

# UACM

Universidad Autónoma  
de la Ciudad de México

NADA HUMANO ME ES AJENO

UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO

COLEGIO DE CIENCIA Y TECNOLOGÍA

POSGRADO EN CIENCIAS GENÓMICAS

**Efecto de la sobreexpresión del factor de transcripción  
EhMyb10 y el factor de *splicing* EhCDC5-like sobre el  
transcriptoma de *Entamoeba histolytica***

TESIS

QUE PARA OPTAR POR EL GRADO DE

**MAESTRA EN CIENCIAS GENÓMICAS**

PRESENTA:

**CLAUDIA CASTILLO MERCADO**

DIRECTORA: **DRA. ELISA IRENE AZUARA LICEAGA**

CODIRECTOR: **DR. JESÚS VALDÉS FLORES**

POSGRADO  
EN CIENCIAS  
GENÓMICAS

Ciudad de México, mayo de 2024.

## SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



## UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

### RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

### DERECHOS RESERVADOS ©

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.

## **INTEGRACIÓN DEL JURADO:**

Presidente: Dra. Silvia Castellanos Castro, UACM  
Secretario: Dr. Jesús Valdés Flores, Cinvestav-IPN  
Vocal: Dra. Elisa Azuara Liceaga, UACM

Plantel de adscripción:

PLANTEL DEL VALLE, UACM.

### **DIRECTOR**

---

Dra. Elisa Irene Azuara Liceaga  
Universidad Autónoma de la Ciudad de México

### **CODIRECTOR**

---

Dr. Jesús Valdés Flores  
Centro de Investigación y de Estudios Avanzados, IPN

## **Asesores**

**Dra. Lilia López Cánovas**

Posgrado en Ciencias Genómicas  
Universidad Autónoma de la Ciudad de México

**Dra. Patricia Areli Cuellar Silva, UACM**

Posgrado en Ciencias Genómicas  
Universidad Autónoma de la Ciudad de México

**Dr. Ricardo Grande Cano**

Instituto de Biotecnología  
Universidad Nacional Autónoma de México

## **Lectores**

**Dra. Lilia López Cánovas**

Posgrado en Ciencias Genómicas  
Universidad Autónoma de la Ciudad de México

**Dra. Silvia Castellanos Castro**

Colegio de Ciencias y Humanidades  
Universidad Autónoma de la Ciudad de México

**Dra. Ángeles Concepción Tecalco Cruz**

Posgrado en Ciencias Genómicas  
Universidad Autónoma de la Ciudad de México

Este trabajo se realizó con el apoyo del Consejo Nacional de Ciencia y Tecnología (CONACYT) con el número de CVU 958316, en el laboratorio 2 del Posgrado en Ciencias Genómicas en el plantel del Valle de la UACM, bajo la dirección de la Dra. Elisa Irene Azuara Liceaga, en colaboración con el Dr. Jesús Valdés Flores, del Departamento de Bioquímica del Cinvestav-IPN. Este trabajo forma parte del proyecto “*Control co-transcripcional de genes relacionados a la virulencia y el enquistamiento de Entamoeba*” (CF-2019-194163), aprobado en la convocatoria Ciencia de Frontera 2019.

## **Agradecimientos**

Al Dr. Jesús Valdés Flores, Departamento de Bioquímica del Centro de Investigación y de Estudios Avanzados, IPN.

A la Dra. Abigaíl Betanzos Fernández, Departamento de Infectómica y Patogénesis Molecular del Centro de Investigación y de Estudios Avanzados, IPN.

A la Dra. Elizabeth Jacqueline Castañeda Ortiz, Posgrado en Ciencias Genómicas, UACM

A la M. en C. Laura Isabel Vázquez Carrillo, Posgrado en Ciencias Genómicas, UACM

Al Dr. Ricardo Alfredo Grande Cano, Unidad Universitaria de Secuenciación Masiva y Bioinformática (UUSMB), UNAM.

Al M. en C. Carlos Ernesto Martínez Rodríguez, Plantel Casa Libertad, UACM.

## **Dedicatoria**

A Roberto Galindo mi compañero de vida y padre de mi hijo, por tu gran apoyo, paciencia y comprensión. Gracias por todo, te amo.

A János Galindo, mi hijo, por ser mi motor y apoyo.

A toda mi familia quien siempre ha estado presente apoyándome y escuchándome.

A la Dra. Elisa Irene Azuara Liceaga y al Dr. Jesús Valdés Flores por su paciencia, orientación y apoyo invaluable en el proceso de esta tesis. Sus conocimientos, paciencia y compromiso han sido fundamentales para mi éxito académico.

A la Dra. Patricia por su paciencia y todo su apoyo en el proceso de esta tesis, así como sus enseñanzas y su amistad.

A la Dra. Jacqueline por su apoyo, experiencia, escucharme y su amistad.

A mi compañero de generación Jorge, que hicimos un buen equipo de apoyo en la maestría.

A los profesores del posgrado que siempre me apoyaron y que dedicaron su tiempo a enseñarnos.

A mis compañeros de laboratorio por escucharme y apoyarme. Fernando, Mariana, Danna, Cesar, Aldair, Alan, Marlen.

## **IN MEMORIAM**

A mi padre Crisóforo Castillo Hernández, un hombre fuerte, trabajador e inquieto. Aunque ya no estés conmigo, tu espíritu vivirá siempre en mi corazón.

# Índice

Resumen.....	1
<b>Abstract.....</b>	<b>2</b>
<b>1. Introducción.....</b>	<b>3</b>
1.1. Expresión génica.....	3
1.2. Transcripción.....	3
1.2.1. Inicio de la transcripción.....	3
1.2.2 Elongación y término de la transcripción.....	5
1.3. <i>Splicing</i> .....	7
1.3.1. NTC.....	8
1.3.2. CDC5.....	10
1.3.3. Retención Intrónica (RI).....	12
1.4. Regulación cotranscripcional.....	14
1.5. <i>Entamoeba histolytica</i> : enfermedad, epidemiología, ciclo de vida.....	16
1.6. Genoma de <i>E. histolytica</i> .....	18
<b>2. Antecedentes particulares.....</b>	<b>20</b>
2.1. Factores de transcripción en <i>E. histolytica</i> .....	20
2.2. <i>Splicing</i> en <i>E. histolytica</i> .....	21
2.3. Poliadenilación.....	23
2.4. Análisis Transcriptómicos en <i>E. histolytica</i> .....	24
2.5. Proteínas Myb como factores de transcripción en <i>E. histolytica</i> .....	27
2.6. EhMyb10.....	29
2.7. EhCDC5.....	32
<b>3. Hipótesis.....</b>	<b>37</b>
<b>4. Justificación.....</b>	<b>37</b>
<b>5. Objetivo general:.....</b>	<b>37</b>
<b>6. Objetivos particulares:.....</b>	<b>37</b>
<b>7. Estrategia experimental.....</b>	<b>38</b>
<b>8. Materiales y Métodos.....</b>	<b>39</b>
8.1. Identificación in silico de los genes blanco que son regulados cotranscripcionalmente por EhMyb10 y EhCDC5 en <i>E. histolytica</i> .....	39
8.2. Análisis ontológico de los genes.....	39
8.3. Cultivo de trofozoítos de <i>E. histolytica</i> .....	39
8.4. Obtención de la construcción pEhEx-EhCDC5.....	40
8.5. Extracción de DNA plasmídico para transfección.....	40
8.6. Transfección de trofozoítos de <i>E. histolytica</i> con las construcciones pEhEx-EhCDC5 y pKT3M-EhMyb10.....	41
8.7. Aislamiento de RNA total de los trofozoítos transfectados.....	42
8.8. Extracción y cuantificación de DNA genómico.....	42
8.9. Análisis de la expresión de <i>ehmyb10</i> y <i>ehcdc5</i> por qRT-PCR.....	42
8.10. Diseño de oligonucleótidos.....	43
8.11. Inmunofluorescencia indirecta (IFI).....	44
8.12. Western Blot.....	44

