5S en su genoma (Hernández-Rivas et al., 1992; Moreno-Campos et al., 2016). Dentro de cada una de las cepas, se han detectado huellas Pr77Sh idénticas tanto en composición de secuencia como en el sitio de inserción considerando las secuencias del genoma que las flanquean. De manera interesante, también se han identificado huellas Pr77Sh idénticas insertadas en los mismos genes en diferentes cepas. Esta conservación del mismo entorno genómico de la huella Pr77Sh entre distintas cepas pertenecientes a diferentes DTUs sugiere que estas inserciones se adquirieron antes de la divergencia de las cepas a lo largo de la evolución. Interesantemente, se detectaron sitios de inserción de huellas Pr77Sh específicos de cepa. Así, un 45,3% de las Pr77Sh identificadas en B. M. López se insertaron en elementos VIPER (*vestigial interposed retroelement*). Los VIPER son retroelementos LTR que codifican tirosina recombinasa que están presentes en los genomas de *Trypanosoma cruzi*, *Trypanosoma brucei* y *Trypanosoma vivax*, pero no en *Leishmania spp* (Lorenzi et al., 2006; Vázquez et al., 2000). Cabe destacar que la huella Pr77 de todas las copias Pr77Sh insertadas en VIPER fueron idénticas a nivel de secuencia, conservando todas ellas la misma identidad con la huella Pr77 de L1Tc/NARTc.

En el caso de la cepa SOL, secuenciada por Ion Torrent y PacBio, la mayoría de estas huellas Pr77Sh también se localizaron en el genoma secuenciado por PacBio, análisis que además reveló la existencia de más de una copia de una misma Pr77Sh localizadas en diferentes *contigs*, indicativo de la existencia de más de una copia de cada una de ellas y su inserción en diferentes copias de genes a lo largo del genoma. La inserción de estas huellas, principalmente, en genes que son relevantes para el parásito sugiere el papel funcional de estas secuencias desde el punto de vista de regulación de la expresión génica.

Como se ha comentado previamente, Pr77 constituye un sistema dual promotor/ribozima, con actividad promotora a nivel de ADN y ribozima de tipo HDV a nivel de ARN (Sánchez-Luque et al., 2012). La implicación de estas actividades en la regulación de la expresión génica de los parásitos que las hospedan podría garantizar la escisión y liberación adecuada de una copia activa del elemento que se cotranscriba como parte de un ARN policistrónico, a la vez que contribuiría al mantenimiento de sus funciones de promotor interno y ribozima; garantizar la transcripción de las copias presentes en orientación antisentido con respecto a los *clusters* policistrónicos y las que se encuentran fuera de ellos; mediar la iniciación de la ARN polimerasa II para la transcripción de grandes *cluster* policistrónicos dada la acumulación de elementos móviles que portan Pr77 en las SSR (*Strand Switch Region*); así como podría prevenir un decaimiento del nivel de transcripción de las regiones distales, asegurando la correcta expresión de los últimos genes de cada *cluster*. Además, las huellas Pr77 de retroelementos insertados en la región 3' UTR de genes somáticos podrían estar regulando a la baja el nivel de ARNm reduciendo la estabilidad del ARN tras el corte de L1TcRz (Macías et al., 2018; Sánchez-Luque et al., 2012).

Para determinar si las Pr77Sh conservaban su actividad ribozima a nivel de ARN, se determinó la estructura secundaria de todas las huellas Pr77Sh identificadas en todas las cepas y se categorizaron en cuatro niveles (I-IV) dependiendo de su similitud o divergencia respecto a la estructura secundaria adoptada por la huella L1TcRz. En Ikiakarora, Y, SOL y CL Brener, se identificó un gran porcentaje de huellas Pr77Sh que presentaban un plegamiento compatible con L1TcRz (tipo I y II). Por el contrario, en la cepa B. M. López, más del 67% de las

huellas no poseían un plegamiento compatible con el de L1TcRz, así como en Dm28, donde el 100% de las Pr77Sh tenían estructura IV (no compatible), frente a sólo el 3% de tipo IV observado en la cepa SOL. De igual manera, se observó que no existía una correlación directa entre el grado de identidad de la secuencia Pr77Sh con la de Pr77 de L1Tc/NARTc y la estructura secundaria L1TcRz. Así, hubo secuencias que manteniendo un 74% de identidad con Pr77 de L1Tc/NARTc adquirieron una estructura ribozima tipo I y secuencias que con un 85,7% de identidad no adoptaron un plegamiento compatible con L1TcRz. Este hecho corroboró la importancia del emparejamiento de bases para adquirir una conformación ribozima tipo HDV adecuada, independientemente de la composición nucleotídica de las secuencias.

Basándonos en la compatibilidad del plegamiento adoptado por las huellas Pr77Sh con L1TcRz y su contexto genómico, se seleccionaron 10 Pr77Sh para estudiar su potencial actividad catalítica mediante ensayos de corte co-transcripcional y post-transcripcional *in vitro*. La transcripción *in vitro* de transcritos quiméricos formados por la secuencia del vector y la secuencia de cada huella, seguida de la cuantificación de los productos generados por el corte que debiera tener lugar en el nucleótido +1 de cada huella permitiría detectar, en caso de actividad ribozima HDV, la secuencia situada corriente arriba del corte y aquella localizada corriente abajo del punto de corte. Los resultados obtenidos revelaron actividad de corte co-transcripcional en todas las huellas seleccionadas y mostraron que IK-51 y BM-1 tenían una actividad de corte co-transcripcional superior, que en el caso de BM-1 era comparable a la detectada en L1TcRz en todas las condiciones ensayadas (25°C, 37°C y 42°C, así como a baja concentración de rNTP). A 37°C, IK-51 y BM-1 presentaron un porcentaje de corte del 4% y 64,5%, respectivamente, siendo el obtenido para L1TcRz del 70,6% y el del resto de huellas ensayadas entre 1,6 y 2,5%. Para determinar cómo afectaba la temperatura a la actividad óptima de las ribozimas, las reacciones también se llevaron a cabo a otras temperaturas. La actividad co-transcripcional disminuyó notablemente en aquellas ribozimas con mayor actividad cuando las reacciones se realizaron a 25°C respecto a 37°C (2,9% *versus* 4% para IK-51, 28,5% *versus* 64,5% para BM-1 y 39,6% *versus* 70,6% para L1TcRz). Por el contrario, a 42°C la actividad aumentó alrededor de un 8% en las ribozimas más activas BM-1 (69,6%) y L1TcRz (75,3%). También se observó un ligero incremento en CL-102 (3,7%). Este aumento de la actividad a 42°C y

disminución a 25°C es consistente con lo observado en ensayos previos en L1TcRz (Sánchez-Luque et al., 2011) y apoya la fuerte dependencia de la temperatura que se ha descrito para ribozimas tipo HDV, como las presentes en otros organismos como el anfibio *Branchiostoma floridae* y la lamprea marina *Petromyzon marinus* (Webb and Lupták, 2011). Asimismo, en otras ribozimas, como la de *Tetrahymena*, los autores también afirman que la temperatura puede alterar la tasa de plegado y elongación del ARN (Heilman-Miller and Woodson, 2003). Sin embargo, más que por la temperatura, la actividad ribozima se vio afectada por la concentración de rNTPs en las reacciones. Así, a 37°C y a una concentración 100 veces menor de rNTPs, se detectó la mayor actividad catalítica de hasta el 4,6% para IK-51 y del 75% para BM-1, siendo de un 81% para L1TcRz. Esto podría deberse al aumento de los intermediarios plegados metaestables producidos durante la transcripción al reducirse la tasa de transcripción de la T7 ARN polimerasa debido a la disminución de la concentración de NTP, como se describió para el ARN del viroide del tubérculo fusiforme de la patata (Repsilber et al., 1999).

Las reacciones de corte post-transcripcional de transcritos sin cortar, procedentes de los ensayos co-transcripcionales, realizadas a diferentes concentraciones de Mg<sup>2+</sup> indicaron que la actividad catalítica de las ribozimas ensayadas es dependiente de magnesio. Aunque la ribozima HDV puede funcionar en ausencia de iones metálicos divalentes, en condiciones fisiológicas de tampón la catálisis depende de la presencia de cationes metálicos divalentes como el Mg<sup>2+</sup> (Been, 2006) y presentan una actividad drásticamente menor en iones monovalentes (Rosenstein and Been, 1990; Wu et al., 1989). Estos cationes divalentes facilitan el plegado del ARN y parecen participar en la catálisis (Nakano et al., 2001, 2003). La ribozima HDV es la única de las pequeñas ribozimas catalíticas ensayadas que coordina un catión Mg<sup>2+</sup> en su centro catalítico (Chen et al., 2009, 2010). Así, la dependencia del Mg<sup>2+</sup> observada es consistente con la naturaleza de la ribozima. Las cinéticas de corte de las ribozimas IK-51 y BM-1 se ajustaban al modelo exponencial *two-phase decay*, característico de las ribozimas tipo HDV (Nakano et al., 2000). BM-1, tal y como se observó en los ensayos co-transcripcionales, fue la ribozima en la que se detectó la mayor actividad catalítica (52,7%), seguida de IK-51 (6,5%). Interessantemente, la mayor eficiencia de corte se alcanzó a 10 mM de Mg<sup>2+</sup> para todas las huellas ensayadas incluido L1TcRz, excepto



para BM-1 que fue a 1 mM, lo que indicaría que para BM-1 esta concentración de iones magnesio podría ser la óptima para la estabilidad del plegamiento. En un estudio anterior, también se observó mayor actividad catalítica a menor concentración de MgCl<sub>2</sub> en la ribozima presente en el elemento SIDER2A de *Leishmania major*, con mayor actividad a 0,1 mM que a 1 o 10 mM de MgCl<sub>2</sub> (Sánchez-Luque et al., 2014). Cabe destacar que tanto BM-1 como L1TcRz fueron más activas en condiciones co-transcripcionales que post-transcripcionales, lo cual es consistente con estudios anteriores que soportan que las ribozimas de *Trypanosoma spp* son más activas a nivel co-transcripcional (Sánchez-Luque et al., 2011, 2014) que las ribozimas de *Leishmania spp* que tienen mayor actividad post-transcripcional que co-transcripcional.

La existencia de actividad HDV de las huellas Pr77Sh no asociadas a retroelementos y que se encuentran insertadas en diferentes localizaciones genómicas, muchas de ellas en genes implicados en procesos de virulencia del parásito, sugiere un importante papel funcional de estas secuencias en la regulación de la expresión génica que tiene lugar en el parásito. Este hecho justificaría que las huellas Pr77, tanto las que forman parte de los retroelementos como las no asociadas a estos, no constituyen elementos fósiles, ni son basura genómica ni elementos egoístas si no que son herramientas activas que conllevan a que los genomas que las portan las hayan mantenido a lo largo de la evolución.

### **3. ESTUDIO DEL CONTENIDO GÉNICO Y EN RETROELEMENTOS DE LA CEPA SOL DE *T. cruzi* SECUENCIADA POR TECNOLOGÍA PACBIO. ANÁLISIS DE EXPRESIÓN GÉNICA DIFERENCIAL ENTRE LAS 3 FORMAS DEL PARÁSITO**

La cepa SOL de *T. cruzi* se aisló de un bebé infectado verticalmente y nacido en España, región no endémica de la enfermedad de Chagas (Murcia et al., 2013). El parásito se aisló de la sangre del bebé al nacer, mediante hemocultivo, en el laboratorio y se secuenció su genoma, además de por la tecnología Ion Torrent, por secuenciación PacBio. Mediante la tecnología PacBio, se generaron lecturas de mayor tamaño que facilitaron el ensamblaje del genoma, obteniéndose *contigs* más grandes que los obtenidos empleando la tecnología Ion Torrent, con un tamaño promedio de 37.646 pb. Esto se tradujo en un genoma menos fragmentado,

ensamblado en menos *contigs* (1.783), lo que permitió un análisis más preciso del contenido génico y de retroelementos los cuales tienen un tamaño de hasta cerca de 5 kb.

Como era de esperar para este kinetoplástido, el contenido de G+C rondó el 50% (51,8%) y la calidad del ensamblaje fue corroborado por análisis BUSCO. Tras la anotación del genoma, se identificaron 31.570 genes, con un tamaño promedio de 1.227 pb, de los cuales, un 93% fueron anotados. De estos, el 68,5% presentaba identidad con proteínas homólogas de función conocida, mientras que el 31,5% correspondían a genes en los que no fue posible determinar su función teórica (codificantes de proteínas hipotéticas). Interesantemente, el porcentaje de estos genes codificantes de proteínas hipotéticas detectado en SOL, mediante la metodología descrita en esta Tesis, fue inferior al detectado en otras cepas de *T. cruzi*, que poseen valores en torno al 50%, incluso respecto a la cepa Bug2148, también de DTU V (53,3%) (Callejas-Hernández, Gironès, et al., 2018; Callejas-Hernández, Rastrojo, et al., 2018).

Cuando se estimó el contenido génico de diferentes genes y familias génicas, se observó que el mayor número de copias identificadas correspondía a genes codificantes de TS y RHS, representado un 5,5% y un 4,1%, respectivamente, del total de genes identificados en este genoma. Del mismo modo, se detectaron numerosas copias de genes codificantes para mucinas, MASP, DGF-1 y GP63. La metaloproteasa GP63 está implicada en el proceso de infección de las células del hospedador (Kulkarni et al., 2009) y DGF-1 se ha asociado con la capacidad de *T. cruzi* para unirse a proteínas de la matriz extracelular (Kawashita et al., 2009). La fiabilidad de esta estimación se corroboró al detectar también un alto número de copias para genes de expresión constitutiva como quinasas, cisteína peptidasa o glicosiltransferasas, entre otras. De igual manera, se identificaron 4 copias del gen codificante para KMP11, consistente con el número de copias de este gen identificadas experimentalmente en tripanosomas, lo cual soporta la buena calidad del genoma ensamblado (Bridge et al., 1998; Thomas et al., 2000). Aunque en menor medida, también se detectaron copias de genes codificantes para EF-1 $\gamma$ , TASV, HSP70 y PFR.

En este genoma también se identificaron las huellas Pr77, así como los retrotransposones L1Tc, NARTc y DIRE que las portan. Se identificaron 282 secuencias Pr77, con una identidad de entre un 89% y 100% con la secuencia Pr77 de L1Tc. De igual manera, se identificaron 78 secuencias del retrotransposón L1Tc (identidad > 92%) y 93 de NARTc (identidad > 80%). La mayoría de estas secuencias mostraron un tamaño de, al menos, el 95% del tamaño de la secuencia L1Tc o NARTc. También, se analizó en este genoma la presencia de elementos DIREs, hallándose 107 copias de estos elementos, 88 de ellos de tamaño completo, con identidad con las secuencias DIREs empleadas de entre 80,4% a 99,7%.

El análisis de la expresión génica diferencial en los 3 estadios de *T. cruzi* mediante ARN-seq, reveló que existía un claro patrón de expresión diferencial entre las formas epimastigote respecto a tripomastigote y amastigote, presentando estas dos últimas mayor similitud entre ellas, aunque con diferencias de expresión importantes en ciertos genes. Así, 18.224 genes se expresaban diferencialmente entre tripomastigote y epimastigote, 17.981 genes entre amastigote y epimastigote y 2.485 genes entre tripomastigote y amastigote. Todas estas diferencias fueron estadísticamente significativas y suponían, al menos, el doble o la mitad de expresión en una forma del parásito respecto a otra. Interesantemente, cuando se comparó el nivel de expresión de la forma tripomastigote *versus* epimastigote y amastigote *versus* epimastigote, alrededor del 50% de los genes diferencialmente expresados en cada caso, mostraban diferencias de expresión del orden de, al menos, 32 veces más o menos expresión. Del mismo modo, aproximadamente 1.000 genes en cada caso, vio modificado su nivel de expresión hasta 1.024 veces, la mayoría infraexpresados en tripomastigote/amastigote *versus* epimastigote. Cuando comparamos las formas tripomastigote *versus* amastigote, además de la disminución en el número de genes diferencialmente expresados, las diferencias de expresión no alcanzaron valores tan dispares, mostrando la mayoría de los genes el doble o la mitad de expresión entre estas dos formas del parásito.

Al analizar la expresión diferencial de genes concretos, se observó un mayor nivel de expresión en el estadio tripomastigote de importantes genes para el parásito asociados a procesos de virulencia como TS, MASP, TASV, mucinas, GP63 y DGF-1. Las diferencias de expresión fueron estadísticamente significativas respecto

a la expresión detectada en epimastigote para todas ellas y también respecto a la expresión en amastigote para TS, MASP, TASV y mucinas.

El empleo de los datos de ARN-seq para determinar la existencia de expresión diferencial de los retroelementos identificados en el genoma, asociada a la forma del parásito, reveló que muchos de los retroelementos identificados mostraban diferencias de expresión entre estadios. La expresión diferencial se detectó entre los estadios tripomastigote *versus* epimastigote o amastigote *versus* epimastigote, con excepción de 7 elementos DIRE diferencialmente expresados entre las formas tripomastigote y amastigote. La ausencia de diferencias de expresión entre estos dos últimos estadios, unido al hecho de que la mayoría de los retroelementos diferencialmente expresados en tripomastigote *versus* epimastigote y amastigote *versus* epimastigote se trataban de los mismos retroelementos, pone de manifiesto la similitud entre los estadios tripomastigote y amastigote en cuanto a la expresión de retroelementos y la diferencia de ambas formas del parásito respecto a epimastigote. Interesantemente, tanto para L1Tc como NARTc diferencialmente expresados, el número de ellos infraexpresadas o sobreexpresadas fue similar, mientras que los DIREs se encontraban en su mayoría infraexpresados en tripomastigote/amastigote *versus* epimastigote.

Cuando se compararon los niveles de expresión de cada retroelemento en cada estadio, se observó que existía una mayor expresión de la secuencia Pr77 y los DIRE\_322.1 y DIRE\_664.1 en el estadio epimastigote en comparación con los estadios tripomastigote y amastigote, siendo las diferencias estadísticamente significativas.

#### **4. ANÁLISIS DE LA SENSIBILIDAD Y ESPECIFICIDAD DE KMP11, HSP70, PFR2 Y 3973<sup>d</sup> DE *T. cruzi* EN PACIENTES CON ENFERMEDAD DE CHAGAS**

Uno de los grandes retos para el control de la enfermedad de Chagas es la disponibilidad de biomarcadores para la evaluación temprana de la respuesta terapéutica (Pinazo et al., 2014). En este sentido, como parte de un ensayo ciego multicéntrico en el que se evaluó la reactividad serológica a lo largo del tiempo de una serie de antígenos del parásito, de diferente procedencia, en una colección retrospectiva de muestras de pacientes con enfermedad de Chagas de diferente

origen, evaluamos la sensibilidad y especificidad de los antígenos KMP11, HSP70, PFR2 y 3973<sup>d</sup> de *T. cruzi*.

Para determinar si un biomarcador podría estar asociado a la infección, se ha determinado en un criterio TPP (*Target Product Profile*) reciente, que un biomarcador aceptable para una prueba de evaluación temprana de eficacia del tratamiento en la enfermedad de Chagas debe poseer una sensibilidad  $\geq 60\%$  y especificidad  $\geq 90\%$  (Alonso-Padilla et al., 2020). Los resultados del presente estudio indican que KMP11, 3973<sup>d</sup> y PFR2 obedecerían a esos criterios establecidos. Además, de todos los antígenos evaluados, KMP11 y 3973<sup>d</sup> fueron los dos antígenos con mejor rendimiento para identificar tanto a los sujetos negativos como a los infectados por *T. cruzi* lo que indica un importante valor diagnóstico de estas proteínas. Así, su sensibilidad fue del 91% y el 84%, y su especificidad del 98% y el 93%, respectivamente, teniendo un AUC ROC de 0,97 y 0,96 (Alonso-Padilla et al., 2021). Estos valores de especificidad, junto a los de las moléculas HSP70 (83%) y 3973<sup>d</sup> (93%) diferían considerablemente de los obtenidos para otros antígenos incluidos en el estudio como F29 e INFynity (27% y 74%, respectivamente) (Alonso-Padilla et al., 2021). Los resultados obtenidos sugieren el potencial de estos 4 antígenos como potenciales biomarcadores para el seguimiento de los sujetos tratados con enfermedad de Chagas.

##### **5. KMP11, HSP70, PFR2 Y 3973<sup>d</sup> DE *T. cruzi* COMO BIOMARCADORES DE EFICACIA TERAPÉUTICA EN PACIENTES CON ENFERMEDAD DE CHAGAS**

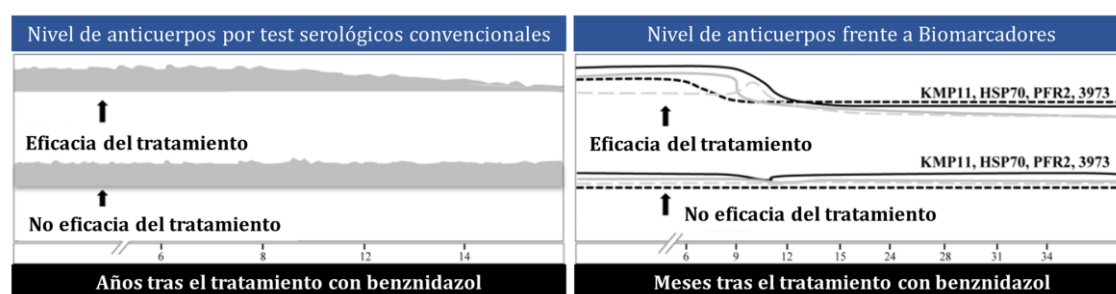
La conversión a serología negativa que se asocia a la cura parasitológica (Cançado, 1999), puede tardar décadas en ocurrir después del tratamiento (Fabbro et al., 2007; Viotti et al., 2006). Así, una seroconversión clínicamente relevante podría medirse como un descenso continuo y sustancial de la reactividad del suero del paciente frente a antígenos específicos de *T. cruzi*, pudiendo ser una herramienta útil para evaluar el efecto del tratamiento. Según el TPP, un biomarcador aceptable de respuesta terapéutica permite detectar cambios en la respuesta observada (descenso en la reactividad) durante un periodo corto de tiempo (de 3 a 24 meses) después del tratamiento (Pinazo et al., 2014). En este sentido, en estudios anteriores del laboratorio, se han descrito una serie de antígenos de *T. cruzi*, incluidos KMP11,

PFR2, HSP70 y 3973<sup>d</sup> como marcadores de seguimiento terapéutico. Estos antígenos fueron reconocidos por los sueros de los pacientes con enfermedad de Chagas en cualquier fase clínica, observándose un descenso significativo de los títulos de anticuerpos frente a ellos poco después del tratamiento con benznidazol (de 6 a 9 meses) (Fernández-Villegas et al., 2011; Thomas et al., 2012).

La disminución de los títulos serológicos de los pacientes frente a antígenos específicos del parásito se asocia con una reducción sustancial de la carga parasitaria y con una mejora del estado clínico de los pacientes tratados (Fragata-Filho et al., 2016; Viotti et al., 2011). En estas circunstancias, el objetivo de este estudio fue establecer el valor predictivo de la eficacia terapéutica de las moléculas KMP11, PFR2, HSP70 y 3973<sup>d</sup> como set de BMK. Así, en un ensayo a ciegas evaluamos y comparamos la reactividad de los sueros de pacientes IND contra estos 4 BMK antes y a los 24 y 48 meses después del tratamiento con benznidazol.