8.13. Análisis in vivo de los patrones de <i>splicing</i> en los trofozoítos que sobreexpresan a las proteínas EhMyb10 y EhCDC5.....	44
8.14. Construcción y secuenciación de bibliotecas y obtención de los transcriptomas.....	45
<b>9. Resultados.....</b>	<b>46</b>
9.1. Selección de genes candidatos de EhMyb10 y EhCDC5 sobre los patrones de <i>splicing</i> de <i>E. histolytica</i> .....	46
9.2. Sobreexpresión de EhMyb10 y EhCDC5 en trofozoítos de <i>E. histolytica</i> ...50	
9.2.1. Clonación del gen <i>ehcdc5</i> en el vector pEhEx.....	50
9.2.2. Análisis de la expresión de los genes <i>ehmyb10</i> y <i>ehcdc5</i> por RT-PCR tiempo real.....	52
9.2.3. Expresión y localización de las proteínas de EhMyb10 y EhCDC5 en los trofozoítos transfectados de <i>E. histolytica</i> .....	53
9.2.4. Expresión de la proteína de EhCDC5 en los trofozoítos transfectados de <i>E. histolytica</i> .....	54
9.3. Efecto de la sobreexpresión de EhMyb10 y EhCDC5 sobre los patrones de <i>splicing in vivo</i> de <i>E. histolytica</i> .....	56
9.4. Efecto de la sobreexpresión de EhMyb10 y EhCDC5 en la regulación de la expresión génica de <i>E. histolytica</i> .....	60
9.4.1. Análisis de calidad e integridad del RNA total.....	60
9.4.2. Análisis de calidad de las lecturas.....	61
9.4.3. Análisis de la sobreexpresión de EhMyb10ox y EhCDC5ox.....	62
9.4.4. Análisis de los transcritos que modifican su expresión en EhMyb10 y EhCDC5.....	64
9.4.5. Condición de sobreexpresión de EhMyb10.....	64
9.4.6. Condición de sobreexpresión de EhCDC5.....	69
9.4.7. Condición de sobreexpresión de EhMyb10+EhCDC5.....	73
9.4.8. Análisis de expresión de los genes seleccionados por medio de análisis Ballgown.....	76
9.4.9. Motivos encontrados en los promotores de los genes que modificaron su expresión al sobreexpresar a las proteínas EhMyb10 y EhCDC5.....	78
<b>10. Discusión.....</b>	<b>80</b>
<b>11. Conclusiones.....</b>	<b>87</b>
<b>12. Perspectivas.....</b>	<b>89</b>
<b>13. Bibliografía.....</b>	<b>90</b>
<b>14. Anexos.....</b>	<b>97</b>

## Índice de tablas

Tabla 1. Composición del complejo NTC.....	10
Tabla 2. Características del genoma de <i>E. histolytica</i> .....	19
Tabla 3. Oligos empleados en este proyecto de investigación.....	43
Tabla 4. Niveles de expresión de genes que contienen el ERM en sus promotores.....	46
Tabla 5. Genes seleccionados (candidatos) para analizar el efecto de EhMyb10 y EhCDC5 sobre los patrones de <i>splicing</i> de <i>E. histolytica</i> .....	50
Tabla 6. Análisis de calidad de RNA total.....	61
Tabla 7. Cobertura del transcriptoma de <i>E. histolytica</i> .....	62
Tabla 8. Genes que modifican su expresión en los trofozoítos que sobreexpresan a las proteínas EhMyb10, EhCDC5 y EhMyb10+EhCDC5.....	63
Tabla 9. Genes diferencialmente expresados en EhMyb10ox.....	68
Tabla 10. Genes diferencialmente expresados en EhCDC5ox.....	72
Tabla 11. Genes diferencialmente expresados en EhMyb10+EhCDC5.....	75
Tabla 12. Motivos conservados en los genes diferencialmente expresados.....	79

## Índice de figuras

Figura 1. Esquema de un factor de transcripción prototipo.....	5
Figura 2. Proceso de la expresión génica en células eucariotas.....	6
Figura 3. Secuencias conservadas encontradas en los intrones del premRNA en metazoos y levadura ( <i>S. cerevisiae</i> ).....	8
Figura 4. Vía canónica de ensamblaje y eliminación de intrones del spliceosoma.....	8
Figura 5. Funciones del complejo NTC.....	9
Figura 6. Diagrama de la proteína CDC5 en <i>Schizosaccharomyces pombe</i> .....	11
Figura 7. Modelo del NTC y CDC5 en <i>S. pombe</i> .....	12
Figura 8. Mecanismo de retención intrónica (RI) producido por <i>splicing</i> alternativo.....	14
Figura 9. Ciclo de vida de <i>E. histolytica</i> .....	17
Figura 10. Conservación de sitios de <i>splicing</i> en <i>E. histolytica</i> .....	22
Figura 11. Moléculas relacionadas con la virulencia reveladas a partir de estudios transcriptómicos.....	26
Figura 12. Clasificación de genes basada en TPM normalizado log2.....	27
Figura 13. Las proteínas que contienen DBD de <i>E. histolytica</i> MYB forman tres familias bien definidas.....	29
Figura 14. EhMyb10 de <i>E. histolytica</i> .....	30
Figura 15. Interactoma <i>in silico</i> de EhMyb10.....	31
Figura 16. Proteínas que interactúan con EhMyb10.....	32
Figura 17. Interactoma de EhCDC5.....	33
Figura 18. Algunas proteínas del complejo de <i>splicing</i> identificadas mediante CLIP.....	33
Figura 19. Proteína EhCDC5 de <i>E. histolytica</i> .....	34
Figura 20. Alineamiento de EhCDC5.....	35
Figura 21. Procesamiento cotranscripcional de pre-mRNA en <i>E. histolytica</i> .....	36
Figura 22. Vectores utilizados para la construcción de EhMyb10 y EhCDC5.....	40
Figura 23. Clasificación de la expresión de los genes que contienen ERM en sus promotores.....	47
Figura 24. Presencia del ERM y secuencias de <i>splicing</i> en los posibles genes blanco de EhMyb10 y EhCDC5.....	48
Figura 25. Análisis ontológico de los posibles genes blanco de EhMyb10 y EhCDC5.....	49
Figura 26. Clonación y subclonación del gen <i>ehcdc5</i> de <i>E. histolytica</i> .....	51
Figura 27. Secuenciación de la construcción pEhEx-EhCDC5.....	51
Figura 28. Análisis de la expresión de los genes <i>ehmyb10</i> y <i>ehcdc5</i> .....	53
Figura 29. Localización subcelular de las proteínas EhMyb10 y EhCDC5 en trofozoítos transfectados.....	54
Figura 30. Localización subcelular de las proteínas en la cotransfección (EhMyb10+EhCDC5) en trofozoítos 72 hpt.....	54
Figura 31. Inmunodetección de la proteína EhMyb10 y EhCDC5 en trofozoítos transfectados.....	56
Figura 32. Análisis de la retención intrónica.....	59

Figura 33. Isoformas del gen que codifica para SNRP70.....	60
Figura 34. Gráficas de calidad de las lecturas obtenidas de RNAseq. ....	62
Figura 35. Diagramas de caja y bigotes de las lecturas obtenidas de RNAseq..	63
Figura 36. Genes diferencialmente expresados en los trofozoítos que sobreexpresan las proteínas EhMyb10, EhCDC5 y EhMyb10+EhCDC5.....	64
Figura 37. Análisis ontológico de los genes con expresión diferencial en trofozoítos transfectados con EhMyb10ox.....	67
Figura 38. Análisis ontológico de los genes con expresión diferencial en trofozoítos transfectados con EhCDC5ox.....	71
Figura 39. Análisis ontológico de los genes con expresión diferencial en trofozoítos transfectados con EhMyb10+EhCDC5.....	74
Figura 40. Niveles de expresión de los transcritos de los genes <i>hmg</i> , <i>sfr1</i> y <i>zinc knukle-domain</i> , procesados en cada condición.....	77
Figura 41. Niveles de expresión de los transcritos del gen <i>snrp70</i> procesados en cada condición.....	78

## Abreviaturas

<b>% RI</b>	Porcentaje de retención intrónica
<b>CAP</b>	Residuo de 7-metilguanosina
<b>CDC5</b>	Proteína Ciclo de División Celular 5
<b>CTD</b>	Dominio carboxilo terminal
<b>DAPI</b>	4, 6-diamino-2-fenilindol diclorhidrato
<b>DBD</b>	Dominio de unión a DNA
<b>DNA</b>	Ácido desoxiribonucleico
<b>ERM</b>	Elemento de reconocimiento a Myb
<b>FC</b>	Fold Change
<b>FT</b>	Factore(s) de transcripción
<b>HA</b>	Hemaglutinina
<b>hpt</b>	horas postransfección
<b>IFI</b>	Inmunofluorescencia indirecta
<b>INR</b>	Inicio de la transcripción
<b>IRI</b>	Isoformas de retención intrónica
<b>kb</b>	Kilobase
<b>kDa</b>	Kilodaltones
<b>LB</b>	Medio Luria Bertani
<b>mRNA</b>	RNA mensajero
<b>Myb</b>	Del inglés, Myeloblastosis
<b>NMD</b>	Por sus siglas en inglés, nonsense-mediated decay
<b>NTC</b>	Por su siglas inglés Nineteen complex
<b>pb</b>	Pares de bases
<b>PCR</b>	Reacción en cadena de la polimerasa
<b>Prp19</b>	Por su siglas en inglés, protein pre-mRNA processing factor 19 o nineteen complex
<b>RI</b>	Retención intrónica
<b>RNA</b>	Ácido Ribonucleico
<b>RNAPII</b>	RNA polimerasa II
<b>RNP</b>	Ribonucleoproteína
<b>RRM</b>	Motivos de reconocimiento a RNA
<b>RT-PCR</b>	Retro Transcripción reversa reacción en cadena de la polimerasa
<b>snRNP</b>	Por sus siglas en inglés small nuclear ribonucleoprotein, Ribonucleoproteínas nucleares pequeñas
<b>TMK</b>	Cinasas transmembrana
<b>TPM</b>	Transcritos por millón
<b>μL</b>	Microlitros

## Resumen

*Entamoeba histolytica*, regula y modifica su transcriptoma en respuesta a diferentes eventos como virulencia, enquistamiento y respuesta al estrés. Esta dinámica implica mecanismos como la transcripción y el *splicing* que pueden llevarse a cabo de manera simultánea, lo que se conoce como regulación cotranscripcional. Las proteínas con dominio de unión a DNA MYB (DBD-MYB) conforman una de las familias más abundantes relacionadas con la regulación transcripcional. De esta familia EhMyb10 y EhCDC5 se seleccionaron en este trabajo, para determinar su papel en la regulación cotranscripcional en *E. histolytica* evaluando sus funciones de factor de transcripción y de *splicing*, respectivamente. Para ello, primeramente, se realizó un análisis *in silico* de los 8,201 genes de este parásito, en los que se identificaron 147 genes que contienen el Elemento de Reconocimiento a Myb (ERM) en sus promotores y presentan intrones, por lo cual podrían ser genes blanco de estas proteínas. Además, sobreexpresamos transitoriamente a estas proteínas de manera independiente y conjunta en los trofozoítos de *E. histolytica*. La sobreexpresión se evaluó mediante RT-qPCR, WB y ensayos de inmunofluorescencia. Mediante estos ensayos se determinó que las proteínas se sobreexpresaron y que EhMyb10 y EhCDC5 colocalizan en el núcleo. EhCDC5 se detectó con un peso molecular de ~ 30 kDa y EhMyb10 de ~40 kDa y, lo cual sugiere la presencia de modificaciones postraduccionales que podrían ser necesarias para su regulación. Posteriormente, se analizaron los patrones de *splicing* de 4 genes seleccionados, los cuales modificaron su perfil de expresión al sobreexpresar a las proteínas en las diferentes condiciones. Finalmente, mediante un análisis de RNAseq se observó que 179, 243 y 115 genes modificaron su expresión en las condiciones de EhMyb10ox, EhCDC5ox y cuando se combinan, respectivamente. De estos genes, 61 están compartidos en todas las condiciones, los cuales podrían ser genes blanco de la regulación cotranscripcional. Estos genes están relacionados con diferentes funciones como: metabolismo de RNA, regulación transcripcional y propiedades de virulencia.

## Abstract

*Entamoeba histolytica*, regulates and modifies its transcriptome in response to different events such as virulence, encystment, and stress responses. This dynamic suggests that splicing and transcription processes may occur concurrently; this is referred to as cotranscriptional regulation. The MYB DNA binding domain proteins (DBD-MYB) belong to the most abundant families related to transcriptional regulation. EhMyb10 and EhCDC5 were selected from this family to determine their role in cotranscriptional regulation in *E. histolytica*, evaluating their functions as transcription and *splicing* factors, respectively. Initially, we performed an *in silico* analysis of the parasite's 8,201 genes, finding 147 genes with introns and the Myb recognition element (MRE) in their promoter region, which suggests that these genes may be targets of these proteins. In addition, we transiently overexpressed these proteins individually and in combination in *E. histolytica* trophozoites. Overexpression was evaluated through RT-qPCR, WB, and immunofluorescence assays. Through these assays, we determined that EhMyb10 and EhCDC5 are localized in the nucleus and cytoplasm and that they colocalize in the nucleus. EhCDC5 was detected with a molecular weight of ~ 30 kDa and EhMyb10 with ~40 kDa, which suggests the presence of postranslational modifications that could be necessary for their regulation. Subsequently, we examined the splicing patterns of four distinct genes that underwent modification in various overexpression conditions. Finally, through an RNAseq analysis, it was observed that 179, 243, and 115 genes modified their expression in EhMyb10ox, EhCDC5ox, and in combined conditions, respectively. From these genes, 61 were shared in all conditions, which could be target genes of cotranscriptional regulation. These genes are related to different functions, such as RNA metabolism, transcriptional regulation, and virulence properties.