La reactividad frente cada uno de los antígenos analizados disminuyó continuamente a lo largo del tiempo después del tratamiento en muchos pacientes IND con significancia estadística. Sin embargo, no se observaron cambios en la reactividad frente a *STcA*. El porcentaje de pacientes en los que la reactividad frente a los antígenos (analizados individualmente) disminuyó fue similar para cada uno de los antígenos y osciló entre el 86% (PFR2 y HSP70) y el 89% (KMP11) a T24 (24 meses post-tratamiento) y entre el 91% (PFR2 y HSP70) y el 97% (KMP11 y 3973<sup>d</sup>) a T48 (48 meses post-tratamiento). Cuando se analizaron los 4 BMKs como set, se detectó una disminución de la reactividad frente a los 4 BMK en el 45,5% de los pacientes a T9 (9 meses post-tratamiento), y en el 68,2% y 81,3% de los pacientes a T24 y T48, respectivamente. Estos porcentajes fueron ligeramente mayores cuando se permitió un ligero aumento de la reactividad del suero a 9 meses (inferior al 20%). La importancia de la contribución de cada uno de los 4 BMK para evaluar el impacto del tratamiento se demostró por PCA, reforzando la idea del uso de los 4 como set. De esta forma, los resultados mostraron que, tras un corto periodo después del tratamiento, el análisis de la reactividad frente a estos 4 BMKs podría constituir un sistema útil de marcadores tempranos para evaluar la influencia del tratamiento.

Dos de los criterios de eficacia terapéutica se establecieron en base a una disminución continua de la reactividad para los 4 BMK a los 24/48 meses del tratamiento y una disminución sustancial de la reactividad para, al menos, dos (STEC, *Standard Therapeutic Efficacy Criteria*) o tres (RTEC, *Restrictive Therapeutic Efficacy Criteria*) de los 4 BMK. El descenso debía ser de, al menos, un 40% para KMP11, PFR2 y 3973<sup>d</sup> y de un 30% para HSP70 en comparación con la reactividad antes del tratamiento (T0). Ambos criterios permitían un ligero aumento de la reactividad poco después del tratamiento (T9), inferior al 20% del valor a T0. El 42,4% (28/66) de los pacientes IND cumplían el criterio STEC a T24. Además, el 68,75% de los pacientes mostraron evidencia de eficacia terapéutica a T48, siendo esta la tasa global de respuesta al benznidazol que se estimó en este estudio. Interesantemente, dado que 12 de los 14 pacientes que cumplían los criterios de eficacia terapéutica a los 24 meses después del tratamiento también los cumplieron a 48 meses, pudimos afirmar que el STEC tenía un valor predictivo del 85,7%. Cuando se consideraron los criterios de eficacia terapéutica más estrictos (RTEC), el valor de eficacia terapéutica obtenido a T48 fue similar al detectado a T24 en las condiciones del STEC (40,6% *versus* 42,4%, respectivamente), indicando que las condiciones más restrictivas (RTEC) tienen el mismo valor predictivo de eficacia terapéutica que la STEC, aunque el uso de RTEC retrasa la detección de la tasa de respuesta al tratamiento. El valor predictivo observado para el set de BMK en un corto periodo de tiempo es crucial ya que otros estudios han observado seroconversión completa o parcial tras periodos 5 veces más largos que los aquí estudiados (**Figura D1**) (Fabbro et al., 2011; Pinazo et al., 2014, 2015).



**Figura D1. Nivel de anticuerpos por test serológicos para la evaluación de la eficacia del tratamiento en la enfermedad de Chagas.** La seroconversión negativa mediante las pruebas serológicas convencionales puede tardar décadas en ser concluyente. Sin embargo, el uso de los antígenos KMP11, HSP70, PFR2 y 3973<sup>d</sup> de *T. cruzi* permite la detección de la disminución de la reactividad después del tratamiento a los 6 meses para KMP11 y a los 9 meses para PFR2, HSP70 y 3973. *Adaptada de Egui et al., 2018.*

El fracaso terapéutico se define como la persistencia del parásito después del tratamiento, detectado por métodos como la PCR (Viotti et al., 2014). En consonancia con la fiabilidad del set de 4 BMKs para detectar el fracaso terapéutico, ninguno de los pacientes que cumplían el STEC/RTEC tuvo un resultado positivo de PCR tras el tratamiento. Tampoco disminuyó en estos pacientes el título de anticuerpos frente a antígenos totales del parásito (kit comercial Biokit) lo que está en concordancia a lo previamente reportado (Fernández-Villegas et al., 2011, 2016).

Dada la necesidad de marcadores tempranos que permitan evaluar la influencia del tratamiento en la enfermedad de Chagas, los resultados aquí mostrados suponen un relevante hallazgo al indicar que el set de biomarcadores KMP11, HSP70, PFR2 y 3973<sup>d</sup> y la aplicación de los criterios establecidos en este estudio podrían ser herramientas útiles para monitorizar la eficacia del tratamiento con benznidazol en pacientes con enfermedad de Chagas crónica, y lo que es más importante, a corto plazo, pudiendo considerarse biomarcadores tempranos de detección de eficacia o, en su caso, fallo terapéutico. Además, resultados de otros miembros del laboratorio mostraron que los pacientes que cumplen los criterios de eficacia terapéutica evaluados por este set de biomarcadores serológicos muestran un aumento notable de la respuesta multifuncional de sus células T CD8<sup>+</sup> específicas frente a antígenos de *T. cruzi* (Egui et al., 2019).

## **6. DETERMINACIÓN DE LA EXPRESIÓN DIFERENCIAL DE 106 GENES IMPLICADOS EN LA RESPUESTA INMUNE EN PACIENTES CRÓNICOS CON ENFERMEDAD DE CHAGAS**

### **6.1. Expresión diferencial de genes implicados en la respuesta inmune en pacientes en fase indeterminada *versus* sujetos sanos**

La infección por *T. cruzi* desencadena en el hospedador múltiples mecanismos inmunitarios para combatir el patógeno, que pueden mantenerse durante décadas haciendo que el sujeto permanezca muchos años, incluso décadas, en la fase indeterminada de la enfermedad. A lo largo de este tiempo, existe un frágil equilibrio entre la replicación del parásito y la respuesta inmunitaria del huésped (dos Santos Virgilio et al., 2014) que, cuando se rompe, conduce a la progresión de



la enfermedad. Es por ello que la comprensión de los mecanismos moleculares de la infección resulta esencial para el control de la enfermedad.

Parte de los resultados incluidos en esta Tesis corresponden a un extenso análisis por RT-qPCR empleando OpenArrays® para identificar los cambios globales en los perfiles de expresión génica de 106 genes relacionados con el sistema inmunitario en pacientes IND en respuesta a las proteínas específicas del parásito (*STcA*) en comparación con sujetos sanos (DS). Los genes se seleccionaron en base a su relevancia como parte de los procesos inmunológicos que se asocian al control de la infección causada por patógenos intracelulares.

En una primera aproximación, se analizó el perfil de expresión de los genes a estudio entre los donantes sanos en respuesta a la estimulación con *STcA*, teniendo en cuenta la influencia del origen del donante, ya fuese de zonas endémicas o no endémicas de la enfermedad de Chagas. Los individuos no se agruparon según su origen en el mapa de calor (**Figura R22**) o análisis de componentes principales (**Figura R23A**) construidos a partir de sus valores de expresión. Además, los análisis estadísticos confirmaron que no existían diferencias estadísticamente significativas entre las puntuaciones de las componentes principales representadas entre ambos grupos de sujetos (PC1  $p = 0,65$ , PC2  $p = 0,23$ ). Por otra parte, el análisis de la expresión génica diferencial reveló que sólo el 4,7% de los genes analizados (5 de 106) se expresaban diferencialmente entre ambos grupos de sujetos sanos con significancia estadística. Estos genes fueron las interleuquinas *IL6*, *IL12B* e *IL17A* y los marcadores de fenotipo *FCER1A* y *CD27*. Así, los resultados indicaron la ausencia de expresión diferencial de la mayoría de los genes a estudio entre los donantes sanos de distinto origen geográfico, motivo por el que se consideraron un único grupo en los análisis comparativos posteriores.

Cuando se comparó el nivel de expresión génica de los 106 genes a estudio de los pacientes IND con el de los sujetos sanos, representando sus valores NRQ en un mapa de calor (**Figura R24**), se detectaron diferencias notables en un gran grupo de genes que suponían más de la mitad de los genes analizados, estando la mayoría sobreexpresados en IND frente a DS. La estructura del conjunto de datos se determinó mediante análisis de componentes principales (PCA), que puede interpretarse como una medida de la expresión génica diferencial entre los sujetos

IND y DS. El análisis PCA mostró que existían diferencias de expresión génica entre los sujetos IND y DS, en los genes correlacionados con PC1. Los 23 genes correlacionados positivamente con PC1 correspondían a genes codificantes para citoquinas/interleuquinas y receptores (*CSF1*, *IFNG*, *IL12A*, *IL12RB1*, *IL12RB2*, *IL18R1*, *IL2RG*, *IL7* y *TNF*), moléculas coestimuladoras (*CD2*, *CD40*, *CD40LG*, *CD69*, *ICOS*), factores de transcripción (*GATA3*, *STAT1*, *TBX21*), moléculas implicadas en la apoptosis (*BCL2*, *CASP3*, *FAS*), marcador de fenotipo (*CD28*), receptor inhibitorio (*BTLA*) y marcador de células dendríticas (*CD83*). Por otro lado, 9 genes mostraron una correlación negativa con PC1 siendo estos, interleuquinas y receptores (*IFNGR1*, *IFNGR2*, *IL17RA*, *IL18*), moléculas de adhesión (*ITGAX*, *ITGB2*), receptor inhibitorio (*HAVCR2*), marcador de células dendríticas (*CCR1*) y molécula coestimuladora (*CD86*).

El análisis de la expresión génica diferencial reveló que el nivel de expresión de un gran número de genes variaba entre los pacientes IND y DS, expresándose el 32% de los genes más del doble o menos de la mitad con significancia estadística ( $p < 0,05$ ) en IND *versus* DS. Además, para el 91% de los genes las diferencias detectadas fueron muy significativas ( $p < 0,001$ ). La mayoría de los genes expresados diferencialmente estaban sobreexpresados en IND (67,6%), siendo estos, citocinas/interleuquinas y receptores (*CSF1*, *CSF2*, *IFNG*, *IL12A*, *IL12B*, *IL12RB2*, *IL2*, *IL27*, *IL5*, *IL5RA*, *IL6*, *IL7* y *TNF*), moléculas coestimuladoras (*CD40*, *CD40LG*), factores de transcripción (*STAT1*, *TBX21*), moléculas implicadas en apoptosis (*BCL2*, *FAS*), receptores inhibitorios (*BTLA*, *CD274*), marcador de fenotipo (*CD27*) y enzimas (*IDO1*). Los genes infraexpresados en IND fueron la interleuquina *IL18* y los receptores *IFNGR1* e *IFNGR2*, moléculas de adhesión (*ITGAX*, *ITGB2*), receptor inhibitorio *HAVCR2*, marcadores de células dendríticas (*CCR1*, *CLEC9A*, *XCR1*), marcador de fenotipo *FCER1A* y molécula coestimuladora *CD86*. Las mayores diferencias de expresión se identificaron para 9 genes. Cuatro de ellos infraexpresados en los pacientes IND (*CD86*, *CLEC9A*, *FCER1A* e *IL18*). La expresión del gen *CLEC9A* no se detectó en ningún paciente IND, mientras que el nivel de ARNm de *CD86*, *FCER1A* e *IL18* fue 5,5, 22 y 9 veces menor en IND que en DS, respectivamente. Los 5 genes restantes (*CD27*, *CSF2*, *IFNG*, *IL5* e *IL6*) estaban sobreexpresados en IND, con diferencias de entre 6,5 a 32 veces más expresión). Cabe destacar que la expresión del gen *IFNG* (*IFN- $\gamma$* ) sufrió la mayor variación entre

IND y DS mostrando un valor de FC > 5, lo que corresponde a 32 veces más expresión en los pacientes IND. El IFN- $\gamma$  es importante para orquestar el desarrollo de la inmunidad adaptativa, contribuyendo a la diferenciación de las células T CD4<sup>+</sup> Th1 y CD8<sup>+</sup> necesarias para controlar la proliferación del parásito que se produce durante la infección aguda (Cerbán et al., 2020). Así, el control de la infección por *T. cruzi* está relacionado con la activación del IFN- $\gamma$  que conduce a la eliminación intracelular de los parásitos (Kulkarni et al., 2015). El gen *TNF* también se encontró sobreexpresado en los pacientes IND frente a DS, lo que soporta estudios previos que describieron la secreción de altos niveles de IFN- $\gamma$  y TNF en pacientes IND (Ferreira et al., 2003; Requena-Méndez et al., 2013), lo que sugiere el relevante papel de estas moléculas en el control de la infección por *T. cruzi*. Además del IFN- $\gamma$  y el *TNF*, el control de la infección por *T. cruzi* se ha asociado al perfil de citoquinas producido por las células Th1 (Hoft and Eickhoff, 2005; Kumar and Tarleton, 2001; Petray et al., 1993; Rodrigues et al., 1999; Tarleton et al., 2000).

Con objeto de conocer los procesos inmunológicos a los que se asocian los perfiles de expresión génica diferenciales observados entre IND y DS, se construyó una red de interacción proteína-proteína (IPP) en STRING a partir de los 34 genes expresados diferencialmente. El análisis IPP mostró que estos genes expresados diferencialmente entre los sujetos IND y DS codifican proteínas que poseen un alto grado de interacción entre sí (valor *p* de enriquecimiento IPP < 1e-16). Los procesos y vías biológicas más significativos en los que participan dichos genes estaban asociados a la respuesta inmunitaria como la ruta de señalización de la familia IL-2, diferenciación de las células Th1 y Th2, señalización de la familia IL-12, vía de señalización JAK-STAT, así como la diferenciación de las células Th17 y la vía de citotoxicidad mediada por células NK. Según el análisis STRING, 20 de los 34 genes expresados diferencialmente (59%) estaban implicados en una o más de estas seis rutas mencionadas, con valores de FDR muy bajos (entre 4,77e-19 y 2,2e-08), lo que indica la fiabilidad de las predicciones.

STAT1 juega un papel central en 5 de estas rutas altamente enriquecidas. La sobreexpresión de *STAT1* observado en los pacientes IND soporta estudios anteriores que describieron la activación de la vía de señalización de STAT1 en las células del huésped tras la infección por *T. cruzi*, lo que conduce a una expresión de

*STAT1* significativamente elevada (De Avalos et al., 2002). Se ha descrito, también, que *STAT1* juega un papel importante en la primera línea de defensa contra la invasión por tripomastigotes de *T. cruzi* (Stahl et al., 2014). Además, *STAT1* se considera un mediador clave en la señalización intracelular del *IFN-γ*, conduciendo la eliminación de esta proteína a la susceptibilidad a varios microbios intracelulares (Kulkarni et al., 2015). De hecho, el efecto protector del *IFNγ*, tanto en la entrada de los tripomastigotes a las células del huésped como frente a la multiplicación intracelular de los amastigotes, se basa en la activación de *STAT1* por fosforilación de la tirosina (Stahl et al., 2014). Estos resultados son consistentes con la sobreexpresión, tanto de *STAT1* como de *IFNG*, observada en la presente Tesis en pacientes IND, lo que sugiere el relevante papel de estas moléculas para controlar la infección por *T. cruzi* en esta fase. Además, se ha detectado la sobreexpresión de los genes *IL12A*, *IL12B* e *IL12RB2* en los pacientes IND, los cuales participan en las rutas “señalización de la familia IL12”, “diferenciación de las células Th1 y Th2” y “señalización JAK-STAT” según STRING. La interleuquina 12 actúa sobre los linfocitos T activados, impulsando su diferenciación hacia la subclase Th1. Esta citoquina se caracteriza por ser un potente inductor de la producción de *IFN-γ* por parte de las células NK y de diferentes subpoblaciones de células T (Gately et al., 1994), lo que concuerda con la expresión diferencial de *IFNG* que detectamos en los pacientes IND. Además, los anticuerpos anti-IL12 aumentan la susceptibilidad a la infección, destacando su importante papel en el control de la parasitemia (Aliberti et al., 1996). Los resultados obtenidos también sugieren la posible activación de la vía de señalización de la familia IL2, detectándose una sobreexpresión de esta interleuquina en los pacientes IND. Se ha descrito que la coproducción antígeno-específica de *IFN-γ*, IL2 y *TNFα* por parte de las células T CD8<sup>+</sup> se encuentra en mayor proporción en pacientes asintomáticos disminuyendo su proporción con la progresión de la severidad de la enfermedad de Chagas cardiaca (Lasso et al., 2015; Mateus et al., 2015).

Ciertos genes expresados diferencialmente en IND participaban en la diferenciación de las células Th17 según STRING. Las células Th17 corresponden a una subpoblación de células T CD4<sup>+</sup> que desempeñan un papel central en la patogénesis de múltiples enfermedades autoinmunes, así como en la defensa contra algunas bacterias y hongos extracelulares (Ishigame et al., 2009; Lin et al., 2009;

Milner et al., 2010; Zielinski et al., 2012). Sin embargo, se ha cuestionado su papel en las infecciones intracelulares (Cai et al., 2016). En contraste con este paradigma, se ha descrito el papel protector de las células Th17 en el control de la parasitemia y la supervivencia de ratones infectados por *T. cruzi* (Cai et al., 2016; Miyazaki, et al., 2010). Además, el perfil Th17 se ha considerado un factor protector para prevenir el daño miocárdico en la enfermedad de Chagas humana (Magalhães et al., 2013; Sousa et al., 2017). Los resultados obtenidos apoyan este hallazgo y sugieren que esta vía puede activarse para controlar la parasitemia y prevenir la progresión de la enfermedad en pacientes IND.

El análisis GSEA reveló que varios conjuntos de genes estaban enriquecidos en los sujetos IND, asociado a la sobreexpresión observada en una serie de genes. De estos, 17 genes se correlacionaban con PC1 (componente principal que separaba sujetos IND de sanos en el PCA) o bien mostraron una expresión génica diferencial estadísticamente significativa en el gráfico de volcán ( $FC > 1$  y  $p < 0,05$ ), cumpliendo ambos criterios un 52,9% de ellos. En cuanto a las rutas enriquecidas, la activación en estos pacientes de las células B dependientes de antígenos fue consecuencia de la sobreexpresión de los genes *CD28*, *CD40*, *CD40LG*, *FAS* e *IL2*. La interacción de CD40 con CD40L y la de CD28 con CD80 proporcionan señales coestimuladoras positivas que estimulan la activación, proliferación y diferenciación de las células B hacia células de memoria (BioCarta - [http://www.gsea-msigdb.org/gsea/msigdb/cards/BIOCARTA\\_ASBCCELL\\_PATHWAY](http://www.gsea-msigdb.org/gsea/msigdb/cards/BIOCARTA_ASBCCELL_PATHWAY)). La inflamación como respuesta protectora a la infección también se observa con el enriquecimiento en la vía "Citoquinas y respuesta inflamatoria" como resultado de la sobreexpresión de los genes *IL2*, *TNF*, *CSF1*, *CSF2*, *INFG*, *IL5*, *IL6* e *IL7*.

El hallazgo de que la ruta "IL12 dependiente de NO2 en las células NK" está enriquecida en los pacientes IND, los cuales permanecen asintomáticos y, por tanto, controlan la progresión de la enfermedad, sugiere que la activación de esta vía puede ser esencial para combatir el parásito. Un estudio anterior describió que la resistencia frente a *T. cruzi* se basa en la liberación de IL12 por parte de los macrófagos infectados, que induce la producción de IFN- $\gamma$  por las células T y NK (Cerbán et al., 2020). En los macrófagos, el IFN- $\gamma$  actúa activando las enzimas óxido nítrico sintasa inducible (iNOS) y NADPH oxidasa para la producción de óxido

nítrico (NO), especies reactivas de oxígeno (ROS) e intermediarios reactivos de nitrógeno (RNI) como el peroxinitrito (ONOO-), los cuales son críticos para la actividad tripanocida (Gazzinelli et al., 1992; Guiñazú et al., 2007; Vespa et al., 1994).

El análisis GSEA también reveló un enriquecimiento en los pacientes IND del conjunto de genes "Inducción de la regulación de las HSP por el estrés" como consecuencia de la sobreexpresión de los genes *FAS*, *BCL2*, *CASP3* y *TNF* incluidos en esta ruta. La activación de estos genes, los 4 implicados en procesos de apoptosis, se esperaba en pacientes de Chagas, dada la infección intracelular con *T. cruzi*. Sin embargo, la activación de esta ruta sugiere, además, que la expresión de estos genes podría estar activando las proteínas de choque térmico. Las proteínas de choque térmico, y en particular la Hsp27, ejercen un fuerte efecto protector sobre las células, principalmente, debido a su función en la regulación de la apoptosis (Garrido et al., 2003; Wang et al., 2014). Además, la Hsp27 ha demostrado tener la capacidad de disminuir los niveles de ROS, permitiendo que las células aumenten su resistencia al estrés oxidativo (Arrigo et al., 2005; Garrido et al., 1997; Rogalla et al., 1999) por lo que esta vía también podría estar actuando como mecanismo de control de la presencia de ROS en las células de estos pacientes. Este hecho parece ser esencial ya que se ha demostrado que cuando estas especies citotóxicas se producen en exceso o durante períodos de tiempo prolongados o cuando hay una respuesta antioxidante inadecuada, pueden acumularse y contribuir a la patogénesis de la enfermedad de Chagas (Zacks et al., 2005).

Los resultados mostrados en esta Tesis indican que la infección por *T. cruzi* induce cambios en el perfil de expresión de distintos genes que parecen estar implicados en vías inmunológicas relevantes. La expresión de estos genes codificantes de proteínas, podría constituir biomarcadores útiles de la forma indeterminada de la enfermedad de Chagas, así como podrían actuar como nuevas dianas terapéuticas en los huéspedes, útiles para prevenir la progresión a la fase crónica sintomática.

## **6.2. Expresión diferencial de genes implicados en la respuesta inmune en pacientes de Chagas en fase cardiaca *versus* fase indeterminada, previamente y tras el tratamiento con benznidazol**

Cuando los análisis comparativos de expresión génica se llevaron a cabo entre pacientes crónicos en fase indeterminada y pacientes con sintomatología cardiaca, se pudo observar que entre ellos también existían patrones de expresión diferenciales en determinados genes implicados en la respuesta inmune celular antígeno-específica, tanto por análisis PCA como en la representación por gráfico de volcán, siendo estas diferencias estadísticamente significativas.

Del mismo modo, los resultados mostrados en esta Tesis también revelan la existencia de patrones de expresión diferenciales en genes implicados en la respuesta inmune celular antígeno-específica tras el tratamiento con benznidazol, tanto en pacientes asintomáticos como en aquellos con sintomatología cardíaca. El análisis de estos genes diferencialmente expresados entre estos grupos de pacientes se abordará próximamente, ya que estos podrían constituir potenciales biomarcadores de progresión de la enfermedad o posibles dianas terapéuticas, así como útiles biomarcadores de eficacia del tratamiento.