# **1. Introducción**

## **1.1. Expresión génica**

La expresión génica es el proceso en el cual la información contenida en los genomas es transferida de DNA a RNA y de RNA a proteínas, involucrando e integrando diferentes eventos que se llevan a cabo en el núcleo y en el citoplasma de la célula. Estos eventos comprenden: 1) la transcripción para generar una molécula de pre-mRNA, 2) la adición del CAP en el extremo 5' del pre-mRNA, 3) la remoción de intrones del pre-mRNA, 4) la poliadenilación del extremo 3', 5) la exportación del mRNA maduro al citoplasma, 6) la traducción del mRNA a proteína y finalmente 7) la degradación del mRNA (Rambout *et al.*, 2017). La expresión de genes ha revelado un papel esencial entre complejos proteicos que llevan a cabo los diferentes eventos los cuales están estrechamente relacionados (Orphanides y Reinberg, 2002).

## **1.2. Transcripción**

La transcripción es un proceso que consiste en la apertura de la doble hélice del DNA para hacer una copia de un gen en una molécula de RNA denominada mRNA (Orphanides y Reinberg, 2002). La transcripción se lleva a cabo en el núcleo y puede iniciar mediante una señal dentro o fuera de la célula; el material genético se encuentra condensado gracias a las proteínas histonas conformando la cromatina. Para que la transcripción se lleve a cabo, la cromatina expone parte del DNA lo cual es facilitado por la acetilación de los extremos amino terminal de las histonas (Deutschman, 2005). La RNA polimerasa II (RNAPII) es la enzima que cataliza las reacciones de polimerización del mRNA. Esta enzima es dependiente del DNA, y está conformada por 12 subunidades (Liu *et al.*, 2013). En el proceso de transcripción participan diferentes proteínas y consta de tres pasos: inicio, elongación y término de la transcripción.

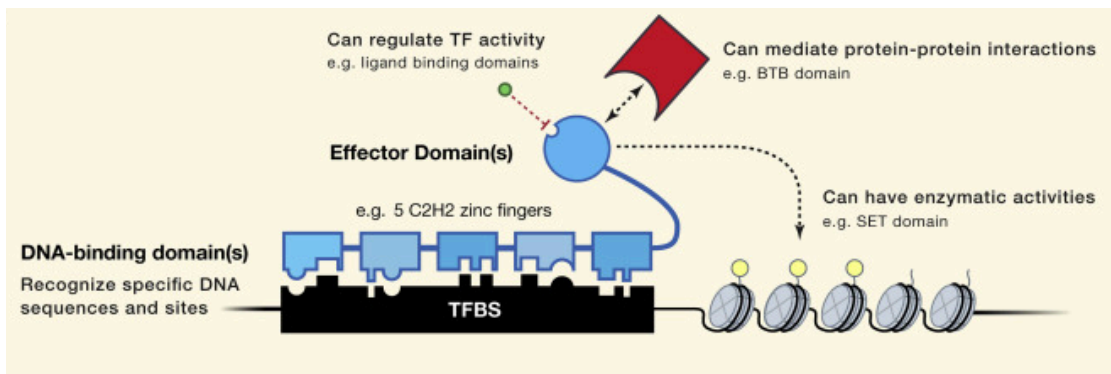
### **1.2.1. Inicio de la transcripción**

La RNAPII es responsable de catalizar la síntesis del mRNA y es dependiente de un complejo de proteínas denominadas factores generales de transcripción (TF, del inglés "Transcription Factor"). Los factores se abrevian anotando con números romanos la polimerasa y con letras el tipo de factor. Éstos se unen a una región del DNA conocida como core del promotor, el cual puede ser rico en timinas y adenosinas. Una de las secuencias más comunes en esta región es la caja TATA

que es reconocida por TFIID, seguida por TFIIB, que funcionan como andamio para la unión de la RNAPII acompañada por TFIIE, TFIIF y TFIIH. El extremo CTD (del inglés “Carboxyl Terminal Domain”) de la RNAPII en mamíferos consta de un repetido de siete aminoácidos (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) que varía en el número de repeticiones dependiendo del organismo. Durante el inicio de la transcripción las serinas y tirosinas son fosforiladas por el factor TFIIH (Schier y Taatjes, 2020; Proudfoot, 2016).

Los FT se unen a secuencias de DNA situadas río arriba del sitio de inicio de la transcripción. Estas secuencias de DNA se conocen como elementos regulatorios y están localizados río arriba del sitio de inicio de la transcripción (dependiendo del organismo y el gen) (Orphanides y Reinberg, 2002). Los FT comúnmente tienen dos regiones o dominios. Una de estas regiones denominada dominio de unión a DNA (DBD, del inglés DNA Binding Domain) reconoce una secuencia específica de DNA de doble cadena (secuencia consenso) mediante uniones covalentes, iónicas y puentes de hidrógeno con el DNA. Los FT comúnmente tienen  $\alpha$ -hélices y hojas  $\beta$ -plegadas que conforman diferentes arquitecturas, Por ejemplo, tienen una estructura de hélice-vuelta-hélice, dedos de zinc, horquilla, cremallera de leucina y hélice bucle hélice. La segunda región es el motivo de activación o efector, esta región favorece el ensamblaje de las proteínas en la región del promotor lo cual estabiliza la interacción de la RNAPII o bien permite el reclutamiento de enzimas modificadoras de cromatina (Deutschman, 2005). Los dominios efectores regulan la actividad del FT al interactuar con sus ligandos, facilitando las interacciones proteína-proteína, o tienen actividad enzimática, como es el caso de los dominios SET (*Su(var)3-9*, *Enhancer-of-zeste and Trithorax*) en proteínas que modulan la organización de la cromatina y, por ende, controlan de forma transcripcional los estados activos o reprimidos.

La unión de los FT al DNA facilita el reclutamiento de la RNAPII para promover o reprimir la transcripción (Lee y Young, 2000). Además del DBD además los FT pueden presentar otros dominios que ayudan a estabilizar o inhibir el ensamblaje de la maquinaria de transcripción. En algunos casos también presentan dominios de activación que pueden ser modificados en diferentes condiciones y cambiar el papel que el FT desempeña en diferentes contextos (Figura 1) (Latchman, 1997).



**Figura 1. Esquema de un factor de transcripción prototipo.** El esquema muestra los dominios activadores o efectores que posee un FT canónico, así como el dominio de unión a DNA. En azul se muestra el dominio de unión al DNA y el dominio efector. Imagen adaptada de Lambert *et al.*, 2018.