Los resultados indican que la innovadora estrategia empleada en los estudios de expresión incluidos en esta Tesis Doctoral podría aplicarse para futuros análisis de expresión génica de más genes implicados en las vías identificadas, así como para la identificación de nuevas rutas. Todo ello permitirá, sin duda, dilucidar la respuesta inmune producida y las vías inmunológicas activadas en los pacientes asintomáticos y sintomáticos de la enfermedad de Chagas.





## **CONCLUSIONES**

---



### ***Biomarcadores genéticos***

1. Los genomas de las cepas de *T. cruzi* estudiadas, B. M. López (DTU Ia), Dm28 (DTU Id), Y (DTU II), Ikiakarora (DTU III), SOL (DTU V) y CL Brener (DTU VI) contienen diferente número de los retroelementos L1Tc y NARTc portadores de la huella Pr77 existiendo, en todas ellas, mayor número de elementos L1Tc que de NARTc.
2. En las mencionadas 6 cepas de *T. cruzi* existen diferencias en el número de copias de genes de función relevante y/o implicados en procesos de virulencia del parásito (TS, MASP, TASV, RHS, etc.) manteniéndose conservado el número de copias de genes constitutivos.
3. En el genoma de las 6 cepas de *T. cruzi* a estudio se han identificado huellas Pr77 no asociadas a los retroelementos L1Tc/NARTc, denominadas *single Pr77 hallmarks* (Pr77Sh), con diferente abundancia, identidad de secuencia y sitio de inserción en el genoma.
4. Los retrotransposones L1Tc y NARTc portadores de Pr77 y, especialmente, las huellas Pr77Sh se asocian frecuentemente con genes que codifican para proteínas de funciones relevantes y/o implicadas en procesos de virulencia (TS, RHS, MASP, etc.).
5. Existen huellas Pr77Sh de idéntica composición nucleotídica e idéntico sitio de inserción dentro de una misma cepa, así como en diferentes cepas del parásito.
6. En el genoma de todas las cepas analizadas, con excepción de Dm28, existen huellas Pr77Sh que mantienen conservado el motivo DPE (*Downstream Promoter Element*) en términos de composición de la secuencia y localización de la misma respecto al sitio de inicio de la transcripción (nucleótido +1 de Pr77).
7. Una gran proporción de las huellas Pr77Sh identificadas en las cepas de *T. cruzi* a estudio adoptan una estructura secundaria compatible con la ribozima de tipo HDV presente en L1Tc, a pesar de presentar algunas divergencias en su secuencia nucleotídica, respecto a L1TcRz, mostrando actividad ribozima en ensayos *in vitro* de corte co-transcripcional y post-transcripcional.
8. El ensamblaje y anotación del genoma de la cepa SOL de *T. cruzi* secuenciado por tecnología PacBio ha permitido la identificación de 78 retroelementos L1Tc, 93 NARTc y 107 elementos DIREs, así como un análisis preciso del contenido génico de esta cepa.

9. Análisis de la expresión génica en las tres formas del parásito (cepa SOL) muestra una expresión diferencial de un elevado número de genes y de retroelementos en la forma infectiva (tripomastigote) e intracelular (amastigote) respecto a la forma epimastigote. La expresión de genes codificantes para TS, MASP, TASV y mucinas es superior con significancia estadística en el estadio tripomastigote y la expresión de GP63 y DGF-1 es superior en las formas tripomastigotes y amastigotes.

### ***Biomarcadores de respuesta serológica***

10. Los antígenos de *T. cruzi* KMP11, PFR2 y 3973<sup>d</sup> son reconocidos con alta especificidad ( $\geq 93\%$ ) y sensibilidad ( $\geq 70\%$ ) por el suero de pacientes con enfermedad de Chagas en fase crónica y obedecen al criterio de biomarcadores de eficacia terapéutica establecido en el *Target product profile* (TPP) recientemente establecido para este tipo de biomarcadores.
11. La determinación de la reactividad del suero de pacientes de Chagas tratados con benznidazol frente al set de moléculas compuesto por KMP11, HSP70, PFR2 y 3973<sup>d</sup> muestra ser una herramienta útil como marcador temprano de eficacia del tratamiento en estos pacientes.

### ***Biomarcadores de respuesta celular***

12. Los pacientes crónicos de Chagas en fase indeterminada *versus* controles sanos, muestran una estadísticamente significativa activación del perfil de expresión de genes implicados en relevantes rutas de respuesta inmune, tras estimulación *in vitro* de las células de sangre periférica con proteínas totales solubles de *T. cruzi*.
13. Los sujetos sanos procedentes de áreas endémicas y no endémicas de la enfermedad de Chagas muestran un similar perfil de expresión de los genes implicados en la respuesta inmune celular específica frente a antígenos del parásito.
14. Los pacientes con enfermedad de Chagas en fase crónica indeterminada y cardiaca muestran patrones diferenciales de expresión de genes implicados en la respuesta inmune celular antígeno-específica.
15. El tratamiento con benznidazol modula el patrón de expresión de los genes implicados en la respuesta inmune celular antígeno-específica, activados en pacientes con enfermedad crónica de Chagas, tanto asintomáticos como con sintomatología cardiaca.

## **BIBLIOGRAFÍA**

---



- Abrahamsohn, I. A. (1998). Cytokines in innate and acquired immunity to *Trypanosoma cruzi* infection. *Brazilian Journal of Medical and Biological Research*, 31(1), 117–121. <https://doi.org/10.1590/S0100-879X1998000100015>
- Abrahamsohn, I. A., and Coffman, R. L. (1996). *Trypanosoma cruzi*: IL-10, TNF, IFN-gamma, and IL-12 regulate innate and acquired immunity to infection. *Experimental Parasitology*, 84(2), 231–244. <https://doi.org/10.1006/EXPR.1996.0109>
- Abramo-Orrego, L., Lansetti, J. C., Bozzini, J. P., and Wynne de Martini, G. J. (1980). Hemocultivo como método de diagnóstico de la enfermedad de Chagas. *Medicina*, 40(Suppl 1), 56–62.
- Acosta-Rodríguez, E. v., Montes, C. L., Motrán, C. C., Zuniga, E. I., Liu, F. T., Rabinovich, G. A., and Gruppi, A. (2004). Galectin-3 mediates IL-4-induced survival and differentiation of B cells: functional cross-talk and implications during *Trypanosoma cruzi* infection. *Journal of Immunology*, 172(1), 493–502. <https://doi.org/10.4049/JIMMUNOL.172.1.493>
- Albareda, M. C., and Laucella, S. A. (2015). Modulation of *Trypanosoma cruzi*-specific T-cell responses after chemotherapy for chronic Chagas disease. *Memorias Do Instituto Oswaldo Cruz*, 110(3), 414–421. <https://doi.org/10.1590/0074-02760140386>
- Albareda, M. C., Olivera, G. C., Laucella, S. A., Alvarez, M. G., Fernandez, E. R., Lococo, B., Viotti, R., Tarleton, R. L., and Postan, M. (2009). Chronic human infection with *Trypanosoma cruzi* drives CD4+ T cells to immune senescence. *Journal of Immunology*, 183(6), 4103–4108. <https://doi.org/10.4049/JIMMUNOL.0900852>
- Aliberti, J. C. S., Cardoso, M. A. G., Martins, G. A., Gazzinelli, R. T., Vieira, L. Q., and Silva, J. S. (1996). Interleukin-12 mediates resistance to *Trypanosoma cruzi* in mice and is produced by murine macrophages in response to live trypomastigotes. *Infection and Immunity*, 64(6), 1961–1967. <https://doi.org/10.1128/iai.64.6.1961-1967.1996>
- Alonso-Padilla, J., Abril, M., de Noya, B. A., Almeida, I. C., Angheben, A., Jorge, T. A., Chatelain, E., Esteva, M., Gascón, J., Grijalva, M. J., Guhl, F., Hasslochermoreno, A. M., López, M. C., Luquetti, A., Noya, O., Pinazo, M. J., Ramsey, J. M., Ribeiro, I., Ruiz, A. M., ... Picado, A. (2020). Target product profile for a test for the early assessment of treatment efficacy in Chagas disease patients: An expert consensus. *PLoS Neglected Tropical Diseases*, 14(4), e0008035. <https://doi.org/10.1371/JOURNAL.PNTD.0008035>
- Alonso-Padilla, J., López, M. C., Esteva, M., Zrein, M., Casellas, A., Gómez, I., Granjon, E., Méndez, S., Benítez, C., Ruiz, A. M., Sanz, S., Gascón, J., Thomas, M. C., Pinazo, M. J., Abril, M., de Noya, B. A., Jorge, T. A., Chatelain, E., Grijalva, M. J., ... Viotti, R. (2021). Serological reactivity against T. cruzi-derived antigens: Evaluation of their suitability for the assessment of response to treatment in chronic Chagas disease. *Acta Tropica*, 221, 105990. <https://doi.org/10.1016/j.actatropica.2021.105990>
- Alonso-Vega, C., Billot, C., and Torrico, F. (2013). Achievements and challenges upon the implementation of a program for national control of congenital Chagas in Bolivia: Results 2004-2009. *PLoS Neglected Tropical Diseases*, 7(7), 2304. <https://doi.org/10.1371/journal.pntd.0002304>
- Altcheh, J., Biancardi, M., Lapeña, A., Ballering, G., and Freilij, H. (2005). Congenital Chagas disease: experience in the Hospital de Niños, Ricardo Gutiérrez, Buenos Aires, Argentina. *Revista Da Sociedade Brasileira de Medicina Tropical*, 38(Suppl 2), 41–45.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

## BIBLIOGRAFÍA

---

- Andersen, C. L., Jensen, J. L., and Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, 64(15), 5245–5250. <https://doi.org/10.1158/0008-5472.CAN-04-0496>
- Andrés-León, E., and Rojas, A. M. (2019). miARma-Seq, a comprehensive pipeline for the simultaneous study and integration of miRNA and mRNA expression data. *Methods*, 152, 31–40. <https://doi.org/10.1016/J.YMETH.2018.09.002>
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Antinori, S., Galimberti, L., Bianco, R., Grande, R., Galli, M., and Corbellino, M. (2017). Chagas disease in Europe: A review for the internist in the globalized world. *European Journal of Internal Medicine*, 43, 6–15. <https://doi.org/10.1016/j.ejim.2017.05.001>
- Araújo, C. A. C., Waniek, P. J., Xavier, S. C. C., and Jansen, A. M. (2011). Genotype variation of *Trypanosoma cruzi* isolates from different Brazilian biomes. *Experimental Parasitology*, 127(1), 308–312. <https://doi.org/10.1016/j.exppara.2010.07.013>
- Arrigo, A. P., Virost, S., Chaufour, S., Firdaus, W., Kretz-Remy, C., and Diaz-Latoud, C. (2005). Hsp27 consolidates intracellular redox homeostasis by upholding glutathione in its reduced form and by decreasing iron intracellular levels. *Antioxidants and Redox Signaling*, 7(3–4), 414–424. <https://doi.org/10.1089/ars.2005.7.414>
- Bartholomeu, D. C., Cerqueira, G. C., Leão, A. C. A., daRocha, W. D., Pais, F. S., Macedo, C., Djikeng, A., Teixeira, S. M. R., and El-Sayed, N. M. (2009). Genomic organization and expression profile of the mucin-associated surface protein (masp) family of the human pathogen *Trypanosoma cruzi*. *Nucleic Acids Research*, 37(10), 3407–3417. <https://doi.org/10.1093/nar/gkp172>
- Been, M. D. (2006). HDV ribozymes. *Current Topics in Microbiology and Immunology*, 307, 47–65. [https://doi.org/10.1007/3-540-29802-9\\_3](https://doi.org/10.1007/3-540-29802-9_3)
- Been, M. D., Perrotta, A. T., and Rosenstein, S. P. (1992). Secondary structure of the self-cleaving RNA of hepatitis delta virus: applications to catalytic RNA design. *Biochemistry*, 31(47), 11843–11852. <https://doi.org/10.1021/BI00162A024>
- Berberich, C., Requena, J. M., and Alonso, C. (1997). Cloning of genes and expression and antigenicity analysis of the *Leishmania infantum* KMP-11 protein. *Experimental Parasitology*, 85(1), 105–108. <https://doi.org/10.1006/EXPR.1996.4120>
- Bern, C. (2015). Chagas' Disease. *New England Journal of Medicine*, 373(5), 456–466. <https://doi.org/10.1056/NEJMra1410150>
- Bern, C., Martin, D. L., and Gilman, R. H. (2011). Acute and congenital Chagas disease. *Advances in Parasitology*, 75, 19–47. <https://doi.org/10.1016/B978-0-12-385863-4.00002-2>
- Bernabó, G., Levy, G., Ziliani, M., Caeiro, L. D., Sánchez, D. O., and Tekiel, V. (2013). TcTASV-C, a protein family in *Trypanosoma cruzi* that is predominantly trypomastigote-stage specific and secreted to the medium. *PLoS One*, 8(7), e71192. <https://doi.org/10.1371/journal.pone.0071192>
- Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., ... 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>
- Bittencourt, A. L. (1992). Possible risk factors for vertical transmission of Chagas' disease. *Revista Do Instituto de Medicina Tropical de São Paulo*, 34(5), 403–408. <https://doi.org/10.1590/S0036-46651992000500006>
- Bouhdidi, A. el, Truyens, C., Rivera, M. T., Bazin, H., and Carlier, Y. (1994). *Trypanosoma cruzi* infection in mice induces a polyisotypic hypergammaglobulinaemia and parasite-specific response involving high IgG2a concentrations and highly avid IgG1 antibodies. *Parasite Immunology*, 16(2), 69–76. <https://doi.org/10.1111/J.1365-3024.1994.TB00325.X>
- Boulesteix, M., Weiss, M., and Biéumont, C. (2006). Differences in genome size between closely related species: The *Drosophila melanogaster* species subgroup. *Molecular Biology and Evolution*, 23(1), 162–167. <https://doi.org/10.1093/molbev/msj012>
- Brener, Z., and Gazzinelli, R. T. (1997). Immunological control of *Trypanosoma cruzi* infection and pathogenesis of Chagas' disease. *International Archives of Allergy and Immunology*, 114(2), 103–110. <https://doi.org/10.1159/000237653>
- Brenière, S. F., Waleckx, E., and Barnabé, C. (2016). Over six thousand *Trypanosoma cruzi* strains classified into Discrete Typing Units (DTUs): Attempt at an inventory. *PLoS Neglected Tropical Diseases*, 10(8), e0004792. <https://doi.org/10.1371/JOURNAL.PNTD.0004792>
- Briceño, L., and Mosca, W. (1996). Defective production of interleukin 2 in patients with Chagas' disease. Purified IL-2 augments in vitro response in patients with chagasic cardiomyopathy. *Memorias Do Instituto Oswaldo Cruz*, 91(5), 601–607. <https://doi.org/10.1590/S0074-02761996000500011>
- Bridge, M. A., Zhou, Q., Koop, B. F., and Pearson, T. W. (1998). Cloning and characterization of the kinetoplastid membrane protein-11 gene locus of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, 91(2), 359–363. [https://doi.org/10.1016/S0166-6851\(97\)00229-6](https://doi.org/10.1016/S0166-6851(97)00229-6)
- Bringaud, F., Bartholomeu, D. C., Blandin, G., Delcher, A., Baltz, T., El-Sayed, N. M. A., and Ghedin, E. (2006). The *Trypanosoma cruzi* L1Tc and NARTc non-LTR retrotransposons show relative site specificity for insertion. *Molecular Biology and Evolution*, 23(2), 411–420. <https://doi.org/10.1093/molbev/msj046>
- Bringaud, F., Berriman, M., and Hertz-Fowler, C. (2009). Trypanosomatid genomes contain several subfamilies of ingi-related retroposons. *Eukaryot Cell*, 8(10), 1532–1542. <https://doi.org/10.1128/EC.00183-09>
- Bringaud, F., Biteau, N., Melville, S. E., Hez, S., El-Sayed, N. M., Leech, V., Berriman, M., Hall, N., Donelson, J. E., and Baltz, T. (2002). A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryotic Cell*, 1(1), 137–151. <https://doi.org/10.1128/EC.1.1.137-151.2002>
- Bringaud, F., García-Pérez, J. L., Heras, S. R., Ghedin, E., El-Sayed, N. M., Andersson, B., Baltz, T., and López, M. C. (2002). Identification of non-autonomous non-LTR retrotransposons in the genome of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, 124(1–2), 73–78. [https://doi.org/10.1016/S0166-6851\(02\)00167-6](https://doi.org/10.1016/S0166-6851(02)00167-6)
- Bringaud, F., Ghedin, E., Blandin, G., Bartholomeu, D. C., Caler, E., Levin, M. J., Baltz, T., and El-Sayed, N. M. (2006). Evolution of non-LTR retrotransposons in the trypanosomatid

- genomes: *Leishmania major* has lost the active elements. *Molecular and Biochemical Parasitology*, 145(2), 158–170. <https://doi.org/10.1016/j.molbiopara.2005.09.017>
- Bringaud, F., Müller, M., Cerqueira, G. C., Smith, M., Rochette, A., El-Sayed, N. M. A., Papadopoulou, B., and Ghedin, E. (2007). Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*. *PLoS Pathogens*, 3(9), 1291–1307. <https://doi.org/10.1371/JOURNAL.PPAT.0030136>
- Bringaud, F., Rogers, M., and Ghedin, E. (2015). Identification and analysis of ingi-related retroposons in the trypanosomatid genomes. *Methods in Molecular Biology*, 1201, 109–122. [https://doi.org/10.1007/978-1-4939-1438-8\\_6](https://doi.org/10.1007/978-1-4939-1438-8_6)
- Brisse, S., Dujardin, J. C., and Tibayrenc, M. (2000). Identification of six *Trypanosoma cruzi* lineages by sequence-characterised amplified region markers. *Molecular and Biochemical Parasitology*, 111(1), 95–105. [https://doi.org/10.1016/S0166-6851\(00\)00302-9](https://doi.org/10.1016/S0166-6851(00)00302-9)
- Brisse, S., Verhoef, J., and Tibayrenc, M. (2001). Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. *International Journal for Parasitology*, 31(11), 1218–1226. [https://doi.org/10.1016/S0020-7519\(01\)00238-7](https://doi.org/10.1016/S0020-7519(01)00238-7)
- Budzko, D. B., Pizzimenti, M. C., and Kierszenbaum, F. (1975). Effects of complement depletion in experimental Chagas disease: immune lysis of virulent blood forms of *Trypanosoma cruzi*. *Infection and Immunity*, 11(1), 86–91. <https://doi.org/10.1128/IAI.11.1.86-91.1975>
- Burleigh, B. A., and Andrews, N. W. (1995). The mechanisms of *Trypanosoma cruzi* invasion of mammalian cells. *Annual Review of Microbiology*, 49(1), 175–200. <https://doi.org/10.1146/annurev.mi.49.100195.001135>
- Bustamante, J. M., Presti, M. S. L., Rivarola, H. W., Fernández, A. R., Enders, J. E., Fretes, R. E., and Paglini-Oliva, P. (2007). Treatment with benznidazole or thioridazine in the chronic phase of experimental Chagas disease improves cardiopathy. *International Journal of Antimicrobial Agents*, 29(6), 733–737. <https://doi.org/10.1016/j.ijantimicag.2007.01.014>
- Caeiro, L. D., Alba-Soto, C. D., Rizzi, M., Solana, M. E., Rodriguez, G., Chidichimo, A. M., Rodriguez, M. E., Sánchez, D. O., Levy, G. v., and Tekiel, V. (2018). The protein family TcTASV-C is a novel *Trypanosoma cruzi* virulence factor secreted in extracellular vesicles by trypomastigotes and highly expressed in bloodstream forms. *PLoS Neglected Tropical Diseases*, 12(5), e0006475. <https://doi.org/10.1371/journal.pntd.0006475>
- Cai, C. W., Blase, J. R., Zhang, X., Eickhoff, C. S., and Hoft, D. F. (2016). Th17 cells are more protective than Th1 cells against the intracellular parasite *Trypanosoma cruzi*. *PLoS Pathogens*, 12(10), e1005902. <https://doi.org/10.1371/journal.ppat.1005902>
- Callejas-Hernández, F., Gironès, N., and Fresno, M. (2018). Genome sequence of *Trypanosoma cruzi* strain Bug2148. *Genome Announcements*, 6(3), e01497-17. <https://doi.org/10.1128/GENOMEA.01497-17>
- Callejas-Hernández, F., Rastrojo, A., Poveda, C., Gironès, N., and Fresno, M. (2018). Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. *Scientific Reports*, 8(1), 14631. <https://doi.org/10.1038/S41598-018-32877-2>

- Camargo, M. E. (1966). Fluorescent antibody test for the serodiagnosis of American trypanosomiasis. Technical modification employing preserved culture forms of *Trypanosoma cruzi* in a slide test. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, 8(5), 227–235.
- Camargo, M. E., Hoshino, S., Corrêa, N. S., and Peres, B. A. (1971). Hemagglutination test for Chagas' disease with chromium chloride, formalin-treated erythrocytes, sensitized with *Trypanosoma cruzi* extracts. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, 13(1), 45–50.
- Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984). Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. *Nature*, 311(5984), 350–355. <https://doi.org/10.1038/311350a0>
- Cançado, J. R. (1999). Criteria of Chagas disease cure. *Memorias Do Instituto Oswaldo Cruz*, 94(Suppl 1), 331–335. <https://doi.org/10.1590/S0074-02761999000700064>
- Cançado, J. R. (2002). Long term evaluation of etiological treatment of Chagas disease with benznidazole. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, 44(1), 29–37. <https://doi.org/10.1590/S0036-46652002000100006>
- Cano, M. I. N. (2001). Telomere biology of Trypanosomatids: More questions than answers. *Trends in Parasitology*, 17(9), 425–429. [https://doi.org/10.1016/S1471-4922\(01\)02014-1](https://doi.org/10.1016/S1471-4922(01)02014-1)
- Cardoni, R. L., Antunez, M. I., and Abrami, A. A. (1999). Respuesta TH1 en la infección experimental con *Trypanosoma cruzi*. *Medicina (Buenos Aires)*, 59(Suppl 2), 84–90.
- Carvalho, L. P., Passos, S., Dutra, W. O., Soto, M., Alonso, C., Gollob, K. J., Carvalho, E. M., and Ribeiro De Jesus, A. (2005). Effect of LACK and KMP11 on IFN-gamma production by peripheral blood mononuclear cells from cutaneous and mucosal leishmaniasis patients. *Scandinavian Journal of Immunology*, 61(4), 337–342. <https://doi.org/10.1111/j.1365-3083.2005.01581.X>
- Castro, J. A., de Mecca, M. M., and Bartel, L. C. (2006). Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Human and Experimental Toxicology*, 25(8), 471–479. <https://doi.org/10.1191/0960327106het653oa>
- Cerbán, F. M., Stempin, C. C., Volpini, X., Carrera Silva, E. A., Gea, S., and Motran, C. C. (2020). Signaling pathways that regulate *Trypanosoma cruzi* infection and immune response. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1866(5), 165707. <https://doi.org/10.1016/j.bbadis.2020.165707>
- Chain, C. Y., Pires Souto, D. E., Sbaraglini, M. L., Labriola, C. A., Daza Millone, M. A., Ramirez, E. A., Cisneros, J. S., Lopez-Albizu, C., Scollo, K., Kubota, L. T., Ruiz, A. M., and Vela, M. E. (2019). *Trypanosoma cruzi* virulence factors for the diagnosis of Chagas' disease. *ACS Infectious Diseases*, 5(11), 1813–1819. <https://doi.org/10.1021/ACSINFECDIS.9B00269>
- Chen, J., Ganguly, A., Miswan, Z., Hammes-Schiffer, S., Bevilacqua, P. C., and Golden, B. L. (2013). Identification of the catalytic Mg<sup>2+</sup> ion in the HDV ribozyme. *Biochemistry*, 52(3), 567. <https://doi.org/10.1021/BI3013092>
- Chen, J. H., Gong, B., Bevilacqua, P. C., Carey, P. R., and Golden, B. L. (2009). A catalytic metal ion interacts with the cleavage site G.U wobble in the HDV ribozyme. *Biochemistry*, 48(7), 1498–1507. <https://doi.org/10.1021/BI8020108>
- Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. (2010). A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid