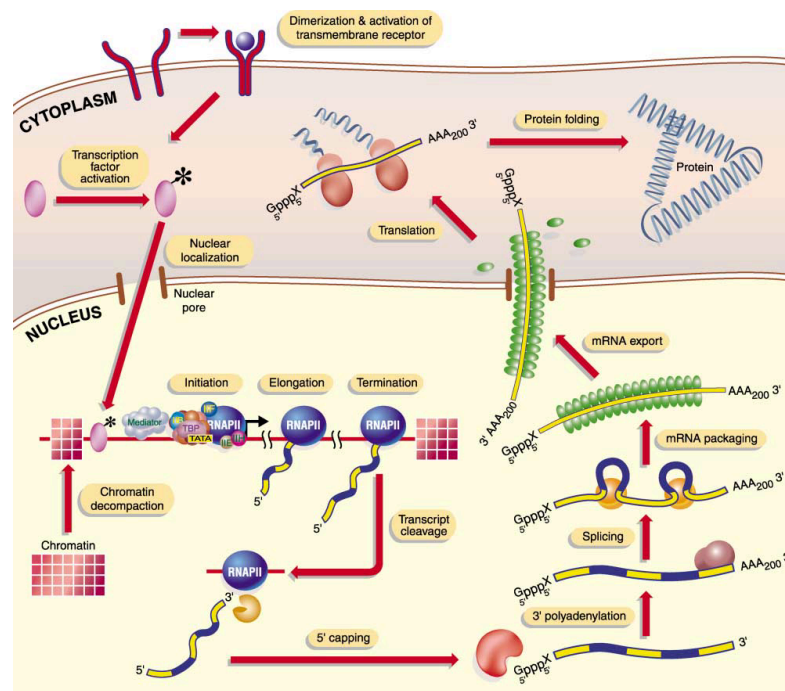
### 1.2.2 Elongación y término de la transcripción

La etapa de elongación comienza cuando al transcrito de una longitud aproximada de 25 nucleótidos, se le adiciona el CAP (residuo de 7-metilguanosa) en el extremo 5' del RNA provocando la liberación de varios FT lo que permite la unión de factores de elongación al extremo CTD de la RNAPII. Estas interacciones coordinan a las proteínas relacionadas en los procesos postranscripcionales del mRNA (Liu *et al.*, 2013). La función de los factores de elongación es reprimir la pausa de la transcripción generada por el complejo RNAPII/TFIIF, por ejemplo, el factor de elongación positivo b (P-TEFb) regula la tasa de transcripción de genes al fosforilar el extremo CTD de la RNAPII (Phatnani *et al.*, 2006).

Para terminar la transcripción, existen secuencias conservadas 10 a 30 nucleótidos del sitio de escisión en el extremo 5', y otra secuencia rica en residuos G y U, de 20 a 40 nucleótidos del sitio de escisión en el extremo 3'. La escisión genera un 3'-hidroxilo libre marcando el final del transcrito, al cual se le adicionan adeninas mediante la poliadenilato polimerasa, el transcrito se escinde en la señal poli(A) para liberar el mRNA (Hu *et al.*, 2005, Proudfoot, 2016). Una vez que se ha completado la transcripción, la RNAPII es desfosforilada y puede iniciar otro ciclo de transcripción (Figura 2).

Otro elemento importante en la transcripción son los *enhancers* o potenciadores, los cuales son secuencias de DNA que se encuentran a varios kilobases de distancia del sitio de inicio de la transcripción (INR) del gen y afectan la tasa de transcripción. Un solo *enhancer* puede afectar la transcripción de múltiples genes (Krivega y Dean, 2012). Estos *enhancers* son modulados por los *insulators* que son elementos de la secuencia de DNA que previenen interacciones inapropiadas

entre dominios de cromatina adyacentes. Un tipo de *insulators* establece dominios que separan *enhancers* y promotores para bloquear su interacción, mientras que un segundo tipo crea una barrera para evitar el efecto de la heterocromatina (Gazner y Felsenfeld, 2006). Cabe mencionar que cada paso que regula la expresión génica (desde la transcripción hasta la traducción) es un proceso continuo, ya que las etapas están conectadas física y funcionalmente con la siguiente, lo que garantiza que haya una interacción eficiente, a lo que se conoce como regulación cotranscripcional. Estudios sugieren que pasos consecutivos como la transcripción y el *splicing* son vías interdependientes o son influenciadas una por otra (Orphanides *et al.*, 2002).



**Figura 2. Proceso de la expresión génica en células eucariotas.** Se muestran los diferentes eventos para que se lleve a cabo la expresión génica desde la activación de reguladores transcripcionales hasta la síntesis de una proteína funcional. Los FT se unen a elementos reguladores de genes, promueven el acceso al DNA y facilitan el reclutamiento de la enzima RNAPII al INR. La transcripción de genes codificadores de proteínas está catalizada por RNAPII. La RNAPII inicia la transcripción, del RNA naciente se modifica mediante la adición de una estructura de "CAP" en su extremo 5', tiene un papel como sitio de unión para las proteínas involucradas en la exportación del mRNA maduro al citoplasma. La RNAPII comienza la síntesis de RNA, y la "elongación" de la transcripción, en la que la polimerasa se mueve de 3' a 5' a lo largo de la secuencia del gen (sobre la cadena de DNA molde) y extiende la transcripción en dirección 5' a 3'. La fase de elongación de la transcripción está sujeta a regulación por una familia de "factores de elongación". Los intrones se eliminan mediante splicing del mRNA. Al llegar al final de un gen, RNAPII detiene la transcripción ("terminación"), el RNA recién sintetizado se escinde ("escisión") y se agrega una cola de poliadenosina [poli(A)] al extremo 3' de la transcripción "poliadenilación". El mRNA procesado se transporta desde el núcleo al citoplasma. Imagen tomada de Orphanides *et al.*, 2002.

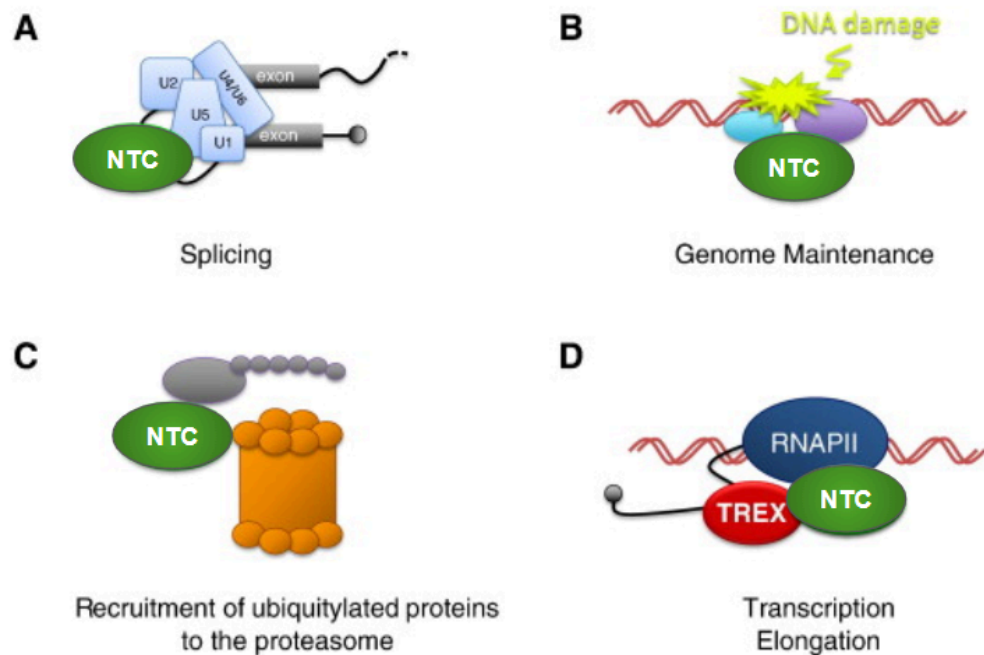
### 1.3. *Splicing*

El *splicing* es un proceso por el cual se remueven intrones en el pre-mRNA y es ejecutado por el spliceosoma a través de dos pasos de transesterificación (Nelson y Cox, 2017). El *splicing* implica además un paso adicional en la regulación de la expresión génica, ya que a partir de un solo gen se pueden obtener diferentes proteínas por un proceso conocido como *splicing* alternativo (Rambout *et al.*, 2017).