- and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*, 49(31), 6508–6518. <https://doi.org/10.1021/BI100670P>
- Chen, Z., Erickson, D. L., and Meng, J. (2021). Polishing the Oxford Nanopore long-read assemblies of bacterial pathogens with Illumina short reads to improve genomic analyses. *Genomics*, 113(3), 1366–1377. <https://doi.org/10.1016/j.YGENO.2021.03.018>
- Chin-Hong, P. v., Schwartz, B. S., Bern, C., Montgomery, S. P., Kontak, S., Kubak, B., Morris, M. I., Nowicki, M., Wright, C., and Ison, M. G. (2011). Screening and treatment of Chagas disease in organ transplant recipients in the United States: recommendations from the Chagas in transplant working group. *American Journal of Transplantation*, 11(4), 672–680. <https://doi.org/10.1111/j.1600-6143.2011.03444.x>
- Cicora, F., Ecurra, V., Silguero, S., González, I. M., and Roberti, J. E. (2014). Use of kidneys from *Trypanosoma cruzi*-infected donors in naive transplant recipients without prophylactic therapy. *Transplantation*, 97(1), e3–e4. <https://doi.org/10.1097/01.TP.0000437673.86339.82>
- Clark, CG., and Pung, O. J. (1994). Host specificity of ribosomal DNA variation in sylvatic *Trypanosoma cruzi* from North America. *Molecular and Biochemical Parasitology*, 66(1), 175–179. [https://doi.org/10.1016/0166-6851\(94\)90052-3](https://doi.org/10.1016/0166-6851(94)90052-3)
- Clayton, C. E. (2002). Life without transcriptional control? From fly to man and back again. *EMBO Journal*, 21(8), 1881–1888. <https://doi.org/10.1093/emboj/21.8.1881>
- Cooley, G., Etheridgel, R. D., Boehlke, C., Bundy, B., Weatherly, D. B., Minning, T., Haney, M., Postan, M., Laucella, S., and Tarleton, R. L. (2008). High throughput selection of effective serodiagnostics for *Trypanosoma cruzi* infection. *PLoS Neglected Tropical Diseases*, 2(10), e316. <https://doi.org/10.1371/journal.pntd.0000316>
- Coura, J. R., and de Castro, S. L. (2002). A critical review on Chagas disease chemotherapy. *Memorias Do Instituto Oswaldo Cruz*, 97(1), 3–24. <https://doi.org/10.1590/S0074-02762002000100001>
- Coura, J. R., and Dias, J. C. P. (2009). Epidemiology, control and surveillance of Chagas disease - 100 years after its discovery. *Memorias Do Instituto Oswaldo Cruz*, 104(Suppl 1), 31–40. <https://doi.org/10.1590/S0074-02762009000900006>
- Cuellar, A., Santander, S. P., Thomas, M. D. C., Guzmán, F., Gómez, A., López, M. C., and Puerta, C. J. (2008). Monocyte-derived dendritic cells from chagasic patients vs healthy donors secrete differential levels of IL-10 and IL-12 when stimulated with a protein fragment of *Trypanosoma cruzi* heat-shock protein-70. *Immunology and Cell Biology*, 86(3), 255–260. <https://doi.org/10.1038/SJ.ICB.7100146>
- Cunha-Neto, E., Dzau, V. J., Allen, P. D., Stamatiou, D., Benvenuto, L., Higuchi, M. L., Koyama, N. S., Silva, J. S., Kalil, J., and Liew, C. C. (2005). Cardiac gene expression profiling provides evidence for cytokinopathy as a molecular mechanism in Chagas' disease cardiomyopathy. *American Journal of Pathology*, 167(2), 305–313. [https://doi.org/10.1016/S0002-9440\(10\)62976-8](https://doi.org/10.1016/S0002-9440(10)62976-8)
- Cura, C. I., Mejía-Jaramillo, A. M., Duffy, T., Burgos, J. M., Rodríguez, M., Cardinal, M. v., Kjos, S., Gurgel-Gonçalves, R., Blanchet, D., de Pablos, L. M., Tomasini, N., da Silva, A., Russomando, G., Cuba, C. A. C., Aznar, C., Abate, T., Levin, M. J., Osuna, A., Gürtler, R. E., ... Schijman, A. G. (2010). *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. *International Journal for Parasitology*, 40(14), 1599–1607. <https://doi.org/10.1016/j.ijpara.2010.06.006>

- de Andrade, A. L., Zicker, F., de Oliveira, R. M., Almeida E Silva, S., Luquetti, A., Travassos, L. R., Almeida, I. C., de Andrade, S. S., Guimarães De Andrade, J., and Martelli, C. M. T. (1996). Randomised trial of efficacy of benznidazole in treatment of early *Trypanosoma cruzi* infection. *Lancet*, 348(9039), 1407–1413. [https://doi.org/10.1016/S0140-6736\(96\)04128-1](https://doi.org/10.1016/S0140-6736(96)04128-1)
- de Andrade, A. L., Zicker, F., and Martelli, C. M. T. (1994). An epidemiological approach to study congenital Chagas' disease. *Cadernos de Saúde Pública*, 10(Suppl 2), S345–S351. <https://doi.org/10.1590/s0102-311x1994000800012>
- De Avalos, S. V., Blader, I. J., Fisher, M., Boothroyd, J. C., and Burleigh, B. A. (2002). Immediate/early response to *Trypanosoma cruzi* infection involves minimal modulation of host cell transcription. *Journal of Biological Chemistry*, 277(1), 639–644. <https://doi.org/10.1074/jbc.M109037200>
- de la Peña, M., García-Robles, I., and Cervera, A. (2017). The Hammerhead ribozyme: A long history for a short RNA. *Molecules*, 22(1), 78. <https://doi.org/10.3390/MOLECULES22010078>
- de Meis, J., Morrot, A., Farias-de-Oliveira, D. A., Villa-Verde, D. M. S., and Savino, W. (2009). Differential regional immune response in Chagas disease. *PLoS Neglected Tropical Diseases*, 3(7), e417. <https://doi.org/10.1371/journal.pntd.0000417>
- de Moraes, C. G. V., Castro Lima, A. K., Terra, R., dos Santos, R. F., Da-Silva, S. A. G., and Dutra, P. M. L. (2015). The dialogue of the host-parasite relationship: *Leishmania* spp. and *Trypanosoma cruzi* infection. *BioMed Research International*, 2015, 324915. <https://doi.org/10.1155/2015/324915>
- de Noya, B. A., Díaz-Bello, Z., Colmenares, C., Ruiz-Guevara, R., Mauriello, L., Muñoz-Calderón, A., and Noya, O. (2015). Update on oral Chagas disease outbreaks in Venezuela: Epidemiological, clinical and diagnostic approaches. *Memorias Do Instituto Oswaldo Cruz*, 110(3), 377–386. <https://doi.org/10.1590/0074-02760140285>
- de Noya, B. A., Díaz-Bello, Z., Colmenares, C., Ruiz-Guevara, R., Mauriello, L., Zavala-Jaspe, R., Suarez, J. A., Abate, T., Naranjo, L., Paiva, M., Rivas, L., Castro, J., Márques, J., Mendoza, I., Acquatella, H., Torres, J., and Noya, O. (2010). Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. *Journal of Infectious Diseases*, 201(9), 1308–1315. <https://doi.org/10.1086/651608>
- de Pablos, L. M., González, G. G., Parada, J. S., Hidalgo, V. S., Lozano, I. M. D., Samblás, M. M. G., Bustos, T. C., and Osuna, A. (2011). Differential expression and characterization of a member of the mucin-associated surface protein family secreted by *Trypanosoma cruzi*. *Infection and Immunity*, 79(10), 3993–4001. <https://doi.org/10.1128/IAI.05329-11>
- de Souza, W. (2009). Structural organization of *Trypanosoma cruzi*. *Memorias Do Instituto Oswaldo Cruz*, 104(Suppl 1), 89–100. <https://doi.org/10.1590/s0074-02762009000900014>
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*, 35(1), 41–48. <https://doi.org/10.1038/ng1223>
- Díaz de Toranzo, E. G., Castro, J. A., Franke de Cazzulo, B. M., and Cazzulo, J. J. (1988). Interaction of benznidazole reactive metabolites with nuclear and kinetoplasmic DNA, proteins and lipids from *Trypanosoma cruzi*. *Experientia*, 44(10), 880–881. <https://doi.org/10.1007/BF01941187>

- d'Imperio Lima, M. R., Eisen, H., Minoprio, P., Joskowicz, M., and Coutinho, A. (1986). Persistence of polyclonal B cell activation with undetectable parasitemia in late stages of experimental Chagas' disease. *The Journal of Immunology*, 137(1), 353–356.
- Docampo, R., and Stoppani, A. O. M. (1979). Generation of superoxide anion and hydrogen peroxide induced by nifurtimox in *Trypanosoma cruzi*. *Archives of Biochemistry and Biophysics*, 197(1), 317–321. [https://doi.org/10.1016/0003-9861\(79\)90251-0](https://doi.org/10.1016/0003-9861(79)90251-0)
- Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M. D. (2001). The “cleavage” activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring “2A-like” sequences. *Journal of General Virology*, 82(5), 1027–1041. <https://doi.org/10.1099/0022-1317-82-5-1027>
- dos Santos, S. L., Freitas, L. M., Lobo, F. P., Rodrigues-Luiz, G. F., Mendes, T. A. de O., Oliveira, A. C. S., Andrade, L. O., Chiari, É., Gazzinelli, R. T., Teixeira, S. M. R., Fujiwara, R. T., and Bartholomeu, D. C. (2012). The MASP family of *Trypanosoma cruzi*: Changes in gene expression and antigenic profile during the acute phase of experimental infection. *PLoS Neglected Tropical Diseases*, 6(8), e1779. <https://doi.org/10.1371/journal.pntd.0001779>
- dos Santos Virgilio, F., Pontes, C., Dominguez, M. R., Ersching, J., Rodrigues, M. M., and Vasconcelos, J. R. (2014). CD8+ T cell-mediated immunity during *Trypanosoma cruzi* infection: A path for vaccine development? *Mediators of Inflammation*, 2014, 243786. <https://doi.org/10.1155/2014/243786>
- Duffy, T., Bisio, M., Altcheh, J., Burgos, J. M., Diez, M., Levin, M. J., Favaloro, R. R., Freilij, H., and Schijman, A. G. (2009). Accurate real-time PCR strategy for monitoring bloodstream parasitic loads in Chagas disease patients. *PLoS Neglected Tropical Diseases*, 3(4), e419. <https://doi.org/10.1371/journal.pntd.0000419>
- Dutra, W. O., Gollob, K. J., Pinto-Dias, J. C., Gazzinelli, G., Correa-Oliveira, R., Coffman, R. L., and Carvalho-Parra, J. F. (1997). Cytokine mRNA profile of peripheral blood mononuclear cells isolated from individuals with *Trypanosoma cruzi* chronic infection. *Scandinavian Journal of Immunology*, 45(1), 74–80. <https://doi.org/10.1046/j.1365-3083.1997.d01-362.x>
- Dutra, W. O., Martins-filho, O. A., Cañado, J. R., Pinto-dias, J. C., Brener, Z., Freeman, G. L., Colley, D. G., Gazzinelli, G., and Parra, J. C. (1994). Activated T and B lymphocytes in peripheral blood of patients with Chagas' disease. *International Immunology*, 6(4), 499–506. <https://doi.org/10.1093/intimm/6.4.499>
- Egui, A., Lasso, P., Pérez-Antón, E., Thomas, M. C., and López, M. C. (2018). Dynamics of T cells repertoire during *Trypanosoma cruzi* infection and its post-treatment modulation. *Current Medicinal Chemistry*, 26(36), 6519–6543. <https://doi.org/10.2174/0929867325666181101111819>
- Egui, A., López, M. C., Gómez, I., Simón, M., Segovia, M., and Thomas, M. C. (2020). Differential phenotypic and functional profile of epitope-specific cytotoxic CD8+ T cells in benznidazole-treated chronic asymptomatic Chagas disease patients. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1866(3), 165629. <https://doi.org/10.1016/j.bbadis.2019.165629>
- Egui, A., Thomas, M. C., Carrilero, B., Segovia, M., Alonso, C., Marañón, C., and López, M. C. (2015). Differential phenotypic and functional profiles of TcCA-2 -specific cytotoxic CD8+ T cells in the asymptomatic versus cardiac phase in Chagasic patients. *PLoS One*, 10(3), e0122115–e0122115. <https://doi.org/10.1371/JOURNAL.PONE.0122115>

- Egui, A., Thomas, M. C., Fernández-Villegas, A., Pérez-Antón, E., Gómez, I., Carrilero, B., del Pozo, Á., Ceballos, M., Andrés-León, E., López-Ruz, M. Á., Gainza, E., Oquiñena, E., Segovia, M., and López, M. C. (2019). A parasite biomarker set for evaluating benznidazole treatment efficacy in patients with chronic asymptomatic *Trypanosoma cruzi* infection. *Antimicrobial Agents and Chemotherapy*, 63(10), e02436-18. <https://doi.org/10.1128/AAC.02436-18>
- Egui, A., Thomas, M. C., Morell, M., Marañón, C., Carrilero, B., Segovia, M., Puerta, C. J., Pinazo, M. J., Rosas, F., Gascón, J., and López, M. C. (2012). *Trypanosoma cruzi* paraflagellar rod proteins 2 and 3 contain immunodominant CD8(+) T-cell epitopes that are recognized by cytotoxic T cells from Chagas disease patients. *Molecular Immunology*, 52(3-4), 289-298. <https://doi.org/10.1016/j.molimm.2012.05.021>
- Eickbush, D. G., and Eickbush, T. H. (2010). R2 Retrotransposons Encode a Self-Cleaving Ribozyme for Processing from an rRNA Cotranscript. *Molecular and Cellular Biology*, 30(13), 3150. <https://doi.org/10.1128/MCB.00300-10>
- Eickbush, TH., and Jamburuthugoda, VK. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research*, 134(1-2), 221-234. <https://doi.org/10.1016/j.virusres.2007.12.010>
- El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., ... Andersson, B. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, 309(5733), 409-415. <https://doi.org/10.1126/science.1112631>
- El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renaud, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., ... Hall, N. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science*, 309(5733), 404-409. <https://doi.org/10.1126/science.1112181>
- Engels, W. R. (1979). Hybrid dysgenesis in *Drosophila melanogaster*: Rules of inheritance of female sterility. *Genetical Research*, 33(3), 219-236. <https://doi.org/10.1017/S001667230001836X>
- Engvall, E., and Perlmann, P. (1971). Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry*, 8(9), 871-874. [https://doi.org/10.1016/0019-2791\(71\)90454-X](https://doi.org/10.1016/0019-2791(71)90454-X)
- Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics*, 24(4), 363-367. <https://doi.org/10.1038/74184>
- Fabbro, D. L., Olivera, V., Bizai, M. L., Denner, S., Diez, C., Mancipar, I., Streiger, M., Arias, E., del Barco, M., Mendicino, D., and Bottasso, O. (2011). Humoral immune response against P2β from *Trypanosoma cruzi* in persons with chronic Chagas disease: Its relationship with treatment against parasites and myocardial damage. *American Journal of Tropical Medicine and Hygiene*, 84(4), 575-580. <https://doi.org/10.4269/ajtmh.2011.10-0261>
- Fabbro, D. L., Streiger, M. L., Arias, E. D., Bizai, M. L., del Barco, M., and Amicone, N. A. (2007). Trypanocide treatment among adults with chronic Chagas disease living in Santa Fe city (Argentina), over a mean follow-up of 21 years: Parasitological, serological and clinical evolution. *Revista Da Sociedade Brasileira de Medicina Tropical*, 40(1), 1-10. <https://doi.org/10.1590/s0037-86822007000100001>

- Fabbro, D. L., Velazquez, E., Bizai, M. L., Denner, S., Olivera, V., Arias, E., Pravia, C., and Ruiz, A. M. (2013). Evaluation of the ELISA-F29 test as an early marker of therapeutic efficacy in adults with chronic Chagas disease. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, 55(3), 167–172. <https://doi.org/10.1590/S0036-46652013000300005>
- Falla, A., Herrera, C., Fajardo, A., Montilla, M., Vallejo, G. A., and Guhl, F. (2009). Haplotype identification within *Trypanosoma cruzi* I in Colombian isolates from several reservoirs, vectors and humans. *Acta Tropica*, 110(1), 15–21. <https://doi.org/10.1016/j.actatropica.2008.12.003>
- Fernandes, O., Mangia, R. H., Lisboa, C. v., Pinho, A. P., Morel, C. M., Zingales, B., Campbell, D. A., and Jansen, A. M. (1999). The complexity of the sylvatic cycle of *Trypanosoma cruzi* in Rio de Janeiro state (Brazil) revealed by the non-transcribed spacer of the mini-exon gene. *Parasitology*, 118(2), 161–166. <https://doi.org/10.1017/S0031182098003709>
- Fernández-Villegas, A., Pinazo, M. J., Marañón, C., Thomas, M. C., Posada, E., Carrilero, B., Segovia, M., Gascon, J., and López, M. C. (2011). Short-term follow-up of chagasic patients after benznidazole treatment using multiple serological markers. *BMC Infectious Diseases*, 11, 206. <https://doi.org/10.1186/1471-2334-11-206>
- Fernández-Villegas, A., Thomas, M. C., Carrilero, B., Lasso, P., Egui, A., Murcia, L., Segovia, M., Alonso, C., and López, M. C. (2016). A 12-mer repetitive antigenic epitope from *Trypanosoma cruzi* is a potential marker of therapeutic efficacy in chronic Chagas' disease. *Journal of Antimicrobial Chemotherapy*, 71(7), 2005–2009. <https://doi.org/10.1093/jac/dkw090>
- Fernández-Villegas, A., Thomas, M. C., Carrilero, B., Téllez, C., Marañón, C., Murcia, L., Moralo, S., Alonso, C., Segovia, M., and López, M. C. (2014). The innate immune response status correlates with a divergent clinical course in congenital Chagas disease of twins born in a non-endemic country. *Acta Tropica*, 140, 84–90. <https://doi.org/10.1016/j.actatropica.2014.08.006>
- Ferreira, R. C., Ianni, B. M., Abel, L. C. J., Buck, P., Mady, C., Kalil, J., and Cunha-Neto, E. (2003). Increased plasma levels of tumor necrosis factor- $\alpha$  in asymptomatic/"Indeterminate" and Chagas disease cardiomyopathy patients. *Memorias Do Instituto Oswaldo Cruz*, 98(3), 407–411. <https://doi.org/10.1590/s0074-02762003000300021>
- Fife, E. H., and Muschel, L. H. (1959). Fluorescent-antibody technic for serodiagnosis of *Trypanosoma cruzi* infection. *Proceedings of the Society for Experimental Biology and Medicine*, 101(3), 540–543. <https://doi.org/10.3181/00379727-101-25009>
- Flechas, I. D., Cuellar, A., Cucunubá, Z. M., Rosas, F., Velasco, V., Steindel, M., Thomas, M. D. C., López, M. C., González, J. M., and Puerta, C. J. (2009). Characterising the KMP-11 and HSP-70 recombinant antigens' humoral immune response profile in chagasic patients. *BMC Infectious Diseases*, 9, 186. <https://doi.org/10.1186/1471-2334-9-186>
- Fragata-Filho, A. A., França, F. F., Fragata, C. da S., Lourenço, A. M., Faccini, C. C., and Costa, C. A. de J. (2016). Evaluation of parasiticide treatment with benznidazol in the electrocardiographic, clinical, and serological evolution of Chagas disease. *PLoS Neglected Tropical Diseases*, 10(3), e0004508. <https://doi.org/10.1371/journal.pntd.0004508>
- Franzen, O., Ochaya, S., Sherwood, E., Lewis, M. D., Llewellyn, M. S., Miles, M. A., and Andersson, B. (2011). Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. *PLoS Neglected Tropical Diseases*, 5(3), e984. <https://doi.org/10.1371/journal.pntd.0000984>



- Frasch, A. C. C. (2000). Functional diversity in the trans-sialidase and mucin families in *Trypanosoma cruzi*. *Parasitology Today*, 16(7), 282–286. [https://doi.org/10.1016/S0169-4758\(00\)01698-7](https://doi.org/10.1016/S0169-4758(00)01698-7)
- Freilij, H., and Altcheh, J. (1995). Congenital Chagas' disease: Diagnostic and clinical aspects. *Clinical Infectious Diseases*, 21(3), 551–555. <https://doi.org/10.1093/clinids/21.3.551>
- Garcia, S., Ramos, C. O., Senra, J. F. V., Vilas-Boas, F., Rodrigues, M. M., Campos-De-Carvalho, A. C., Ribeiro-Dos-Santos, R., and Soares, M. B. P. (2005). Treatment with benznidazole during the chronic phase of experimental Chagas' disease decreases cardiac alterations. *Antimicrobial Agents and Chemotherapy*, 49(4), 1521–1528. <https://doi.org/10.1128/AAC.49.4.1521-1528.2005>
- García-Pérez, J. L., González, C. I., Thomas, M. C., Olivares, M., and López, M. C. (2003). Characterization of reverse transcriptase activity of the L1Tc retroelement from *Trypanosoma cruzi*. *Cellular and Molecular Life Sciences*, 60(12), 2692–2701. <https://doi.org/10.1007/s00018-003-3342-y>
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., ... Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906), 498–511. <https://doi.org/10.1038/nature01097>
- Garrido, C., Ottavi, P., Fromentin, A., Hammann, A., Arrigo, A. P., Chauffert, B., and Mehlen, P. (1997). HSP27 as a mediator of confluence-dependent resistance to cell death induced by anticancer drugs. *Cancer Research*, 57(13), 2661–2667.
- Garrido, C., Schmitt, E., Candé, C., Vahsen, N., Parcellier, A., and Kroemer, G. (2003). HSP27 and HSP70: potentially oncogenic apoptosis inhibitors. *Cell Cycle*, 2(6), 578–583. <https://doi.org/10.4161/cc.2.6.521>
- Gately, M. K., Warriar, R. R., Honasoge, S., Carvajal, D. M., Faherty, D. A., Connaughton, S. E., Anderson, T. D., Sarmiento, U., Hubbard, B. R., and Murphy, M. (1994). Administration of recombinant IL-12 to normal mice enhances cytolytic lymphocyte activity and induces production of IFN- $\gamma$  in vivo. *International Immunology*, 6(1), 157–167. <https://doi.org/10.1093/intimm/6.1.157>
- Gazzinelli, R. T., Oswald, I. P., Hieny, S., James, S. L., and Sher, A. (1992). The microbicidal activity of interferon- $\gamma$ -treated macrophages against *Trypanosoma cruzi* involves an L-arginine-dependent, nitrogen oxide-mediated mechanism inhibitable by interleukin-10 and transforming growth factor- $\beta$ . *European Journal of Immunology*, 22(10), 2501–2506. <https://doi.org/10.1002/eji.1830221006>
- 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>
- Goel, A., Vohra, H., and Varshney, G. C. (1999). Strain-specific recognition of live *Leishmania donovani* promastigotes by homologous antiserum raised against a crude membrane fraction of infected macrophages. *Parasitology Research*, 85(1), 19–24. <https://doi.org/10.1007/S004360050501>
- Gómez, I., López, M. C., Rastrojo, A., Lorenzo-Díaz, F., Requena, J. M., Aguado, B., Valladares, B., and Thomas, M. C. (2021). Variability of the Pr77-hallmark dual system among six

- T. cruzi strains belonging to different discrete typing units (DTUs). *Acta Tropica*, 222, 106053. <https://doi.org/10.1016/j.actatropica.2021.106053>
- Gómez, I., Rastrojo, A., Lorenzo-Díaz, F., Sánchez-Luque, F. J., Macías, F., Aguado, B., Valladares, B., Requena, J. M., López, M. C., and Thomas, M. C. (2020). Trypanosoma cruzi Ikiakarora (TcIII) draft genome sequence. *Microbiology Resource Announcements*, 9(27), e00453-20. <https://doi.org/10.1128/MRA.00453-20>
- Gómez, I., Rastrojo, A., Sánchez-Luque, F. J., Lorenzo-Díaz, F., Macías, F., Valladares, B., Aguado, B., Requena, J. M., López, M. C., and Thomas, M. C. (2020). Draft genome sequence of the Trypanosoma cruzi B. M. López strain (TcIa), isolated from a Colombian patient. *Microbiology Resource Announcements*, 9(18), e00031-20. <https://doi.org/10.1128/MRA.00031-20>
- Gómez, I., Thomas, M. C., Palacios, G., Egui, A., Carrilero, B., Simón, M., Valladares, B., Segovia, M., Carmelo, E., and López, M. C. (2021). Differential expression of immune response genes in asymptomatic chronic Chagas disease patients versus healthy subjects. *Frontiers in Cellular and Infection Microbiology*, 11, 722984. <https://doi.org/10.3389/FCIMB.2021.722984>
- Gorla, D., and Noireau, F. (2017). Geographic distribution of Triatominae vectors in America. In *American Trypanosomiasis Chagas Disease* (pp. 197–221). Elsevier. <https://doi.org/10.1016/B978-0-12-801029-7.00009-5>
- Gourbière, S., Dorn, P., Tripet, F., and Dumonteil, E. (2012). Genetics and evolution of triatomines: From phylogeny to vector control. *Heredity*, 108(3), 190–202. <https://doi.org/10.1038/hdy.2011.71>
- Guiñazú, N., Pellegrini, A., Carrera-Silva, E. A., Aoki, M. P., Cabanillas, A. M., Gironés, N., Fresno, M., Cano, R., and Gea, S. (2007). Immunisation with a major Trypanosoma cruzi antigen promotes pro-inflammatory cytokines, nitric oxide production and increases TLR2 expression. *International Journal for Parasitology*, 37(11), 1243–1254. <https://doi.org/10.1016/j.ijpara.2007.03.010>
- Hamuy, R., Vera, B. N., Ferreira, M. E., Acosta, N., and Lopez, E. (2013). Determination of the in vitro sensitivity of different Trypanosoma cruzi strains to benznidazole and the leaf extract of the plant Zanthoxylum chiloperone. *Memorias Del Instituto de Investigaciones En Ciencias de La Salud*, 9(2), 16–25.
- Heilman-Miller, S. L., and Woodson, S. A. (2003). Effect of transcription on folding of the Tetrahymena ribozyme. *RNA*, 9(6), 733. <https://doi.org/10.1261/RNA.5200903>
- Heras, S. R., López, M. C., García-Pérez, J. L., Martin, S. L., and Thomas, M. C. (2005). The L1Tc C-terminal domain from Trypanosoma cruzi non-long terminal repeat retrotransposon codes for a protein that bears two C2H2 zinc finger motifs and is endowed with nucleic acid chaperone activity. *Molecular and Cellular Biology*, 25(21), 9209–9220. <https://doi.org/10.1128/mcb.25.21.9209-9220.2005>
- Heras, S. R., López, M. C., Olivares, M., and Thomas, M. C. (2007). The L1Tc non-LTR retrotransposon of Trypanosoma cruzi contains an internal RNA-pol II-dependent promoter that strongly activates gene transcription and generates unspliced transcripts. *Nucleic Acids Research*, 35(7), 2199–2214. <https://doi.org/10.1093/nar/gkl1137>
- Heras, S. R., Thomas, M. C., García-Canadas, M., de Felipe, P., García-Pérez, J. L., Ryan, M. D., and López, M. C. (2006). L1Tc non-LTR retrotransposons from Trypanosoma cruzi contain a functional viral-like self-cleaving 2A sequence in frame with the active

- proteins they encode. *Cellular and Molecular Life Sciences*, 63(12), 1449–1460. <https://doi.org/10.1007/S00018-006-6038-2>
- Heras, S. R., Thomas, M. C., Macias, F., Patarroyo, M. E., Alonso, C., and López, M. C. (2009). Nucleic-acid-binding properties of the C2-L1Tc nucleic acid chaperone encoded by L1Tc retrotransposon. *Biochemical Journal*, 424(3), 479–490. <https://doi.org/10.1042/BJ20090766>
- Hernández, C., Cucunubá, Z., Flórez, C., Olivera, M., Valencia, C., Zambrano, P., León, C., and Ramírez, J. D. (2016). Molecular diagnosis of Chagas disease in Colombia: Parasitic loads and discrete typing units in patients from acute and chronic phases. *PLoS Neglected Tropical Diseases*, 10(9), e0004997. <https://doi.org/10.1371/journal.pntd.0004997>
- Hernández-Rivas, R., Martínez-Calvillo, S., Romero, M., and Hernández, R. (1992). Trypanosoma cruzi 5S rRNA genes: molecular cloning, structure and chromosomal organization. *FEMS Microbiology Letters*, 71(1), 63–67. <https://doi.org/10.1111/J.1574-6968.1992.TB05235.X>
- Herwaldt, B. L. (2001). Laboratory-acquired parasitic infections from accidental exposures. *Clinical Microbiology Reviews*, 14(4), 659–688. <https://doi.org/10.1128/CMR.14.3.659-688.2001>
- Higuera, S. L., Guhl, F., and Ramírez, J. D. (2013). Identification of Trypanosoma cruzi Discrete Typing Units (DTUs) through the implementation of a High-Resolution Melting (HRM) genotyping assay. *Parasites and Vectors*, 6(1), 1–6. <https://doi.org/10.1186/1756-3305-6-112>
- Hoff, K. J., and Stanke, M. (2013). WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, 41(Web Server issue), W123–8. <https://doi.org/10.1093/nar/gkt418>
- Hoft, D. F., and Eickhoff, C. S. (2005). Type 1 immunity provides both optimal mucosal and systemic protection against a mucosally invasive, intracellular pathogen. *Infection and Immunity*, 73(8), 4934–4940. <https://doi.org/10.1128/IAI.73.8.4934-4940.2005>
- Hunter, C. A., Ellis-Neyes, L. A., Slifer, T., Kanaly, S., Grünig, G., Fort, M., Rennick, D., and Araujo, F. G. (1997). IL-10 is required to prevent immune hyperactivity during infection with Trypanosoma cruzi. *Journal of Immunology*, 158(7), 3311–3316.
- Huprikar, S., Bosserman, E., Patel, G., Moore, A., Pinney, S., Anyanwu, A., Neofytos, D., Ketterer, D., Striker, R., Silveira, F., Qvarnstrom, Y., Steurer, F., Herwaldt, B., and Montgomery, S. (2013). Donor-derived Trypanosoma cruzi infection in solid organ recipients in the United States, 2001–2011. *American Journal of Transplantation*, 13(9), 2418–2425. <https://doi.org/10.1111/ajt.12340>
- Ishigame, H., Kakuta, S., Nagai, T., Kadoki, M., Nambu, A., Komiyama, Y., Fujikado, N., Tanahashi, Y., Akitsu, A., Kotaki, H., Sudo, K., Nakae, S., Sasakawa, C., and Iwakura, Y. (2009). Differential roles of interleukin-17A and -17F in host defense against mucosal bacterial infection and allergic responses. *Immunity*, 30(1), 108–119. <https://doi.org/10.1016/j.immuni.2008.11.009>
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., ... Myler, P. J. (2005). The genome of the kinetoplastid parasite, Leishmania major. *Science*, 309(5733), 436–442. <https://doi.org/10.1126/science.1112680>

- Jardim, A., Hanson, S., Ullman, B., McCubbin, W. D., Kay, C. M., and Olafson, R. W. (1995). Cloning and structure-function analysis of the *Leishmania donovani* kinetoplastid membrane protein-11. *Biochemical Journal*, 305(Pt 1), 315–320. <https://doi.org/10.1042/BJ3050315>
- Kajikawa, M., and Okada, N. (2002). LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell*, 111(3), 433–444. [https://doi.org/10.1016/S0092-8674\(02\)01041-3](https://doi.org/10.1016/S0092-8674(02)01041-3)
- Kapitonov, V. v, and Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics*, 9(5), 411–412. <https://doi.org/10.1038/nrg2165-c1>
- Kawashita, S. Y., da Silva, C. v., Mortara, R. A., Burleigh, B. A., and Briones, M. R. S. (2009). Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins. *Molecular and Biochemical Parasitology*, 165(1), 19–31. <https://doi.org/10.1016/j.molbiopara.2008.12.010>
- Keating, S. M., Deng, X., Fernandes, F., Cunha-Neto, E., Ribeiro, A. L., Adesina, B., Beyer, A. I., Contestable, P., Custer, B., Busch, M. P., and Sabino, E. C. (2015). Inflammatory and cardiac biomarkers are differentially expressed in clinical stages of Chagas disease. *International Journal of Cardiology*, 199, 451–459. <https://doi.org/10.1016/j.IJCARD.2015.07.040>
- Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 115(1), 49–63. <https://doi.org/10.1023/A:1016072014259>
- Kierszenbaum, F., Gottlieb, C. A., and Budzko, D. B. (1983). Exacerbation of *Trypanosoma cruzi* infection in mice treated with the immunoregulatory agent cyclosporin A. *Tropenmedizin Und Parasitologie*, 34(1), 4–6.
- Kierszenbaum, F., and Lima, M. F. (1983). Susceptibility of insect-borne, metacyclic forms of *Trypanosoma cruzi* to antibody-mediated mechanisms of destruction. *The American Journal of Tropical Medicine and Hygiene*, 32(6), 1236–1241. <https://doi.org/10.4269/AJTMH.1983.32.1236>
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kinoshita-Yanaga, A. T., Toledo, M. J. D. O., de Araújo, S. M., Vier, B. P., and Gomes, M. L. (2009). Accidental infection by *Trypanosoma cruzi* follow-up by the polymerase chain reaction: Case report. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, 51(5), 295–298. <https://doi.org/10.1590/S0036-46652009000500011>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5), 722–736. <https://doi.org/10.1101/GR.215087.116>
- Krautz, G. M., Galvao, L. M. C., Cancado, J. R., Guevara-Espinoza, A., Ouaiissi, A., and Krettli, A. U. (1995). Use of a 24-kilodalton *Trypanosoma cruzi* recombinant protein to monitor cure of human Chagas' disease. *Journal of Clinical Microbiology*, 33(8), 2086–2090. <https://doi.org/10.1128/jcm.33.8.2086-2090.1995>
- Krettli, A. U., and Brener, Z. (1976). Protective effects of specific antibodies in *Trypanosoma cruzi* infections. *Journal of Immunology*, 116(3), 755–760.

- Kulkarni, M. M., Olson, C. L., Engman, D. M., and McGwire, B. S. (2009). Trypanosoma cruzi GP63 proteins undergo stage-specific differential posttranslational modification and are important for host cell infection. *Infection and Immunity*, 77(5), 2193–2200. <https://doi.org/10.1128/IAI.01542-08>
- Kulkarni, M. M., Varikuti, S., Terrazas, C., Kimble, J. L., Satoskar, A. R., and McGwire, B. S. (2015). Signal transducer and activator of transcription 1 (STAT-1) plays a critical role in control of Trypanosoma cruzi infection. *Immunology*, 145(2), 225–231. <https://doi.org/10.1111/imm.12438>
- Kumar, S., and Tarleton, R. L. (1998). The relative contribution of antibody production and CD8+ T cell function to immune control of Trypanosoma cruzi. *Parasite Immunology*, 20(5), 207–216. <https://doi.org/10.1046/j.1365-3024.1998.00154.x>
- Kumar, S., and Tarleton, R. L. (2001). Antigen-specific Th1 but not Th2 cells provide protection from lethal Trypanosoma cruzi infection in mice. *The Journal of Immunology*, 166(7), 4596–4603. <https://doi.org/10.4049/jimmunol.166.7.4596>
- Kun, H., Moore, A., Mascola, L., Steurer, F., Lawrence, G., Kubak, B., Radhakrishna, S., Leiby, D., Herron, R., Mone, T., Hunter, R., and Kuehnert, M. (2009). Transmission of Trypanosoma cruzi by heart transplantation. *Clinical Infectious Diseases*, 48(11), 1534–1540. <https://doi.org/10.1086/598931>
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. v., Caldwell, B., ... Danchin, A. (1997). The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature*, 390(6657), 249–256. <https://doi.org/10.1038/36786>
- Kuo, M. Y., Sharmeen, L., Dinter-Gottlieb, G., and Taylor, J. (1988). Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *Journal of Virology*, 62(12), 4439–4444. <https://doi.org/10.1128/JVI.62.12.4439-4444.1988>
- Kuschnir, E., Sgammini, H., Castro, R., Evequoz, C., Ledesma, R., and Brunetto, J. (1985). Valoración de la función cardíaca por angiografía radioisotópica, en pacientes con cardiopatía chagásica crónica. *Arquivos Brasileiros de Cardiologia*, 45(4), 249–256.
- la Fuente Zeraín, C. A. (2006). Inmunología de la infección por T. cruzi y de la enfermedad de Chagas. *Enfermedades Emergentes*, 8(Suppl 1), 17–19.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., ... Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. <https://doi.org/10.1038/35057062>
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lasso, P., Mateus, J., Pavía, P., Rosas, F., Roa, N., Thomas, M. C., López, M. C., González, J. M., Puerta, C. J., and Cuéllar, A. (2015). Inhibitory receptor expression on CD8 + T cells is linked to functional responses against Trypanosoma cruzi antigens in chronic chagasic patients. *The Journal of Immunology*, 195(8), 3748–3758. <https://doi.org/10.4049/jimmunol.1500459>

- Lasso, P., Mesa, D., Cuéllar, A., Guzmán, F., Bolaños, N., Rosas, F., Velasco, V., Thomas, M. C., López, M. C., González, J. M., and Puerta, C. J. (2010). Frequency of specific CD8+ T cells for a promiscuous epitope derived from *Trypanosoma cruzi* KMP-11 protein in chagasic patients. *Parasite Immunology*, 32(7), 494–502. <https://doi.org/10.1111/J.1365-3024.2010.01206.X>
- Laucella, S. A., de Titto, E. H., and Segura, E. L. (1996). Epitopes common to *Trypanosoma cruzi* and mammalian tissues are recognized by sera from Chagas' disease patients: prognosis value in Chagas disease. *Acta Tropica*, 62(3), 151–162. [https://doi.org/10.1016/S0001-706X\(96\)00032-0](https://doi.org/10.1016/S0001-706X(96)00032-0)
- Laucella, S. A., Postan, M., Martin, D., Fralish, B. H., Albareda, M. C., Alvarez, M. G., Lococo, B., Barbieri, G., Viotti, R. J., and Tarleton, R. L. (2004). Frequency of interferon- $\gamma$ -producing T cells specific for *Trypanosoma cruzi* inversely correlates with disease severity in chronic human Chagas disease. *Journal of Infectious Diseases*, 189(5), 909–918. <https://doi.org/10.1086/381682>
- Lewis, M. D., Llewellyn, M. S., Gaunt, M. W., Yeo, M., Carrasco, H. J., and Miles, M. A. (2009). Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *International Journal for Parasitology*, 39(12), 1305–1317. <https://doi.org/10.1016/j.ijpara.2009.04.001>
- Li, L., Stoeckert Jr., C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9), 2178–2189. <https://doi.org/10.1101/gr.1224503>
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, 1(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics*, 27(12), 1739–1740. <https://doi.org/10.1093/bioinformatics/btr260>
- Lima, L., Espinosa-Álvarez, O., Ortiz, P. A., Trejo-Varón, J. A., Carranza, J. C., Pinto, C. M., Serrano, M. G., Buck, G. A., Camargo, E. P., and Teixeira, M. M. G. (2015). Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). *Acta Tropica*, 151(1), 166–177. <https://doi.org/10.1016/j.actatropica.2015.07.015>
- Lin, L., Ibrahim, A. S., Xu, X., Farber, J. M., Avanesian, V., Baquir, B., Fu, Y., French, S. W., Edwards, J. E., and Spellberg, B. (2009). Th1-Th17 cells mediate protective adaptive immunity against *Staphylococcus aureus* and *Candida albicans* infection in mice. *PLoS Pathogens*, 5(12), e1000703. <https://doi.org/10.1371/journal.ppat.1000703>
- Lisboa, C. V., Pinho, A. P., Monteiro, R. V., and Jansen, A. M. (2007). *Trypanosoma cruzi* (kinetoplastida Trypanosomatidae): Biological heterogeneity in the isolates derived from wild hosts. *Experimental Parasitology*, 116(2), 150–155. <https://doi.org/10.1016/j.exppara.2006.12.005>
- Lorena, V. M. B., Lorena, I. M. B., Braz, S. C. M., Melo, A. S., Melo, M. F. A. D., Melo, M. G. A. C., Silva, E. D., Ferreira, A. G. P., Morais, C. N. L., Costa, V. M. A., Correa-Oliveira, R., and Gomes, Y. M. (2010). Cytokine levels in serious cardiopathy of Chagas disease after in vitro stimulation with recombinant antigens from *Trypanosoma cruzi*. *Scandinavian Journal of Immunology*, 72(6), 529–539. <https://doi.org/10.1111/J.1365-3083.2010.02462.X>

- Lorenzi, H. A., Robledo, G., and Levin, M. J. (2006). The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. *Molecular and Biochemical Parasitology*, 145(2), 184–194. <https://doi.org/10.1016/j.molbiopara.2005.10.002>
- Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*, 72(4), 595–605. [https://doi.org/10.1016/0092-8674\(93\)90078-5](https://doi.org/10.1016/0092-8674(93)90078-5)
- Luhrs, K. A., Fouts, D. L., and Manning, J. E. (2003). Immunization with recombinant paraflagellar rod protein induces protective immunity against *Trypanosoma cruzi* infection. *Vaccine*, 21(21–22), 3058–3069. [https://doi.org/10.1016/S0264-410X\(03\)00108-7](https://doi.org/10.1016/S0264-410X(03)00108-7)
- Macedo, A. M., Machado, C. R., Oliveira, R. P., and Pena, S. D. J. (2004). *Trypanosoma cruzi*: Genetic structure of populations and relevance of genetic variability to the pathogenesis of Chagas disease. *Memórias Do Instituto Oswaldo Cruz*, 99(1), 1–12. <https://doi.org/10.1590/S0074-02762004000100001>
- Macedo, A. M., and Pena, S. D. J. (1998). Genetic variability of *Trypanosoma cruzi*: Implications for the pathogenesis of Chagas disease. *Parasitology Today*, 14(3), 119–124. [https://doi.org/10.1016/S0169-4758\(97\)01179-4](https://doi.org/10.1016/S0169-4758(97)01179-4)
- Macías, F., Afonso-Lehmann, R., Carreira, P. E., and Thomas, M. C. (2021). TBP and SNAP50 transcription factors bind specifically to the Pr77 promoter sequence from trypanosomatid non-LTR retrotransposons. *Parasites & Vectors*, 14(1), 313. <https://doi.org/10.1186/S13071-021-04803-5>
- Macías, F., Afonso-Lehmann, R., López, M. C., Gómez, I., and Thomas, M. C. (2018). Biology of *Trypanosoma cruzi* retrotransposons: From an enzymatic to a structural point of view. *Current Genomics*, 19(2), 110–118. <https://doi.org/10.2174/1389202918666170815150738>
- Macías, F., López, M. C., and Thomas, M. C. (2016). The Trypanosomatid Pr77-hallmark contains a downstream core promoter element essential for transcription activity of the *Trypanosoma cruzi* L1Tc retrotransposon. *BMC Genomics*, 17(1), 105. <https://doi.org/10.1186/s12864-016-2427-6>
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), W636–W641. <https://doi.org/10.1093/NAR/GKZ268>
- Magalhães, L. M. D., Villani, F. N. A., Nunes, M. do C. P., Gollob, K. J., Rocha, M. O. C., and Dutra, W. O. (2013). High interleukin 17 expression is correlated with better cardiac function in human Chagas disease. *The Journal of Infectious Diseases*, 207(4), 661–665. <https://doi.org/10.1093/INFDIS/JIS724>
- Malik, H. S., Burke, W. D., and Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Molecular Biology and Evolution*, 16(6), 793–805.
- Manne-Goehler, J., Umeh, C. A., Montgomery, S. P., and Wirtz, V. J. (2016). Estimating the burden of Chagas disease in the United States. *PLoS Neglected Tropical Diseases*, 10(11), e0005033. <https://doi.org/10.1371/journal.pntd.0005033>
- Marañón, C., Egui, A., Carrilero, B., Thomas, M. C., Pinazo, M. J., Gascón, J., Segovia, M., and López, M. C. (2011). Identification of HLA-A\*02:01-restricted CTL epitopes in