El spliceosoma es un complejo ribonucleoproteico que consiste en 5 ribonucleoproteínas nucleares (snRNP), U1, U2, U4, U5 y U6. Cada snRNP está formada por varias proteínas y una molécula de RNA rica en uracilos (Shukla y Oberdoerffer, 2012). En algunas células eucariotas coexisten dos tipos de spliceosoma, uno dependiente de U2 (spliceosoma mayor) y otro menos abundante dependiente de U12 (spliceosoma menor) (Patel y Steitz, 2003). El procesamiento de exones en el mRNA ocurre en dirección 5' - 3' durante el cual U1 reconoce al dinucleótido GU del extremo 5' del intrón, en tanto que los factores SF1, U2AF2 y U2AF1 reconocen el sitio de ramificación (típicamente 18-40 nt río arriba del sitio de *splicing* 3', 3' ss), el tracto de pirimidinas subsecuente y la secuencia consenso AG del extremo 3' respectivamente (complejo E). La Figura 3 ejemplifica un comparativo de las secuencias consenso que presentan los intrones de eucariontes superiores y levaduras y que son reconocidas por el spliceosoma. U2 snRNP se asocia al sitio de ramificación (complejo A) y se une el complejo U4/U6U5 tri-snRNP (precomplejo B). Debido al reacomodo y las interacciones, se desestabilizan los snRNP U1 y U4, dando lugar al spliceosoma activado (complejo B activo). Posteriormente, la actividad catalítica de la RNA helicasa Prp2 cataliza el primer paso de *splicing*; esto a su vez produce el complejo C, el cual cataliza el segundo paso de *splicing*, después el complejo se disocia para participar nuevamente en el *splicing* de otro transcrito (Figura 4) (Will y Lührmann, 2011).

La activación del *spliceosoma* es un proceso coordinado y dinámico que requiere el ensamblaje de las snRNP y otras proteínas y factores sobre el pre-mRNA para la eliminación de las secuencias no codificantes. Un ejemplo de este dinamismo se observa en la formación del complejo Prp19 o NTC (del inglés nineteen complex).





**Figura 5. Funciones del complejo NTC.** A) Complejo de *splicing*. B) mantenimiento del genoma, C) degradación de proteínas por el proteosoma y D) elongación de la transcripción. Tomado de Chanarat *et al.*, 2013.

Este complejo está compuesto por cuatro subunidades centrales, incluidas Prp19, CDC5, PRL1 y SPF27 (Grote *et al.*, 2010) y estabiliza la formación del spliceosoma. En *Saccharomyces cerevisiae*, este complejo está conformado por 19 proteínas y en eucariotes superiores incluyendo animales y plantas por más de 30 proteínas (Will y Lührmann, 2011; Koncz *et al.*, 2012). En *S. cerevisiae*, el NTC consta de ocho subunidades centrales y al menos hasta 19 proteínas asociadas en total. En humanos se han aislado al menos tres complejos proteicos diferentes (Tabla 1). El complejo Prp19/CDC5 se asocia con otras proteínas, incluidas CTNNBL1, AD-002/CWC15 y PLRG1, lo cual mejora su función en el *splicing* y la respuesta al daño del DNA (Chanarat *et al.*, 2013; Grote *et al.*, 2010). Durante la activación catalítica del spliceosoma, la interacción entre U4/U6 se interrumpe, lo que conduce a la hibridación del snRNA U6 con el pre-mRNA y el snRNA U2. Se ha demostrado en levadura que NTC es esencial para una interacción estable de los snRNP U5 y U6 con el spliceosoma después de la activación (Chan *et al.*, 2003).

**Tabla 1. Composición del complejo NTC**

Organismo	<i>S. cerevisiae</i>	<i>H. sapiens</i>
Complejo	NTC	PRP19/CDC5L
Función conocida	<i>Splicing</i> , transcripción, Reparación de DNA	<i>Splicing</i>
Componentes	Prp19/Pso4E	PRP19/PSO4/SNEV
	Syf1/Ntc90E	–
	Isy1/Ntc30	–
	Cef1/Ntc85E	CDC5L
	Prp46/Ntc50E	PRL1/PRP46/CWC1
	Snt309/Ntc25	SPF27/SNT309/DAM1
	Cwc15	AD002/CWF15
	Clf1/Ntc77/Syf3E	–
	Syf2/Ntc31	–
	Cwc2/Ntc40E	–
	Ntc20	–
	Yju2E	–
	–	CTNNB1
		HSP73

Tomado de Chanarat *et al.*, 2013.

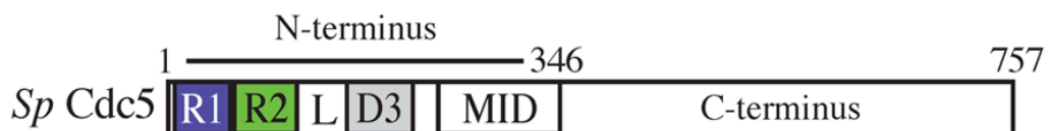
Un estudio sugiere que el NTC puede funcionar como una ubiquitina ligasa E3 que regula la elongación de la transcripción al ubiquitinar a la RNAPII y promover su degradación *in vitro* (Hatakeyama *et al.*, 2001; Ohi *et al.*, 2003). Prp19 ubiquitina a Prp3, un componente del snRNP U4 (Nottrott *et al.* 2002); esta modificación de Prp3 aumenta la afinidad de Prp3 por el componente U5/Prp8 necesario para estabilizar U4/U6U5 snRNP. Además, el NTC puede regular la elongación de la transcripción al promover el reclutamiento de TREX en los genes transcritos que es un complejo proteico que se acopla a la transcripción y la exportación de mRNA en *S. cerevisiae*. Uno de los componentes esenciales del NTC es la proteína de la familia Myb, una proteína con el dominio DBD-MYB denominada CDC5 (Chanarat *et al.*, 2013).