- Trypanosoma cruzi heat shock protein-70 recognized by Chagas disease patients. *Microbes and Infection*, 13(12-13), 1025-1032. <https://doi.org/10.1016/j.micinf.2011.05.010>
- Marañón, C., Thomas, M. C., Planelles, L., and López, M. C. (2001). The immunization of A2/K(b) transgenic mice with the KMP11-HSP70 fusion protein induces CTL response against human cells expressing the T. cruzi KMP11 antigen: identification of A2-restricted epitopes. *Molecular Immunology*, 38(4), 279-287. [https://doi.org/10.1016/S0161-5890\(01\)00059-1](https://doi.org/10.1016/S0161-5890(01)00059-1)
- Marcili, A., Lima, L., Cavazzana, M., Junqueira, A. C. V., Veludo, H. H., Maia Da Silva, F., Campaner, M., Paiva, F., Nunes, V. L. B., and Teixeira, M. M. G. (2009). A new genotype of Trypanosoma cruzi associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitology*, 136(6), 641-655. <https://doi.org/10.1017/S0031182009005861>
- Martin, D., and Tarleton, R. (2005). Antigen-specific T cells maintain an effector memory phenotype during persistent Trypanosoma cruzi infection. *Journal of Immunology*, 174(3), 1594-1601. <https://doi.org/10.4049/JIMMUNOL.174.3.1594>
- Martin, F., Marañón, C., Olivares, M., Alonso, C., and López, M. C. (1995). Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from Trypanosoma cruzi: Homology of the first ORF with the Ape family of DNA repair enzymes. *Journal of Molecular Biology*, 247(1), 49-59. <https://doi.org/10.1006/jmbi.1994.0121>
- Martín, F., Puertas, C., Thomas, M. C., Marañón, C., Patarroyo, M. E., Martín, J., Alonso, C., and López, M. C. (1997). Identification of a Trypanosoma cruzi antigenic epitope implicated in the infectivity of fibroblast LLC-MK2 cells. *Parasitology Research*, 83(3), 226-232. <https://doi.org/10.1007/S004360050238>
- Martínez-Calvillo, S., Nguyen, D., Stuart, K., and Myler, P. J. (2004). Transcription initiation and termination on Leishmania major chromosome 3. *Eukaryotic Cell*, 3(2), 506-517. <https://doi.org/10.1128/EC.3.2.506-517.2004>
- Mateus, J., Lasso, P., Pavia, P., Rosas, F., Roa, N., Valencia-Hernández, C. A., González, J. M., Puerta, C. J., and Cuéllar, A. (2015). Low frequency of circulating CD8+ T stem cell memory cells in chronic chagasic patients with severe forms of the disease. *PLoS Neglected Tropical Diseases*, 9(1), e3432. <https://doi.org/10.1371/journal.pntd.0003432>
- Mathiopoulos, K. D., della Torre, A., Predazzi, V., Petrarca, V., and Coluzzi, M. (1998). Cloning of inversion breakpoints in the Anopheles gambiae complex traces a transposable element at the inversion junction. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), 12444-12449. <https://doi.org/10.1073/pnas.95.21.12444>
- Maya, J. D., Cassels, B. K., Iturriaga-Vásquez, P., Ferreira, J., Faúndez, M., Galanti, N., Ferreira, A., and Morello, A. (2007). Mode of action of natural and synthetic drugs against Trypanosoma cruzi and their interaction with the mammalian host. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 146(4), 601-620. <https://doi.org/10.1016/j.cbpa.2006.03.004>
- Mayer, M. P., and Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62(6), 670-684. <https://doi.org/10.1007/S00018-004-4464-6>



- Meymandi, S., Hernandez, S., Park, S., Sanchez, D. R., and Forsyth, C. (2018). Treatment of Chagas disease in the United States. *Current Treatment Options in Infectious Diseases*, 10(3), 373–388. <https://doi.org/10.1007/s40506-018-0170-z>
- Michels, P. A. M., and Hannaert, V. (1994). The evolution of kinetoplastid glycosomes. *Journal of Bioenergetics and Biomembranes*, 26(2), 213–219. <https://doi.org/10.1007/BF00763070>
- Miles, M. A., Póvoa, M. M., Prata, A., Cedillos, R. A., de Souza, A. A., and Macedo, V. (1981). Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? *The Lancet*, 317(8234), 1338–1340. [https://doi.org/10.1016/S0140-6736\(81\)92518-6](https://doi.org/10.1016/S0140-6736(81)92518-6)
- Miller, M. J., Wrightsman, R. A., Stryker, G. A., and Manning, J. E. (1997). Protection of mice against *Trypanosoma cruzi* by immunization with paraflagellar rod proteins requires T cell, but not B cell, function. *The Journal of Immunology*, 158(11), 5330–5337.
- Milner, J. D., Sandler, N. G., and Douek, D. C. (2010). Th17 cells, Job's syndrome and HIV: Opportunities for bacterial and fungal infections. *Current Opinion in HIV and AIDS*, 5(2), 179–183. <https://doi.org/10.1097/COH.0b013e328335ed3e>
- Minoprio, P., Eisen, H., Joskowicz, M., Pereira, P., and Coutinho, A. (1987). Suppression of polyclonal antibody production in *Trypanosoma cruzi*-infected mice by treatment with anti-L3T4 antibodies. *The Journal of Immunology*, 139(2), 545–550.
- Miyazaki, Y., Hamano, S., Wang, S., Shimano, Y., Iwakura, Y., and Yoshida, H. (2010). IL-17 is necessary for host protection against acute-phase *Trypanosoma cruzi* infection. *The Journal of Immunology*, 185(2), 1150–1157. <https://doi.org/10.4049/jimmunol.0900047>
- Moncayo, A. (2003). Chagas Disease: Current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. *Memorias Do Instituto Oswaldo Cruz*, 98(5), 577–591. <https://doi.org/10.1590/S0074-02762003000500001>
- Monteiro, W. M., Magalhães, L. K. C., de Sá, A. R. N., Gomes, M. L., Toledo, M. J. de O., Borges, L., Pires, I., Guerra, J. A. de O., Silveira, H., and Barbosa, M. das G. V. (2012). *Trypanosoma cruzi* IV causing outbreaks of acute Chagas disease and infections by different haplotypes in the Western Brazilian Amazonia. *PLoS One*, 7(7), e41284. <https://doi.org/10.1371/journal.pone.0041284>
- Monteiro, W. M., Magalhães, L. K., Santana Filho, F. S., Borborema, M., Silveira, H., and Barbosa, M. D. G. v. (2010). *Trypanosoma cruzi* TcIII/Z3 genotype as agent of an outbreak of Chagas disease in the Brazilian Western Amazonia: Short communication. *Tropical Medicine and International Health*, 15(9), 1049–1051. <https://doi.org/10.1111/j.1365-3156.2010.02577.x>
- Montiel, G., and Díaz, G. (2002). Respuesta inmune de las células del hospedero a la infección por *Trypanosoma cruzi*. *Revista Médica Del Hospital Nacional de Niños Dr. Carlos Sáenz Herrera*, 37(1–2), 57–63.
- Moore, K. W., de Waal Malefyt, R., Coffman, R. L., and O'Garra, A. (2001). Interleukin-10 and the interleukin-10 receptor. *Annual Review of Immunology*, 19, 683–765. <https://doi.org/10.1146/ANNUREV.IMMUNOL.19.1.683>
- Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., ...

- Groop, L. C. (2003). PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, 34(3), 267–273. <https://doi.org/10.1038/ng1180>
- Morell, M., Thomas, M. C., Caballero, T., Alonso, C., and López, M. C. (2006). The genetic immunization with paraflagellar rod protein-2 fused to the HSP70 confers protection against late *Trypanosoma cruzi* infection. *Vaccine*, 24(49–50), 7046–7055. <https://doi.org/10.1016/J.VACCINE.2006.07.006>
- Moreno-Campos, R., Florencio-Martínez, L. E., Nepomuceno-Mejía, T., Rojas-Sánchez, S., Vélez-Ramírez, D. E., Padilla-Mejía, N. E., Figueroa-Angulo, E., Manning-Cela, R., and Martínez-Calvillo, S. (2016). Molecular characterization of 5S ribosomal RNA genes and transcripts in the protozoan parasite *Leishmania major*. *Parasitology*, 143(14), 1917–1929. <https://doi.org/10.1017/S0031182016001712>
- Moretti, E., Basso, B., Castro, I., Carrizo Paez, M., Chaul, M., Barbieri, G., Canal Feijoo, D., Sartori, M. J., and Carrizo Paez, R. (2005). Chagas' disease: Study of congenital transmission in cases of acute maternal infection. *Revista Da Sociedade Brasileira de Medicina Tropical*, 38(1), 53–55. <https://doi.org/10.1590/S0037-86822005000100010>
- Moretti, E., Cervetta, L., Basso, B., Castro, I., and Santamarina, N. (1998). Enfermedad de Chagas crónica: efectos del tratamiento en los niveles de anticuerpos hacia antígenos crudos y semipurificados del *Trypanosoma cruzi*. *Boletín Chileno de Parasitología*, 53(1–2), 3–9.
- Morillo, C. A., Marin-Neto, J. A., Avezum, A., Sosa-Estani, S., Rassi, A., Rosas, F., Villena, E., Quiroz, R., Bonilla, R., Britto, C., Guhl, F., Velazquez, E., Bonilla, L., Meeks, B., Rao-Melacini, P., Pogue, J., Mattos, A., Lazdins, J., Rassi, A., ... Yusuf, S. (2015). Randomized trial of benznidazole for chronic Chagas' cardiomyopathy. *New England Journal of Medicine*, 373(14), 1295–1306. <https://doi.org/10.1056/nejmoa1507574>
- Morris, S. A., Tanowitz, H. B., Wittner, M., and Bilezikian, J. P. (1990). Pathophysiological insights into the cardiomyopathy of Chagas' disease. *Circulation*, 82(6), 1900–1909. <https://doi.org/10.1161/01.CIR.82.6.1900>
- Mosca, W., and Briceño, L. (2000). Etiopatología de la enfermedad de Chagas: una nueva perspectiva. *Archivos Del Hospital Vargas*, 42(1), 11–17.
- Muñoz, J., Coll, O., Juncosa, T., Vergés, M., Pino, M. del, Fumado, V., Bosch, J., Posada, E. J., Hernandez, S., Fisa, R., Bogaña, J. M., Gállego, M., Sanz, S., Portús, M., and Gascón, J. (2009). Prevalence and vertical transmission of *Trypanosoma cruzi* infection among pregnant Latin American women attending 2 maternity clinics in Barcelona, Spain. *Clinical Infectious Diseases*, 48(12), 1736–1740. <https://doi.org/10.1086/599223>
- Murcia, L., Carrilero, B., Munoz-Davila, M. J., Thomas, M. C., López, M. C., and Segovia, M. (2013). Risk factors and primary prevention of congenital Chagas disease in a nonendemic country. *Clinical Infectious Diseases*, 56(4), 496–502. <https://doi.org/10.1093/cid/cis910>
- Murcia, L., Carrilero, B., Viñas, P. A., and Segovia, M. (2012). Nifurtimox chemotherapy: Collateral effects in treated *Trypanosoma cruzi* infected patients. *Revista Española de Quimioterapia*, 25(1), 74–75.
- Murcia, L., Simón, M., Carrilero, B., Roig, M., and Segovia, M. (2017). Treatment of infected women of childbearing age prevents congenital *Trypanosoma cruzi* infection by eliminating the parasitemia detected by PCR. *The Journal of Infectious Diseases*, 215(9), 1452–1458. <https://doi.org/10.1093/INFDIS/JIX087>

- Murta, S. M. F., Ropert, C., Alves, R. O., Gazzinelli, R. T., and Romanha, A. J. (1999). In-vivo treatment with benznidazole enhances phagocytosis, parasite destruction and cytokine release by macrophages during infection with a drug-susceptible but not with a derived drug-resistant *Trypanosoma cruzi* population. *Parasite Immunology*, *21*(10), 535–544. <https://doi.org/10.1046/J.1365-3024.1999.00251.X>
- Nakano, S., Cerrone, A. L., and Bevilacqua, P. C. (2003). Mechanistic characterization of the HDV genomic ribozyme: classifying the catalytic and structural metal ion sites within a multichannel reaction mechanism. *Biochemistry*, *42*(10), 2982–2994. <https://doi.org/10.1021/BI026815X>
- Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. (2000). General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, *287*(5457), 1493–1497. <https://doi.org/10.1126/SCIENCE.287.5457.1493>
- Nakano, S., Proctor, D. J., and Bevilacqua, P. C. (2001). Mechanistic characterization of the HDV genomic ribozyme: Assessing the catalytic and structural contributions of divalent metal ions within a multichannel reaction mechanism. *Biochemistry*, *40*(40), 12022–12038. <https://doi.org/10.1021/BI011253N>
- Negrette, O. S., Valdéz, F. J. S., Lacunza, C. D., Bustos, M. F. G., Mora, M. C., Uncos, A. D., and Basombrío, M. Á. (2008). Serological evaluation of specific-antibody levels in patients treated for chronic Chagas' disease. *Clinical and Vaccine Immunology*, *15*(2), 297–302. <https://doi.org/10.1128/CVI.00106-07>
- Niang, M., Yan Yam, X., and Preiser, P. R. (2011). Correction: The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathogens*, *7*(9), e1000307. <https://doi.org/10.1371/annotation/6b4264ca-dd6a-4255-9f91-c0fb6692cc26>
- Nikolayeva, O., and Robinson, M. D. (2014). edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods in Molecular Biology*, *1150*, 45–79. [https://doi.org/10.1007/978-1-4939-0512-6\\_3](https://doi.org/10.1007/978-1-4939-0512-6_3)
- Novy, P., Quigley, M., Huang, X., and Yang, Y. (2007). CD4 T cells are required for CD8 T cell survival during both primary and memory recall responses. *The Journal of Immunology*, *179*(12), 8243–8251. <https://doi.org/10.4049/JIMMUNOL.179.12.8243>
- Olivares, M., Alonso, C., and López, M. C. (1997). The open reading frame 1 of the L1Tc retrotransposon of *Trypanosoma cruzi* codes for a protein with apurinic-apyrimidinic nuclease activity. *Journal of Biological Chemistry*, *272*(40), 25224–25228. <https://doi.org/10.1074/jbc.272.40.25224>
- Olivares, M., García-Pérez, J. L., Thomas, M. C., Heras, S. R., and López, M. C. (2002). The non-LTR (Long Terminal Repeat) retrotransposon L1Tc from *Trypanosoma cruzi* codes for a protein with RNase H activity. *Journal of Biological Chemistry*, *277*(31), 28025–28030. <https://doi.org/10.1074/jbc.M202896200>
- Olivares, M., López, M. C., García-Pérez, J. L., Briones, P., Pulgar, M., and Thomas, M. C. (2003). The endonuclease NL1Tc encoded by the LINE L1Tc from *Trypanosoma cruzi* protects parasites from daunorubicin DNA damage. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1626*(1–3), 25–32. [https://doi.org/10.1016/S0167-4781\(03\)00022-8](https://doi.org/10.1016/S0167-4781(03)00022-8)
- Olivares, M., Thomas, M. C., Alonso, C., and López, M. C. (1999). The L1Tc, long interspersed nucleotide element from *Trypanosoma cruzi*, encodes a protein with 3'-phosphatase and 3'-phosphodiesterase enzymatic activities. *Journal of Biological Chemistry*, *274*(34), 23883–23886. <https://doi.org/10.1074/jbc.274.34.23883>

- Olivares, M., Thomas, M. C., López-Barajas, A., Requena, J. M., García-Pérez, J. L., Angel, S., Alonso, C., and López, M. C. (2000). Genomic clustering of the Trypanosoma cruzi nonlong terminal L1Tc retrotransposon with defined interspersed repeated DNA elements. *Electrophoresis*, 21(14), 2973–2982. [https://doi.org/10.1002/1522-2683\(20000801\)21:14<2973::AID-ELPS2973>3.0.CO;2-4](https://doi.org/10.1002/1522-2683(20000801)21:14<2973::AID-ELPS2973>3.0.CO;2-4)
- Padilla, A. M., Bustamante, J. M., and Tarleton, R. L. (2009). CD8+ T cells in Trypanosoma cruzi infection. *Current Opinion in Immunology*, 21(4), 385–390. <https://doi.org/10.1016/j.COI.2009.07.006>
- Pan American Health Organization (PAHO), and World Health Organization (WHO). (2021). *Enfermedad de Chagas en las Américas - Hoja informativa para los trabajadores de salud*. [https://www3.paho.org/hq/index.php?option=com\\_content&view=article&id=13568:chagas-in-americas-health-workers-2017&Itemid=40721&lang=es](https://www3.paho.org/hq/index.php?option=com_content&view=article&id=13568:chagas-in-americas-health-workers-2017&Itemid=40721&lang=es)
- Pasion, S. G., Brown, G. W., Brown, L. M., and Ray, D. S. (1994). Periodic expression of nuclear and mitochondrial DNA replication genes during the trypanosomatid cell cycle. *Journal of Cell Science*, 107(12), 3515–3520. <https://doi.org/10.1242/jcs.107.12.3515>
- Pereira, K. S., Schmidt, F. L., Guaraldo, A. M. A., Franco, R. M. B., Dias, V. L., and Passos, L. A. C. (2009). Chagas' disease as a foodborne illness. *Journal of Food Protection*, 72(2), 441–446. <https://doi.org/10.4315/0362-028X-72.2.441>
- Pérez Yanez, L. M., Gutiérrez López, A., Rodríguez Blanco, S., and Gil Sarduy, A. (2017). Enfermedad de Chagas. Amenaza en sombras para los corazones de la América Latina. *Revista Cubana de Medicina*, 56(1), 50–68.
- Pérez-Antón, E., Thomas, M. C., Egui, A., and López, M. C. (2019). T-cell exhaustion process during chronic infection caused by intracellular trypanosomatids. *Ars Pharmaceutica*, 60(2), 65–78. <https://doi.org/10.30827/ARS.V60I2.9432>
- Pérez-Molina, J. A., and Molina, I. (2018). Chagas disease. *The Lancet*, 391(10115), 82–94. [https://doi.org/10.1016/S0140-6736\(17\)31612-4](https://doi.org/10.1016/S0140-6736(17)31612-4)
- Petray, P. B., Rottenberg, M. E., Bertot, G., Corral, R. S., Diaz, A., Örn, A., and Grinstein, S. (1993). Effect of anti- $\gamma$ -interferon and anti-interleukin-4 administration on the resistance of mice against infection with reticulotropic and myotropic strains of Trypanosoma cruzi. *Immunology Letters*, 35(1), 77–80. [https://doi.org/10.1016/0165-2478\(93\)90151-Q](https://doi.org/10.1016/0165-2478(93)90151-Q)
- Pfaffl, M. W., Tichopad, A., Prgomet, C., and Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters*, 26(6), 509–515. <https://doi.org/10.1023/B:BILE.0000019559.84305.47>
- Pinazo, M. J., Espinosa, G., Cortes-Lletget, C., de Posada, E. J., Aldasoro, E., Oliveira, I., Muñoz, J., Gállego, M., and Gascon, J. (2013). Immunosuppression and Chagas disease: A management challenge. *PLoS Neglected Tropical Diseases*, 7(1), e1965. <https://doi.org/10.1371/journal.pntd.0001965>
- Pinazo, M. J., Tassies, D., Muñoz, J., Fisa, R., Posada, E., Monteagudo, J., Ayala, E., Gállego, M., Reverter, J.-C., and Gascon, J. (2011). Hypercoagulability biomarkers in Trypanosoma cruzi-infected patients. *Thrombosis and Haemostasis*, 106(10), 617–623. <https://doi.org/10.1160/TH11-04-0251>
- Pinazo, M. J., Thomas, M. C., Bua, J., Perrone, A., Schijman, A. G., Viotti, R. J., Ramsey, J. M., Ribeiro, I., Sosa-Estani, S., López, M. C., and Gascon, J. (2014). Biological markers for evaluating therapeutic efficacy in Chagas disease, a systematic review. *Expert Review*

- of *Anti-Infective Therapy*, 12(4), 479–496.  
<https://doi.org/10.1586/14787210.2014.899150>
- Pinazo, M. J., Thomas, M. C., Bustamante, J., de Almeida, I. C., López, M. C., and Gascon, J. (2015). Biomarkers of therapeutic responses in chronic Chagas disease: State of the art and future perspectives. *Memorias Do Instituto Oswaldo Cruz*, 110(3), 422–432. <https://doi.org/10.1590/0074-02760140435>
- Pinto Dias, J. C. (1989). The indeterminate form of human chronic Chagas' disease. A clinical epidemiological review. *Revista Da Sociedade Brasileira de Medicina Tropical*, 22(3), 147–156. <https://doi.org/10.1590/s0037-86821989000300007>
- Planelles, L., Thomas, M. C., Alonso, C., and López, M. C. (2001). DNA immunization with *Trypanosoma cruzi* HSP70 fused to the KMP11 protein elicits a cytotoxic and humoral immune response against the antigen and leads to protection. *Infection and Immunity*, 69(10), 6558–6563. <https://doi.org/10.1128/IAI.69.10.6558-6563.2001>
- Planelles, L., Thomas, M., Pulgar, M., Marañón, C., Grabbe, S., and López, M. C. (2002). *Trypanosoma cruzi* heat-shock protein-70 kDa, alone or fused to the parasite KMP11 antigen, induces functional maturation of murine dendritic cells. *Immunology and Cell Biology*, 80(3), 241–247. <https://doi.org/10.1046/J.1440-1711.2002.01081.X>
- Prata, A. (2001). Clinical and epidemiological aspects of Chagas disease. *Lancet Infectious Diseases*, 1(2), 92–100. [https://doi.org/10.1016/S1473-3099\(01\)00065-2](https://doi.org/10.1016/S1473-3099(01)00065-2)
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ramírez, J. C., Cura, C. I., da Cruz Moreira, O., Lages-Silva, E., Juiz, N., Velázquez, E., Ramírez, J. D., Alberti, A., Pavia, P., Flores-Chávez, M. D., Muñoz-Calderón, A., Pérez-Morales, D., Santalla, J., Marcos Da Matta Guedes, P., Peneau, J., Marcet, P., Padilla, C., Cruz-Robles, D., Valencia, E., ... Schijman, A. G. (2015). Analytical validation of quantitative real-time PCR methods for quantification of *trypanosoma cruzi* DNA in blood samples from Chagas disease patients. *Journal of Molecular Diagnostics*, 17(5), 605–615. <https://doi.org/10.1016/j.jmoldx.2015.04.010>
- Ramírez, J. D., Guhl, F., Rendón, L. M., Rosas, F., Marin-Neto, J. A., and Morillo, C. A. (2010). Chagas cardiomyopathy manifestations and *Trypanosoma cruzi* genotypes circulating in chronic chagasic patients. *PLoS Neglected Tropical Diseases*, 4(11), e899. <https://doi.org/10.1371/JOURNAL.PNTD.0000899>
- Ramírez, J. R., Berberich, C., Jaramillo, A., Alonso, C., and Vélez, I. D. (1998). Molecular and antigenic characterization of the *Leishmania* (Viannia) *panamensis* kinetoplastid membrane protein-11. *Memórias Do Instituto Oswaldo Cruz*, 93(2), 247–254. <https://doi.org/10.1590/S0074-02761998000200022>
- Rassi, A., Rassi, A., and Marcondes de Rezende, J. (2012). American Trypanosomiasis (Chagas Disease). *Infectious Disease Clinics of North America*, 26(2), 275–291. <https://doi.org/10.1016/j.idc.2012.03.002>
- Rassi, A., Rassi, A., and Marin-Neto, J. A. (2010). Chagas disease. *The Lancet*, 375(9723), 1388–1402. [https://doi.org/10.1016/S0140-6736\(10\)60061-X](https://doi.org/10.1016/S0140-6736(10)60061-X)
- Reeb, P. D., Bramardi, S. J., and Steibel, J. P. (2015). Assessing dissimilarity measures for sample-based hierarchical clustering of RNA sequencing data using plasmode datasets. *PLoS One*, 10(7), e0132310. <https://doi.org/10.1371/JOURNAL.PONE.0132310>