### 1.3.2. CDC5

La proteína CDC5, también conocida como Ciclo de División Celular 5, es una proteína multifacética que se encuentra conservada en varios organismos, incluidos plantas, animales y hongos (Ohi *et al.*, 1998). En *Arabidopsis thaliana*,

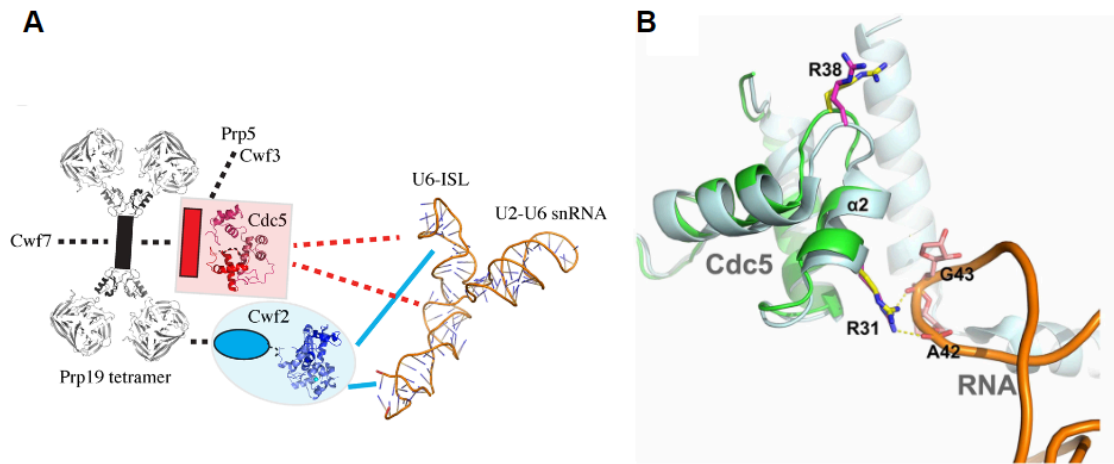
CDC5 es un componente del complejo MAC (MOS4-Associated Complex) ortólogo a NTC, que probablemente regula las respuestas de defensa a través del control transcripcional y es esencial para la inmunidad innata de la planta (Palma *et al.*, 2007). Además, CDC5 es una proteína de unión al DNA que regula positivamente el procesamiento postranscripcional y/o la transcripción de microRNA primarios en *A. thaliana*, afectando la acumulación de microRNAs que controlan diversos procesos biológicos, incluido el desarrollo y las adaptaciones ambientales en las plantas (Zhang *et al.*, 2013).

Como se mencionó anteriormente, CDC5 es miembro del NTC y participa en una variedad de actividades, incluido el *splicing* del mRNA, la biosíntesis de lípidos, la reparación del DNA (Chanarat *et al.*, 2013) y es esencial para la progresión de G<sub>2</sub>/M en la mitosis (Gonzalez-Arranz *et al.*, 2021). El extremo N-terminal de CDC5 está altamente conservado y contiene dos repetidos canónicos de Myb (R1 y R2) y un tercer dominio imperfecto o incompleto denominado repetido similar a Myb (D3) (Ohi *et al.*, 1998) (Figura 6). En *A. thaliana* estos repetidos se unen preferentemente a la secuencia de DNA [CTCAGCG] (Hirayama y Shinozaki, 1996).



**Figura 6. Diagrama de la proteína CDC5 en *Schizosaccharomyces pombe*.** Arquitectura del dominio de CDC5 de *S. pombe*. El N terminal de CDC5 (aa 1–346) contiene dos repetidos Myb canónicos (R1, R2) y un tercer repetido denominado D3 que está predicho como un repetido Myb-like. R2 y D3 están separados por una región predicha en forma de bucle (L, Loop). La región media de CDC5 (MID) es predicha por PSI-PRED para ser una  $\alpha$ -hélice. La posición de estos dominios es la siguiente: R1 (aa 5–55), R2 (aa 58–111), L (aa 111–146), D3 (aa 155–214), y MID (aa 237–346). El C-terminal de CDC5 (aa 347–757) no tiene motivos predichos. Imagen tomada de Collier *et al.*, 2014.

En un estudio realizado en *Schizosaccharomyces pombe* se demostró que el extremo amino de CDC5 también se une al RNA, concluyendo que CDC5 en *S. pombe* desempeña un papel crucial en el *splicing* del pre-mRNA. A pesar de que el dominio CDC5-D3 no adopta una conformación similar a la de los otros repetidos Myb, se une preferentemente al RNA bicatenario a través de su extremo N-terminal por lo que CDC5, interactúa tanto con las proteínas del NTC como con el RNA, funcionando como un andamio importante que facilita los cambios conformacionales para la formación de un spliceosoma catalíticamente activo (Collier *et al.*, 2014) (Figura 7).



**Figura 7. Modelo del NTC y CDC5 en *S. pombe*.** A) Modelo de interacción de Prp19/CDC5 con snRNA. El tetrámero Prp19 se une directamente a los componentes de NTC CDC5, Cwf7 y Cwf2 (líneas sólidas), enlaces cruzados con regiones de los snRNA U2 y U6. El extremo C de CDC5 se une directamente a Prp5 y Cwf3, mientras que el extremo N se une al RNA *in vitro*. Las líneas discontinuas indican un estudio directo *in vitro*. Las líneas discontinuas indican la interacción directa entre componentes *in vitro* y las líneas continuas representan entrecruzamiento proteína-RNA *in vitro*. Imagen de Collier *et al.*, 2014. B) Dos residuos de aminoácidos específicos de MoCDC5 (*Magnaporthe oryzae*), R31 y R38, se encuentran en la periferia del surco cargado positivamente. Los residuos se muestran en un modelo de barra mediante superposición estructural de MoCDC5-DBD (verde) y levadura CDC5 (plata). El residuo R31 de MoCDC5-DBD interactúa con A42 y G43 del RNA U6. Imagen tomada de Wang *et al.*, 2019.

### 1.3.3. Retención Intrónica (RI)

Una forma de *splicing* alternativo es la Retención Intrónica (RI), que como el nombre indica, resulta de la retención de un intrón en el mRNA maduro en lugar de su eliminación por la vía canónica (Mancini, 2016). La RI se ha observado como un mecanismo de regulación génica, dando lugar a diferentes isoformas, lo que puede generar una proteína trunca no funcional o una proteína nueva funcional (Ner-Gaon *et al.*, 2004). La RI también está asociada con el desarrollo de enfermedades complejas. Por ejemplo, muchos intrones que se retienen en cánceres primarios, se pueden detectar en el citoplasma de las células cancerosas (Dvinge y Bradley, 2015).

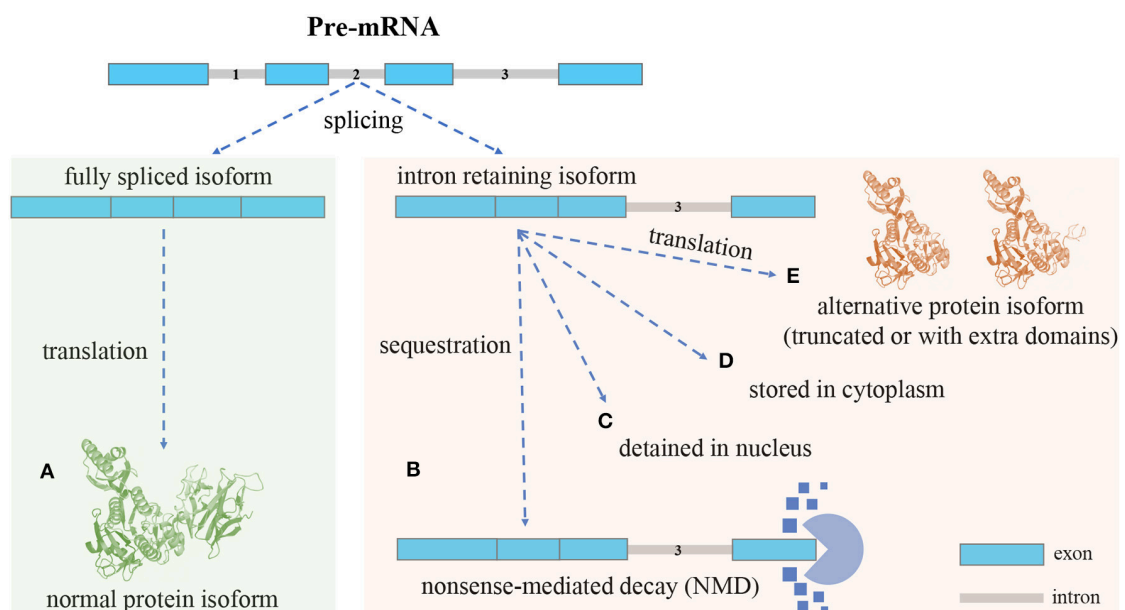
El destino de un mRNA con RI depende de varios factores: la ubicación del intrón retenido dentro del transcrito; retención y degradación nuclear; retención y almacenamiento nuclear en espera de *splicing* inducido por señales, entre otras. La RI en el extremo UTR 5' puede introducir un marco de lectura abierto (Open Reading Frame, ORF) ascendente u otras características estructurales que pueden activar o reprimir la eficiencia de iniciación de la traducción. Un intrón retenido en el ORF principal puede originar codones de paro prematuros que