- Reis-Cunha, J. L., and Bartholomeu, D. C. (2019). Trypanosoma cruzi genome assemblies: Challenges and milestones of assembling a highly repetitive and complex genome. *Methods in Molecular Biology*, 1955(5), 1–22. [https://doi.org/10.1007/978-1-4939-9148-8\\_1](https://doi.org/10.1007/978-1-4939-9148-8_1)
- Reis-Cunha, J. L., Rodrigues-Luiz, G. F., Valdivia, H. O., Baptista, R. P., Mendes, T. A. O., de Moraes, G. L., Guedes, R., Macedo, A. M., Bern, C., Gilman, R. H., Lopez, C. T., Andersson, B., Vasconcelos, A. T., and Bartholomeu, D. C. (2015). Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct Trypanosoma cruzi strains. *BMC Genomics*, 16(1), 499. <https://doi.org/10.1186/s12864-015-1680-4>
- Repsilber, D., Wiese, S., Rachen, M., Schröder, A. W., Riesner, D., and Steger, G. (1999). Formation of metastable RNA structures by sequential folding during transcription: time-resolved structural analysis of potato spindle tuber viroid (-)-stranded RNA by temperature-gradient gel electrophoresis. *RNA*, 5(4), 574–584. <https://doi.org/10.1017/S1355838299982018>
- Requena, J. M., Jimenez-Ruiz, A., Soto, M., Assiego, R., Santarén, J. F., López, M. C., Patarroyo, M. E., and Alonso, C. (1992). Regulation of hsp70 expression in Trypanosoma cruzi by temperature and growth phase. *Molecular and Biochemical Parasitology*, 53(1–2), 201–211. [https://doi.org/10.1016/0166-6851\(92\)90022-C](https://doi.org/10.1016/0166-6851(92)90022-C)
- Requena-Méndez, A., López, M. C., Angheben, A., Izquierdo, L., Ribeiro, I., Pinazo, M. J., Gascon, J., and Muñoz, J. (2013). Evaluating Chagas disease progression and cure through blood-derived biomarkers: a systematic review. *Expert Review of Anti-Infective Therapy*, 11(9), 957–976. <https://doi.org/10.1586/14787210.2013.824718>
- Revollo, S., Oury, B., Laurent, J. P., Barnabé, C., Quesney, V., Carrière, V., Noël, S., and Tibayrenc, M. (1998). Trypanosoma cruzi: Impact of clonal evolution of the parasite on its biological and medical properties. *Experimental Parasitology*, 89(1), 30–39. <https://doi.org/10.1006/expr.1998.4216>
- Riarte, A., Luna, C., Sabatiello, R., Sinagra, A., Schiavelli, R., de Rissio, A., Maiolo, E., García, M. M., Jacob, N., Pattin, M., Lauricella, M., Segura, E. L., and Vázquez, M. (1999). Chagas' disease in patients with kidney transplants: 7 years of experience, 1989-1996. *Clinical Infectious Diseases*, 29(3), 561–567. <https://doi.org/10.1086/598634>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47. <https://doi.org/10.1093/NAR/GKV007>
- Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <https://doi.org/10.1186/GB-2010-11-3-R25>
- Rocha, M. O. C., Ribeiro, A. L. P., and Teixeira, M. M. (2003). Clinical management of chronic Chagas cardiomyopathy. *Frontiers in Bioscience*, 8, e44-54. <https://doi.org/10.2741/926>
- Rodrigues, M. M., Ribeirão, M., Pereira-Chioccola, V., Renia, L., and Costa, F. (1999). Predominance of CD4 Th1 and CD8 Tc1 cells revealed by characterization of the cellular immune response generated by immunization with a DNA vaccine containing a Trypanosoma cruzi gene. *Infection and Immunity*, 67(8), 3855–3863. <https://doi.org/10.1128/iai.67.8.3855-3863.1999>
- Rodriguez, H. O., Guerrero, N. A., Fortes, A., Santi-Rocca, J., Gironès, N., and Fresno, M. (2014). Trypanosoma cruzi strains cause different myocarditis patterns in infected mice. *Acta Tropica*, 139, 57–66. <https://doi.org/10.1016/j.actatropica.2014.07.005>

- Rodríguez, P., Montilla, M., Nicholls, S., Zarante, I., and Puerta, C. (1998). Isoenzymatic characterization of Colombian strains of *Trypanosoma cruzi*. *Memórias Do Instituto Oswaldo Cruz*, 93(6), 739–740. <https://doi.org/10.1590/S0074-02761998000600008>
- Rogalla, T., Ehrnsperger, M., Preville, X., Kotlyarov, A., Lutsch, G., Ducasse, C., Paul, C., Wieske, M., Arrigo, A. P., Buchner, J., and Gaestel, M. (1999). Regulation of Hsp27 oligomerization, chaperone function, and protective activity against oxidative stress/tumor necrosis factor by phosphorylation. *Journal of Biological Chemistry*, 274(27), 18947–18956. <https://doi.org/10.1074/jbc.274.27.18947>
- Romay-Barja, M., Boquete, T., Martínez, O., González, M., Álvarez-Del Arco, D., Benito, A., and Blasco-Hernández, T. (2019). Chagas screening and treatment among Bolivians living in Madrid, Spain: The need for an official protocol. *PLOS One*, 14(3), e0213577. <https://doi.org/10.1371/journal.pone.0213577>
- Rosenstein, S. P., and Been, M. D. (1990). Self-cleavage of hepatitis delta virus genomic strand RNA is enhanced under partially denaturing conditions. *Biochemistry*, 29(35), 8011–8016. <https://doi.org/10.1021/B100487A002>
- Rottenberg, M. E., Rodríguez, D. A., and Örn, A. (1992). Control of *Trypanosoma cruzi* infection in mice deprived of T-cell help. *Scandinavian Journal of Immunology*, 36(2), 261–268. <https://doi.org/10.1111/J.1365-3083.1992.TB03098.X>
- Rowland, E. C., Mikhail, K. S., and McCormick, T. S. (1992). Isotype determination of anti-*Trypanosoma cruzi* antibody in murine Chagas' disease. *Journal of Parasitology*, 78(3), 557–561. <https://doi.org/10.2307/3283671>
- Ruiz, J. H., and Becker, I. (2007). CD8 cytotoxic T cells in cutaneous leishmaniasis. *Parasite Immunology*, 29(12), 671–678. <https://doi.org/10.1111/J.1365-3024.2007.00991.X>
- Russomando, G., de Tomassone, M. M. C., de Guillen, I., Acosta, N., Vera, N., Almiron, M., Candia, N., Calcena, M. F., and Figueredo, A. (1998). Treatment of congenital Chagas' disease diagnosed and followed up by the polymerase chain reaction. *American Journal of Tropical Medicine and Hygiene*, 59(3), 487–491. <https://doi.org/10.4269/ajtmh.1998.59.487>
- Salehi-Ashtiani, K., Lupták, A., Litovchick, A., and Szostak, J. W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, 313(5794), 1788–1792. <https://doi.org/10.1126/SCIENCE.1129308>
- Sales, P. A., Molina, I., Murta, S. M. F., Sánchez-Montalvá, A., Salvador, F., Corrêa-Oliveira, R., and Carneiro, C. M. (2017). Experimental and clinical treatment of Chagas disease: A review. *American Journal of Tropical Medicine and Hygiene*, 97(5), 1289–1303. <https://doi.org/10.4269/ajtmh.16-0761>
- Salomão, K., and de Castro, S. L. (2017). Recent advances in drug development for Chagas disease: Two magic words, combination and repositioning. In *Different Aspects on Chemotherapy of Trypanosomatids* (Issue January).
- Salvador, F., Mego, M., Sánchez-Montalvá, A., Morís, M., Ramírez, K., Accarino, A., Malagelada, J. R., Azpiroz, F., and Molina, I. (2015). Assessment of rectocolonic morphology and function in patients with Chagas disease in Barcelona (Spain). *American Journal of Tropical Medicine and Hygiene*, 92(5), 898–902. <https://doi.org/10.4269/ajtmh.14-0546>
- Salvatella, R., Irabedra, P., and Castellanos, L. G. (2014). Interruption of vector transmission by native vectors and “the art of the possible.” *Memorias Do Instituto Oswaldo Cruz*, 109(1), 122–130. <https://doi.org/10.1590/0074-0276140338>

- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. In *Molecular cloning: a laboratory manual*. (2nd ed., Issue Ed. 2). Cold Spring Harbor Laboratory Press.
- San Francisco, J., Barría, I., Gutiérrez, B., Neira, I., Muñoz, C., Sagua, H., Araya, J. E., Andrade, J. C., Zailberger, A., Catalán, A., Remonsellez, F., Vega, J. L., and González, J. (2017). Decreased cruzipain and gp85/trans-sialidase family protein expression contributes to loss of *Trypanosoma cruzi* trypomastigote virulence. *Microbes and Infection*, *19*(1), 55–61. <https://doi.org/10.1016/J.MICINF.2016.08.003>
- Sánchez-Luque, F. J., López, M. C., Carreira, P. E., Alonso, C., and Thomas, M. C. (2014). The wide expansion of hepatitis delta virus-like ribozymes throughout trypanosomatid genomes is linked to the spreading of L1Tc/ingi clade mobile elements. *BMC Genomics*, *15*(1), 340. <https://doi.org/10.1186/1471-2164-15-340>
- Sánchez-Luque, F. J., López, M. C., Macías, F., Alonso, C., and Thomas, M. C. (2011). Identification of an hepatitis delta virus-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from *Trypanosoma cruzi*. *Nucleic Acids Research*, *39*(18), 8065–8077. <https://doi.org/10.1093/NAR/GKR478>
- Sánchez-Luque, F. J., López, M., Macías, F., Alonso, C., and Thomas, M. C. (2012). Pr77 and L1TcRz: A dual system within the 5'-end of L1Tc retrotransposon, internal promoter and HDV-like ribozyme. *Mobile Genetic Elements*, *2*(1), 1–7. <https://doi.org/10.4161/mge.19233>
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. <https://doi.org/10.1073/PNAS.74.12.5463>
- Santamaria, C., Chatelain, E., Jackson, Y., Miao, Q., Ward, B. J., Chappuis, F., and Ndao, M. (2014). Serum biomarkers predictive of cure in Chagas disease patients after nifurtimox treatment. *BMC Infectious Diseases*, *14*(1), 302. <https://doi.org/10.1186/1471-2334-14-302>
- Sardinha, L. R., Elias, R. M., Mosca, T., Bastos, K. R. B., Marinho, C. R. F., D'Império Lima, M. R., and Álvarez, J. M. (2006). Contribution of NK, NK T, gamma delta T, and alpha beta T cells to the gamma interferon response required for liver protection against *Trypanosoma cruzi*. *Infection and Immunity*, *74*(4), 2031–2042. <https://doi.org/10.1128/IAI.74.4.2031-2042.2006>
- Savino, W., Villa-Verde, D. M. S., Mendes-da-Cruz, D. A., Silva-Monteiro, E., Perez, A. R., Aoki, M. del P., Bottasso, O., Guiñazú, N., Silva-Barbosa, S. D., and Gea, S. (2007). Cytokines and cell adhesion receptors in the regulation of immunity to *Trypanosoma cruzi*. *Cytokine & Growth Factor Reviews*, *18*(1–2), 107–124. <https://doi.org/10.1016/J.CYTOGFR.2007.01.010>
- Schenone, H., Gaggero, M., Sapunar, J., Contreras, M. D. C., and Rojas, A. (2001). Congenital Chagas disease of second generation in Santiago, Chile. Report of two cases. *Revista Do Instituto de Medicina Tropical de Sao Paulo*, *43*(4), 231–232. <https://doi.org/10.1590/S0036-46652001000400011>
- Schenten, D., and Medzhitov, R. (2011). The control of adaptive immune responses by the innate immune system. *Advances in Immunology*, *109*, 87–124. <https://doi.org/10.1016/B978-0-12-387664-5.00003-0>
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863–864. <https://doi.org/10.1093/bioinformatics/btr026>



- Schmunis, G. A. (2007). Epidemiology of Chagas disease in non-endemic countries: The role of international migration. *Memorias Do Instituto Oswaldo Cruz*, 102(Suppl 1), 75–85. <https://doi.org/10.1590/s0074-02762007005000093>
- Schmunis, G. A., and Yadon, Z. E. (2010). Chagas disease: A Latin American health problem becoming a world health problem. *Acta Tropica*, 115(1–2), 14–21. <https://doi.org/10.1016/j.actatropica.2009.11.003>
- Seco-Hidalgo, V., de Pablos, L. M., and Osuna, A. (2015). Transcriptional and phenotypical heterogeneity of *Trypanosoma cruzi* cell populations. *Open Biology*, 5(12), 150190. <https://doi.org/10.1098/rsob.150190>
- Sen, S. K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P. A., Dyer, M., Cordaux, R., Liang, P., and Batzer, M. A. (2006). Human genomic deletions mediated by recombination between Alu elements. *American Journal of Human Genetics*, 79(1), 41–53. <https://doi.org/10.1086/504600>
- Seppely, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962, 227–245. [https://doi.org/10.1007/978-1-4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14)
- Sepulveda, P., Hontebeyrie, M., Liegeard, P., Mascilli, A., and Norris, K. A. (2000). DNA-Based immunization with *Trypanosoma cruzi* complement regulatory protein elicits complement lytic antibodies and confers protection against *Trypanosoma cruzi* infection. *Infection and Immunity*, 68(9), 4986–4991. <https://doi.org/10.1128/IAI.68.9.4986-4991.2000>
- Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G., and Taylor, J. (1988). Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *Journal of Virology*, 62(8), 2674–2679. <https://doi.org/10.1128/JVI.62.8.2674-2679.1988>
- Shevach, E. M. (2001). Certified professionals: CD4+CD25+ suppressor T cells. *The Journal of Experimental Medicine*, 193(11), F41–46. <https://doi.org/10.1084/JEM.193.11.F41>
- Shih, I., and Been, M. (2002). Catalytic strategies of the hepatitis delta virus ribozymes. *Annual Review of Biochemistry*, 71, 887–917. <https://doi.org/10.1146/ANNUREV.BIOCHEM.71.110601.135349>
- Silva, J. S., Machado, F. S., and Martins, G. A. (2003). The role of nitric oxide in the pathogenesis of Chagas disease. *Frontiers in Bioscience*, 8, s314–325. <https://doi.org/10.2741/1012>
- Silva, J. S., Morrissey, P. J., Grabstein, K. H., Mohler, K. M., Anderson, D., and Reed, S. G. (1992). Interleukin 10 and interferon gamma regulation of experimental *Trypanosoma cruzi* infection. *The Journal of Experimental Medicine*, 175(1), 169–174. <https://doi.org/10.1084/JEM.175.1.169>
- Silver, N., Best, S., Jiang, J., and Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*, 7, 33. <https://doi.org/10.1186/1471-2199-7-33>
- Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: Genomic organization, transcription, replication, and evolution. *Annual Review of Microbiology*, 41(1), 363–382. <https://doi.org/10.1146/annurev.mi.41.100187.002051>
- Singh-Jasuja, H., Hilf, N., Arnold-Schild, D., and Schild, H. (2001). The role of heat shock proteins and their receptors in the activation of the immune system. *Biological Chemistry*, 382(4), 629–636. <https://doi.org/10.1515/BC.2001.074>

- Sosa-Estani, S., Cura, E., Velazquez, E., Yampotis, C., and Segura, E. L. (2009). Etiological treatment of young women infected with *Trypanosoma cruzi*, and prevention of congenital transmission. *Revista Da Sociedade Brasileira de Medicina Tropical*, 42(5), 484–487. <https://doi.org/10.1590/S0037-86822009000500002>
- Sosa-Estani, S., Segura, E. L., Ruiz, A. M., Velazquez, E., Porcel, B. M., and Yampotis, C. (1998). Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas' disease. *American Journal of Tropical Medicine and Hygiene*, 59(4), 526–529. <https://doi.org/10.4269/ajtmh.1998.59.526>
- Sousa, G. R., Gomes, J. A. S., Damasio, M. P. S., Nunes, M. C. P., Costa, H. S., Medeiros, N. I., Fares, R. C. G., Chaves, A. T., Corrêa-Oliveira, R., and Rocha, M. O. C. (2017). The role of interleukin 17-mediated immune response in Chagas disease: High level is correlated with better left ventricular function. *PLoS ONE*, 12(3), e0172833. <https://doi.org/10.1371/journal.pone.0172833>
- Stahl, P., Ruppert, V., Schwarz, R. T., and Meyer, T. (2014). *Trypanosoma cruzi* evades the protective role of interferon-gamma-signaling in parasite-infected cells. *PLoS ONE*, 9(10), e110512. <https://doi.org/10.1371/journal.pone.0110512>
- Stoco, P. H., Wagner, G., Talavera-Lopez, C., Gerber, A., Zaha, A., Thompson, C. E., Bartholomeu, D. C., Lückemeyer, D. D., Bahia, D., Loreto, E., Prestes, E. B., Lima, F. M., Rodrigues-Luiz, G., Vallejo, G. A., Filho, J. F. da S., Schenkman, S., Monteiro, K. M., Tyler, K. M., Almeida, L. G. P. de, ... Grisard, E. C. (2014). Genome of the avirulent human-infective *Trypanosoma*—*Trypanosoma rangeli*. *PLoS Neglected Tropical Diseases*, 8(9), e3176. <https://doi.org/10.1371/journal.pntd.0003176>
- Strout, R. G. (1962). A method for concentrating hemoflagellates. *The Journal of Parasitology*, 48, 100. <https://doi.org/10.2307/3275421>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545–15550. <https://doi.org/10.1073/PNAS.0506580102>
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. v. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>
- Tarleton, R. L. (1990). Depletion of CD8+ T cells increases susceptibility and reverses vaccine-induced immunity in mice infected with *Trypanosoma cruzi*. *The Journal of Immunology*, 144(2), 717–724.
- Tarleton, R. L. (2001). Parasite persistence in the aetiology of Chagas disease. *International Journal for Parasitology*, 31(5–6), 550–554. [https://doi.org/10.1016/S0020-7519\(01\)00158-8](https://doi.org/10.1016/S0020-7519(01)00158-8)
- Tarleton, R. L. (2007). Immune system recognition of *Trypanosoma cruzi*. *Current Opinion in Immunology*, 19(4), 430–434. <https://doi.org/10.1016/j.coi.2007.06.003>
- Tarleton, R. L., Grusby, M. J., and Zhang, L. (2000). Increased susceptibility of Stat4-deficient and enhanced resistance in Stat6-deficient mice to infection with *Trypanosoma cruzi*. *The Journal of Immunology*, 165(3), 1520–1525. <https://doi.org/10.4049/jimmunol.165.3.1520>

- Teston, A. P. M., Monteiro, W. M., Reis, D., Bossolani, G. D. P., Gomes, M. L., de Araújo, S. M., Bahia, M. T., Barbosa, M. G. v., and Toledo, M. J. O. (2013). In vivo susceptibility to benznidazole of *Trypanosoma cruzi* strains from the western Brazilian Amazon. *Tropical Medicine & International Health*, 18(1), 85–95. <https://doi.org/10.1111/tmi.12014>
- Thill, G., Vasseur, M., and Tanner, N. K. (1993). Structural and sequence elements required for the self-cleaving activity of the hepatitis delta virus ribozyme. *Biochemistry*, 32(16), 4254–4262. <https://doi.org/10.1021/Bi00067A013>
- Thomas, M. C., Fernández-Villegas, A., Carrilero, B., Marañón, C., Saura, D., Noya, O., Segovia, M., de Noya, B. A., Alonso, C., and López, M. C. (2012). Characterization of an immunodominant antigenic epitope from *Trypanosoma cruzi* as a biomarker of chronic Chagas' disease pathology. *Clinical and Vaccine Immunology*, 19(2), 167–173. <https://doi.org/10.1128/CVI.05566-11>
- Thomas, M. C., García-Pérez, J. L., Alonso, C., and López, M. C. (2000). Molecular characterization of KMP11 from *Trypanosoma cruzi*: a cytoskeleton-associated protein regulated at the translational level. *DNA and Cell Biology*, 19(1), 47–57. <https://doi.org/10.1089/104454900314708>
- Thomas, M. C., Longobardo, M. v., Carmelo, E., Marañón, C., Planelles, L., Patarroyo, M. E., Alonso, C., and López, M. C. (2001). Mapping of the antigenic determinants of the *T. cruzi* kinetoplastid membrane protein-11. Identification of a linear epitope specifically recognized by human Chagasic sera. *Clinical and Experimental Immunology*, 123(3), 465–471. <https://doi.org/10.1046/J.1365-2249.2001.01478.X>
- Thomas, M. C., Macias, F., Alonso, C., and López, M. C. (2010). The biology and evolution of transposable elements in parasites. *Trends in Parasitology*, 26(7), 350–362. <https://doi.org/10.1016/j.pt.2010.04.001>
- Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Torrice, F., Alonso Vega, C., Billot, C., Truyens, C., and Carlier, Y. (2007). Relaciones materno-fetales en la infección con *T. cruzi* y la implementación de un programa nacional de detección y tratamiento de Chagas congénito en Bolivia. *Enfermedades Emergentes*, 9(Suppl 1), 9–16.
- Torrice, F., Alonso-Vega, C., Suarez, E., Rodriguez, P., Torrice, M. C., Dramaix, M., Truyens, C., and Carlier, Y. (2004). Maternal *Trypanosoma cruzi* infection, pregnancy outcome, morbidity, and mortality of congenitally infected and non-infected newborns in Bolivia. *American Journal of Tropical Medicine and Hygiene*, 70(2), 201–209. <https://doi.org/10.4269/ajtmh.2004.70.201>
- Trinchieri, G. (1995). Interleukin-12: a proinflammatory cytokine with immunoregulatory functions that bridge innate resistance and antigen-specific adaptive immunity. *Annual Review of Immunology*, 13, 251–276. <https://doi.org/10.1146/ANNUREV.IY.13.040195.001343>
- Vallejo, A., Monge-Maillo, B., Gutiérrez, C., Norman, F. F., López-Vélez, R., and Pérez-Molina, J. A. (2016). Changes in the immune response after treatment with benznidazole versus no treatment in patients with chronic indeterminate Chagas disease. *Acta Tropica*, 164, 117–124. <https://doi.org/10.1016/j.actatropica.2016.09.010>
- Vandesompele, J., de Preter, K., Pattyn, F., Poppe, B., van Roy, N., de Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric

- averaging of multiple internal control genes. *Genome Biology*, 3(7), research0034.1-0034.11. <https://doi.org/10.1186/gb-2002-3-7-research0034>
- Vargas, N., Pedroso, A., and Zingales, B. (2004). Chromosomal polymorphism, gene synteny and genome size in *T. cruzi* I and *T. cruzi* II groups. *Molecular and Biochemical Parasitology*, 138(1), 131–141. <https://doi.org/10.1016/j.molbiopara.2004.08.005>
- Vasconcelos, R. H. T., Amaral, F. N., Cavalcanti, M. G. A. M., Silva, E. D., Ferreira, A. G. P., Morais, C. N. L., and Gomes, Y. M. (2010). Increased levels of IgA antibodies against CRA and FRA recombinant antigens of *Trypanosoma cruzi* differentiate digestive forms of Chagas disease. *Human Immunology*, 71(10), 964–967. <https://doi.org/10.1016/j.HUMIMM.2010.07.004>
- Vázquez, M., Ben-Dov, C., Lorenzi, H., Moore, T., Schijman, A., and Levin, M. J. (2000). The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. *Proceedings of the National Academy of Sciences of the United States of America*, 97(5), 2128–2133. <https://doi.org/10.1073/PNAS.050578397>
- Verçosa, A. F. A., Lorena, V. M. B., Carvalho, C. L., Melo, M. F. A. D., Cavalcanti, M. G. A., Silva, E. D., Ferreira, A. G. P., Pereira, V. R. A., Souza, W. v., and Gomes, Y. M. (2007). Chagas' disease: IgG isotypes against cytoplasmic (CRA) and flagellar (FRA) recombinant repetitive antigens of *Trypanosoma cruzi* in chronic chagasic patients. *Journal of Clinical Laboratory Analysis*, 21(5), 271–276. <https://doi.org/10.1002/JCLA.20186>
- Vespa, G. N. R., Cunha, F. Q., and Silva, J. S. (1994). Nitric oxide is involved in control of *Trypanosoma cruzi*-induced parasitemia and directly kills the parasite in vitro. *Infection and Immunity*, 62(11), 5177–5182. <https://doi.org/10.1128/iai.62.11.5177-5182.1994>
- Viotti, R., Alarcón De Noya, B., Araujo-Jorge, T., Grijalva, M. J., Guhl, F., López, M. C., Ramsey, J. M., Ribeiro, I., Schijman, A. G., Sosa-Estani, S., Torrico, F., and Gascon, J. (2014). Towards a paradigm shift in the treatment of chronic Chagas disease. *Antimicrobial Agents and Chemotherapy*, 58(2), 635–639. <https://doi.org/10.1128/AAC.01662-13>
- Viotti, R., Vigliano, C., Álvarez, M. G., Lococo, B., Petti, M., Bertocchi, G., Armenti, A., de Rissio, A. M., Cooley, G., Tarleton, R., and Laucella, S. (2011). Impact of aetiological treatment on conventional and multiplex serology in chronic Chagas disease. *PLoS Neglected Tropical Diseases*, 5(9), e1314. <https://doi.org/10.1371/journal.pntd.0001314>
- Viotti, R., Vigliano, C., Lococo, B., Alvarez, M. G., Petti, M., Bertocchi, G., and Armenti, A. (2009). Side effects of benzimidazole as treatment in chronic Chagas disease: Fears and realities. *Expert Review of Anti-Infective Therapy*, 7(2), 157–163. <https://doi.org/10.1586/14787210.7.2.157>
- Viotti, R., Vigliano, C., Lococo, B., Bertocchi, G., Petti, M., Alvarez, M. G., Postan, M., and Armenti, A. (2006). Long-term cardiac outcomes of treating chronic Chagas disease with benzimidazole versus no treatment: A nonrandomized trial. *Annals of Internal Medicine*, 144(10), 724–734. <https://doi.org/10.7326/0003-4819-144-10-200605160-00006>
- Vitte, C., and Bennetzen, J. L. (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), 17638–17643. <https://doi.org/10.1073/pnas.0605618103>

- Voller, A., Draper, C., Bidwell, D. E., and Bartlett, A. (1975). Microplate enzyme-linked immunosorbent assay for Chagas' disease. *The Lancet*, 305(7904), 426–428. [https://doi.org/10.1016/S0140-6736\(75\)91492-0](https://doi.org/10.1016/S0140-6736(75)91492-0)
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11), e112963. <https://doi.org/10.1371/JOURNAL.PONE.0112963>
- Wang, X., Chen, M., Zhou, J., and Zhang, X. (2014). HSP27, 70 and 90, anti-apoptotic proteins, in clinical cancer therapy (review). *International Journal of Oncology*, 45(1), 18–30. <https://doi.org/10.3892/ijo.2014.2399>
- Webb, C.-H. T., and Lupták, A. (2011). HDV-like self-cleaving ribozymes. *RNA Biology*, 8(5), 719–727. <https://doi.org/10.4161/RNA.8.5.16226>
- Webb, C.-H. T., Nguyen, D., Myszka, M., and Lupták, A. (2016). Topological constraints of structural elements in regulation of catalytic activity in HDV-like self-cleaving ribozymes. *Scientific Reports*, 6, 28179. <https://doi.org/10.1038/srep28179>
- Webb, C.-H. T., Riccitelli, N. J., Ruminski, D. J., and Lupták, A. (2009). Widespread occurrence of self-cleaving ribozymes. *Science*, 326(5955), 953. <https://doi.org/10.1126/science.1178084>
- Weiner, A. M. (2002). SINEs and LINEs: The art of biting the hand that feeds you. *Current Opinion in Cell Biology*, 14(3), 343–350. [https://doi.org/10.1016/S0955-0674\(02\)00338-1](https://doi.org/10.1016/S0955-0674(02)00338-1)
- Wickstead, B., Ersfeld, K., and Gull, K. (2003). Repetitive elements in genomes of parasitic protozoa. *Microbiology and Molecular Biology Reviews*, 67(3), 360–375. <https://doi.org/10.1128/mubr.67.3.360-375.2003>
- Wincker, P., Murto-Dovalés, A. C., and Goldenberg, S. (1992). Nucleotide sequence of a representative member of a Trypanosoma cruzi dispersed gene family. *Molecular and Biochemical Parasitology*, 55(1–2), 217–220. [https://doi.org/10.1016/0166-6851\(92\)90142-7](https://doi.org/10.1016/0166-6851(92)90142-7)
- World Health Organization (WHO). (2002). *Control of Chagas disease: second report of the WHO expert committee*. <https://apps.who.int/iris/handle/10665/42443>
- World Health Organization (WHO). (2017). *Integrating neglected tropical diseases into global health and development: fourth WHO report on neglected tropical diseases*. <https://www.who.int/publications/i/item/9789241565448>
- World Health Organization (WHO). (2021). *Chagas disease (also known as American trypanosomiasis)*. [https://www.who.int/en/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/en/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))
- World Health Organization (WHO), and International Programme on Chemical Safety (IPCS). (1993). *Biomarkers in risk assessment: concepts and principles*. <https://apps.who.int/iris/bitstream/handle/10665/39037/9241571551-eng.pdf?sequence=1&isAllowed=y>
- World Health Organization (WHO), and International Programme on Chemical Safety (IPCS). (2001). *Biomarkers in risk assessment: validity and validation*. World Health Organization.
- Wrightsmann, R. A., and Manning, J. E. (2000). Paraflagellar rod proteins administered with alum and IL-12 or recombinant adenovirus expressing IL-12 generates antigen-

- specific responses and protective immunity in mice against *Trypanosoma cruzi*. *Vaccine*, *18*(14), 1419–1427. [https://doi.org/10.1016/S0264-410X\(99\)00380-1](https://doi.org/10.1016/S0264-410X(99)00380-1)
- Wu, H. N., Lin, Y. J., Lin, F. P., Makino, S., Chang, M. F., and Lai, M. M. (1989). Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(6), 1831–1835. <https://doi.org/10.1073/PNAS.86.6.1831>
- Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. (2012). miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Molecular Biology*, *80*(1), 75–84. <https://doi.org/10.1007/s11103-012-9885-2>
- Yamagata, Y., and Nakagawa, J. (2006). Control of Chagas disease. *Advances in Parasitology*, *61*, 129–165. [https://doi.org/10.1016/S0065-308X\(05\)61004-4](https://doi.org/10.1016/S0065-308X(05)61004-4)
- Zacks, M. A., Wen, J. J., Vyatkina, G., Bhatia, V., and Garg, N. (2005). An overview of chagasic cardiomyopathy: Pathogenic importance of oxidative stress. *Anais Da Academia Brasileira de Ciencias*, *77*(4), 695–715. <https://doi.org/10.1590/S0001-37652005000400009>
- Zielinski, C. E., Mele, F., Aschenbrenner, D., Jarrossay, D., Ronchi, F., Gattorno, M., Monticelli, S., Lanzavecchia, A., and Sallusto, F. (2012). Pathogen-induced human TH17 cells produce IFN- $\gamma$  or IL-10 and are regulated by IL-1 $\beta$ . *Nature*, *484*(7395), 514–518. <https://doi.org/10.1038/nature10957>
- Zingales, B. (2018). *Trypanosoma cruzi* genetic diversity: Something new for something known about Chagas disease manifestations, serodiagnosis and drug sensitivity. *Acta Tropica*, *184*, 38–52. <https://doi.org/10.1016/j.actatropica.2017.09.017>
- Zingales, B., Miles, M. A., Campbell, D. A., Tibayrenc, M., Macedo, A. M., Teixeira, M. M. G., Schijman, A. G., Llewellyn, M. S., Lages-Silva, E., Machado, C. R., Andrade, S. G., and Sturm, N. R. (2012). The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. *Infection, Genetics and Evolution*, *12*(2), 240–253. <https://doi.org/10.1016/j.meegid.2011.12.009>
- Zingales, B., Stolf, B. S., Souto, R. P., Fernandes, O., and Briones, M. R. S. (1999). Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. *Memorias Do Instituto Oswaldo Cruz*, *94*(Suppl 1), 159–164. <https://doi.org/10.1590/S0074-02761999000700020>

## **ANEXOS**

---





**Anexo 1. Secuencias empleadas para la identificación y validación de retroelementos por BLASTN en los genomas de *Trypanosoma cruzi*.**

<b>Pr77</b>	<a href="https://saco.csic.es/index.php/s/SamL9KaPfrjszQi">https://saco.csic.es/index.php/s/SamL9KaPfrjszQi</a>
<b>L1Tc</b>	<a href="https://saco.csic.es/index.php/s/Roz3pbHiiqrP43X">https://saco.csic.es/index.php/s/Roz3pbHiiqrP43X</a>
<b>NARTc</b>	<a href="https://saco.csic.es/index.php/s/oqTQJzTrKz8oSNS">https://saco.csic.es/index.php/s/oqTQJzTrKz8oSNS</a>
<b>DIREs</b>	<a href="https://saco.csic.es/index.php/s/DpHT3bwdEYBknHC">https://saco.csic.es/index.php/s/DpHT3bwdEYBknHC</a>

**Anexo 2. Ensayos TaqMan® usados para el análisis por RT-qPCR usando el sistema *QuantStudio™ 12K Flex Real-Time PCR* (n=112).** En cursiva se señalan los genes típicamente utilizados como genes de referencia según la bibliografía. Las letras en negrita corresponden a genes que no amplificaron en ninguna de las muestras analizadas.

Nº	Símbolo del gen	ID del ensayo	Nombre del gen
1	<i>ACTB</i>	Hs99999903_m1	<i>actin beta</i>
2	<i>B2M</i>	Hs99999907_m1	<i>beta-2-microglobulin</i>
3	B3GAT1	Hs00218629_m1	<i>beta-1,3-glucuronyltransferase 1</i>
4	BCL2	Hs00153350_m1	<i>BCL2, apoptosis regulator</i>
5	BTLA	Hs00699198_m1	<i>B and T lymphocyte associated</i>
6	CASP3	Hs00234387_m1	<i>caspase 3</i>
7	CCL5	Hs00982282_m1	<i>C-C motif chemokine ligand 5</i>
8	CCR1	Hs00174298_m1	<i>C-C motif chemokine receptor 1</i>
9	CCR5	Hs00152917_m1	<i>C-C motif chemokine receptor 5</i>
10	CCR7	Hs01013469_m1	<i>C-C motif chemokine receptor 7</i>
11	CD160	Hs00199894_m1	<i>CD160 molecule</i>
12	CD19	Hs00174333_m1	<i>CD19 molecule</i>
13	<b>CD1A</b>	Hs00233332_m1	<i>CD1a molecule</i>
14	CD2	Hs00233515_m1	<i>CD2 molecule</i>
15	CD27	Hs00386811_m1	<i>CD27 molecule</i>
16	CD274	Hs01125301_m1	<i>CD274 molecule</i>
17	CD28	Hs01007422_m1	<i>CD28 molecule</i>
18	CD3E	Hs01062241_m1	<i>CD3e molecule</i>
19	CD4	Hs01058407_m1	<i>CD4 molecule</i>
20	CD40	Hs00374176_m1	<i>CD40 molecule</i>
21	CD40LG	Hs00163934_m1	<i>CD40 ligand</i>
22	CD48	Hs00152927_m1	<i>CD48 molecule</i>
23	CD58	Hs00156385_m1	<i>CD58 molecule</i>
24	CD69	Hs00934033_m1	<i>CD69 molecule</i>
25	CD80	Hs00175478_m1	<i>CD80 molecule</i>
26	CD83	Hs00188486_m1	<i>CD83 molecule</i>
27	CD86	Hs01567026_m1	<i>CD86 molecule</i>

28	CD8A	Hs00233520_m1	<i>CD8a molecule</i>
29	CLEC9A	Hs01651638_m1	<i>C-type lectin domain family 9 member A</i>
30	CR2	Hs00153398_m1	<i>complement component 3d receptor 2</i>
31	CSF1	Hs00174164_m1	<i>colony stimulating factor 1</i>
32	CSF2	Hs00929873_m1	<i>colony stimulating factor 2</i>
33	CTLA4	Hs00175480_m1	<i>cytotoxic T-lymphocyte associated protein 4</i>
34	CXCR3	Hs00171041_m1	<i>C-X-C motif chemokine receptor 3</i>
35	FAS	Hs00236330_m1	<i>Fas cell surface death receptor</i>
36	FASLG	Hs00181225_m1	<i>Fas ligand</i>
37	FCER1A	Hs00758599_m1	<i>Fc fragment of IgE receptor Ia</i>
38	FCER2	Hs00233627_m1	<i>Fc fragment of IgE receptor II</i>
39	FOXP3	Hs01085834_m1	<i>forkhead box P3</i>
40	GAPDH	Hs99999905_m1	<i>glyceraldehyde-3-phosphate dehydrogenase</i>
41	GATA3	Hs00231122_m1	<i>GATA binding protein 3</i>
42	GZMA	Hs00196206_m1	<i>granzyme A</i>
43	GZMB	Hs00188051_m1	<i>granzyme B</i>
44	GZMH	Hs00277212_m1	<i>granzyme H</i>
45	GZMK	Hs00157878_m1	<i>granzyme K</i>
46	GZMM	Hs00193417_m1	<i>granzyme M</i>
47	HAVCR2	Hs00262170_m1	<i>hepatitis A virus cellular receptor 2</i>
48	HPRT1	Hs99999909_m1	<i>hypoxanthine phosphoribosyltransferase 1</i>
49	ICAM1	Hs00164932_m1	<i>intercellular adhesion molecule 1</i>
50	ICOS	Hs00359999_m1	<i>inducible T-cell costimulator</i>
51	ICOSLG	Hs00323621_m1	<i>inducible T-cell costimulator ligand</i>
52	IDO1	Hs00984148_m1	<i>indoleamine 2,3-dioxygenase 1</i>
53	IFNAR1	Hs01066116_m1	<i>interferon alpha and beta receptor subunit 1</i>
54	IFNG	Hs00989291_m1	<i>interferon gamma</i>
55	IFNGR1	Hs00988304_m1	<i>interferon gamma receptor 1</i>
56	IFNGR2	Hs00194264_m1	<i>interferon gamma receptor 2</i>
57	IL10	Hs00961622_m1	<i>interleukin 10</i>
58	IL10RA	Hs00155485_m1	<i>interleukin 10 receptor subunit alpha</i>
59	IL10RB	Hs00175123_m1	<i>interleukin 10 receptor subunit beta</i>
60	IL12A	Hs00168405_m1	<i>interleukin 12A</i>
61	IL12B	Hs00233688_m1	<i>interleukin 12B</i>
62	IL12RB1	Hs00234651_m1	<i>interleukin 12 receptor subunit beta 1</i>
63	IL12RB2	Hs00155486_m1	<i>interleukin 12 receptor subunit beta 2</i>
64	IL13	Hs00174379_m1	<i>interleukin 13</i>
65	IL17A	Hs00174383_m1	<i>interleukin 17A</i>
66	IL17RA	Hs01064648_m1	<i>interleukin 17 receptor A</i>
67	IL18	Hs01038788_m1	<i>interleukin 18</i>
68	IL18R1	Hs00175381_m1	<i>interleukin 18 receptor 1</i>
69	IL1B	Hs01555410_m1	<i>interleukin 1 beta</i>
70	IL2	Hs00174114_m1	<i>interleukin 2</i>

71	IL23A	Hs00372324_m1	<i>interleukin 23 subunit alpha</i>
72	IL23R	Hs00332759_m1	<i>interleukin 23 receptor</i>
73	<b>IL25</b>	Hs03044841_m1	<i>interleukin 25</i>
74	IL27	Hs00377366_m1	<i>interleukin 27</i>
75	IL2RA	Hs00166229_m1	<i>interleukin 2 receptor subunit alpha</i>
76	IL2RG	Hs00953624_m1	<i>interleukin 2 receptor subunit gamma</i>
77	IL4	Hs00174122_m1	<i>interleukin 4</i>
78	IL4R	Hs00965056_m1	<i>interleukin 4 receptor</i>
79	IL5	Hs00174200_m1	<i>interleukin 5</i>
80	IL5RA	Hs00602482_m1	<i>interleukin 5 receptor subunit alpha</i>
81	IL6	Hs00985639_m1	<i>interleukin 6</i>
82	IL6R	Hs01075666_m1	<i>interleukin 6 receptor</i>
83	IL7	Hs00174202_m1	<i>interleukin 7</i>
84	IL7R	Hs00233682_m1	<i>interleukin 7 receptor</i>
85	ITGA4	Hs00168433_m1	<i>integrin subunit alpha 4</i>
86	ITGAL	Hs00158218_m1	<i>integrin subunit alpha L</i>
87	ITGAX	Hs00174217_m1	<i>integrin subunit alpha X</i>
88	ITGB2	Hs00164957_m1	<i>integrin subunit beta 2</i>
89	KLRG1	Hs00195153_m1	<i>killer cell lectin like receptor G1</i>
90	LAG3	Hs00158563_m1	<i>lymphocyte activating 3</i>
91	LGALS9	Hs00247135_m1	<i>galectin 9</i>
92	MKI67	Hs01032443_m1	<i>marker of proliferation Ki-67</i>
93	NCAM1	Hs00941830_m1	<i>neural cell adhesion molecule 1</i>
94	NOS2	Hs01075529_m1	<i>nitric oxide synthase 2</i>
95	PDCD1	Hs01550088_m1	<i>programmed cell death 1</i>
96	PDCD1LG2	Hs01057777_m1	<i>programmed cell death 1 ligand 2</i>
97	<i>PGK1</i>	Hs99999906_m1	<i>phosphoglycerate kinase 1</i>
98	PRF1	Hs00169473_m1	<i>perforin 1</i>
99	SELL	Hs01046459_m1	<i>selectin L</i>
100	STAT1	Hs01013996_m1	<i>signal transducer and activator of transcription 1</i>
101	STAT3	Hs00374280_m1	<i>signal transducer and activator of transcription 3</i>
102	<i>TBP</i>	Hs99999910_m1	<i>TATA-box binding protein</i>
103	TBX21	Hs00203436_m1	<i>T-box 21</i>
104	TGFB1	Hs00998133_m1	<i>transforming growth factor beta 1</i>
105	TGFB2	Hs00234244_m1	<i>transforming growth factor beta 2</i>
106	TGFBR1	Hs00610320_m1	<i>transforming growth factor beta receptor 1</i>
107	TGFBR2	Hs00234253_m1	<i>transforming growth factor beta receptor 2</i>
108	TNF	Hs00174128_m1	<i>tumor necrosis factor</i>
109	TNFRSF14	Hs00998604_m1	<i>TNF receptor superfamily member 14</i>
110	TNFRSF1A	Hs00533560_m1	<i>TNF receptor superfamily member 1A</i>
111	TNFSF10	Hs00921974_m1	<i>tumor necrosis factor superfamily member 10</i>
112	XCR1	Hs00245540_s1	<i>X-C motif chemokine receptor 1</i>



## ANEXO DE PUBLICACIONES

Macías, F., Afonso-Lehmann, R., López, M. C., **Gómez, I.**, and Thomas, M. C. (2018). Biology of *Trypanosoma cruzi* retrotransposons: From an enzymatic to a structural point of view. *Current Genomics*, 19(2), 110–118. <https://doi.org/10.2174/1389202918666170815150738>

Egui, A., Thomas, M.C., Fernández-Villegas, A., Pérez-Antón, E., **Gómez, I.**, Carrilero, B., del Pozo, Á., Ceballos, M., Andrés-León, E., López-Ruz, M. Á., Gainza, E., Oquiñena, E., Segovia, M., and López, M. C. (2019). A parasite biomarker set for evaluating benznidazole treatment efficacy in patients with chronic asymptomatic *Trypanosoma cruzi* infection. *Antimicrobial Agents and Chemotherapy*, 63(10), e02436-18. <https://doi.org/10.1128/AAC.02436-18>

**Gómez, I.**, Rastrojo, A., Sánchez-Luque, F. J., Lorenzo-Díaz, F., Macías, F., Valladares, B., Aguado, B., Requena, J. M., López, M. C., and Thomas, M. C. (2020). Draft genome sequence of the *Trypanosoma cruzi* B. M. López strain (Tcla), isolated from a Colombian patient. *Microbiology Resource Announcements*, 9(18), e00031-20. <https://doi.org/10.1128/MRA.00031-20>

**Gómez, I.**, Rastrojo, A., Lorenzo-Díaz, F., Sánchez-Luque, F. J., Macías, F., Aguado, B., Valladares, B., Requena, J. M., López, M. C., and Thomas, M. C. (2020). *Trypanosoma cruzi* Ikiakarora (TcIII) draft genome sequence. *Microbiology Resource Announcements*, 9(27), e00453-20. <https://doi.org/10.1128/MRA.00453-20>

**Capítulo de libro:** Thomas, M.C., Egui, A., Pérez-Antón, E., **Gómez, I.**, and López, M.C. (2020). Challenges in response to treatment evaluation and progression of the disease. In: Pinazo Delgado M.J., Gascón J. (eds) *Chagas Disease*. Springer, Cham. [https://doi.org/10.1007/978-3-030-44054-1\\_11](https://doi.org/10.1007/978-3-030-44054-1_11)

**Gómez, I.**, López, M. C., Rastrojo, A., Lorenzo-Díaz, F., Requena, J. M., Aguado, B., Valladares, B., and Thomas, M. C. (2021). Variability of the Pr77-hallmark dual system among six *T. cruzi* strains belonging to different discrete typing units (DTUs). *Acta Tropica*, 222, 106053. <https://doi.org/10.1016/j.actatropica.2021.106053>

Alonso-Padilla, J., López, M. C., Esteva, M., Zrein, M., Casellas, A., **Gómez, I.**, Granjon, E., Méndez, S., Benítez, C., Ruiz, A. M., Sanz, S., Gascón, J., Thomas, M. C., Pinazo, M. J., Abril, M., de Noya, B. A., Jorge, T. A., Chatelain, E., Grijalva, M. J., ... Viotti, R. (2021). Serological reactivity against *T. cruzi*-derived antigens: Evaluation of their suitability for the assessment of response to treatment in chronic Chagas disease. *Acta Tropica*, 221, 105990. <https://doi.org/10.1016/j.actatropica.2021.105990>

**Gómez, I.**, Thomas, M. C., Palacios, G., Egui, A., Carrilero, B., Simón, M., Valladares, B., Segovia, M., Carmelo, E., and López, M. C. (2021). Differential expression of immune response genes in asymptomatic chronic Chagas disease patients versus healthy subjects. *Frontiers in Cellular and Infection Microbiology*, 11, 722984. <https://doi.org/10.3389/FCIMB.2021.722984>