conduzcan a la degradación vía NMD (por sus siglas en inglés Nonsense-Mediated Decay) o mantener el marco de lectura permitiendo la producción de isoformas de proteínas con dominios adicionales. Los mensajeros con RI pueden también escapar de la vía NMD (Lykke-Andersen y Jensen, 2015) y traducirse a isoformas de proteínas truncas (Lindeboom *et al.*, 2016; Ottens y Gehring, 2016). Se ha demostrado que aproximadamente el 10 % de genes con RI en humanos se traducen en proteínas truncas (de Lima Morais y Harrison, 2010). Las isoformas de proteínas generadas a través de mRNA con RI pueden ser más cortas (tener menos dominios) o incluir más dominios que la proteína normal (Gontijo *et al.*, 2011; Rekosh y Hammarskjold, 2018) (Figura 8). Si el intrón está a más de ~55 nt del extremo UTR 3', la RI puede estabilizar el mRNA evitando la NMD. Un intrón retenido en el extremo UTR 3' puede introducir elementos en *cis* que afectan la estabilidad o la eficiencia traduccional del mRNA (Jacob y Smith, 2017).

Jacob y Smith (2017) clasificaron a la RI en tres grupos que varían según: el porcentaje de retención intrónica (% RI), el contenido de GC, la longitud del intrón, la historia evolutiva y sus efectos sobre el ORF. La clase A es el grupo más abundante, muestra un % RI bajo, un contenido de GC intermedio e intrones largos que derivan de intrones ancestrales. La clase B o exitrones ocurre dentro de exones anotados, tienen un alto % RI y GC, son cortos y parecen derivar de la intronización de exones ancestrales (Braunschweig *et al.*, 2014; Márquez *et al.*, 2015). Estos intrones tienen características tanto de intrones como de exones. Los exitrones encontrados en *Arabidopsis* y humanos tienen tamaños en múltiplos de tres nucleótidos. El *splicing* de estos exitrones da como resultado proteínas truncas y afecta ampliamente la función de las proteínas al modificar dominios proteicos, regiones desordenadas y sitios de modificación postraduccional. Finalmente, la clase C se caracteriza por estar adyacente a exones de cassette anotados, tienen un % RI intermedio, con % de GC bajo e intrones largos que sugieren que los eventos de *splicing* alternativo se llevan a cabo más lentamente (postranscripcionalmente) (Pandya-Jones *et al.*, 2013; Pandya-Jones y Black, 2009). Las tres clases de RI tienen sitios de *splicing* más débiles que los intrones constitutivos. Los sitios de *splicing* entre intrones y exones dependen de varias características como la longitud del exón y del intrón y también su contenido relativo de GC (Amit *et al.*, 2012; Berget, 1995). Intrones cortos con un alto contenido de GC son reconocidos como una unidad, mientras que exones cortos

rodeados por intrones largos con bajo contenido de GC tienden a ser reconocidos como un exón. La RI puede ser detectada por medio de la secuenciación del transcriptoma (RNAseq) ya que permite determinar las isoformas de un transcrito además del nivel de expresión (Grabski *et al.*, 2020).

La longitud media del extremo 3' UTR de los genes con RI es significativamente más larga que la de los genes que no retienen intrones en todos los vertebrados, y la diferencia aumenta en especies que están evolutivamente más cerca de los humanos (Schmitz *et al.*, 2017).



**Figura 8. Mecanismo de retención intrónica (RI) producido por *splicing* alternativo.** A) Isoformas con intrones completamente empalmados se envían fuera del núcleo para su traducción. B) Isoformas con RI degradadas por la vía de NMD. C) Arresto en núcleo que, en respuesta a estímulos, pueden sufrir más *splicing* para eliminar el intrón retenido, antes de ser exportado fuera del núcleo para su traducción. D) Isoforma con RI transportada al citoplasma para preservación. E) escape de la vía NMD y traducción a proteínas truncadas, o con dominios adicionales formados por las secuencias de aminoácidos traducidas de intrones retenidos (con RI). Imagen de Zheng *et al.*, 2020.

#### 1.4. Regulación cotranscripcional

La regulación cotranscripcional es un evento donde el *splicing* se produce mientras el mRNA es sintetizado por la RNAPII. Este proceso está funcionalmente acoplado con la elongación de la transcripción y ambos pueden influir entre sí (Shenasa y Bentley, 2023).

Una de las primeras evidencias de la regulación cotranscripcional se determinó en micrografías electrónicas en embriones de *Drosophila melanogaster* donde se observó la formación de estructuras en forma de bucles (lariat) asociados a complejos de ribonucleoproteínas. Esto sugirió que el *splicing* ocurre antes de la

escisión del transcrito (Beyer *et al.*, 1981).

El *splicing* y la elongación durante la regulación cotranscripcional pueden ser influenciados por los promotores y por activadores o represores, ya que, en algunos ensayos, al intercambiar promotores, se seleccionó un sitio diferente de *splicing* (Cramer *et al.*, 1997). Huang y colaboradores (2012) demostraron por medio de cromatografía de afinidad y espectrometría de masas que el regulador cotranscripcional MED23 (subunidad del complejo Mediador) se une a factores de procesamiento de mRNA, entre ellos hnRNP (ribonucleoproteínas nucleares heterogéneas).

Además, se ha observado que la cinética de elongación de la transcripción también afecta el *splicing*, cuanto más lenta es la elongación, más se permite el reclutamiento de factores del *splicing*, mientras que una elongación más rápida reduce la posibilidad de reconocer un sitio de *splicing* proximal débil al promotor (Zhu *et al.*, 2020). En *S. cerevisiae* se observó que el *splicing* se produce de forma cotranscripcional para la mayoría de los genes que contienen intrones, ya que la RNAPII hace una pausa dentro de los exones terminales de estos genes. Los genes sin intrones y empalmados de manera ineficiente carecen de esta pausa que proporciona tiempo suficiente para el *splicing* antes de la terminación. El descubrimiento de la pausa del exón terminal demuestra el acoplamiento funcional de la transcripción y el *splicing* cerca de los extremos de los genes (Oesterreich *et al.*, 2010).

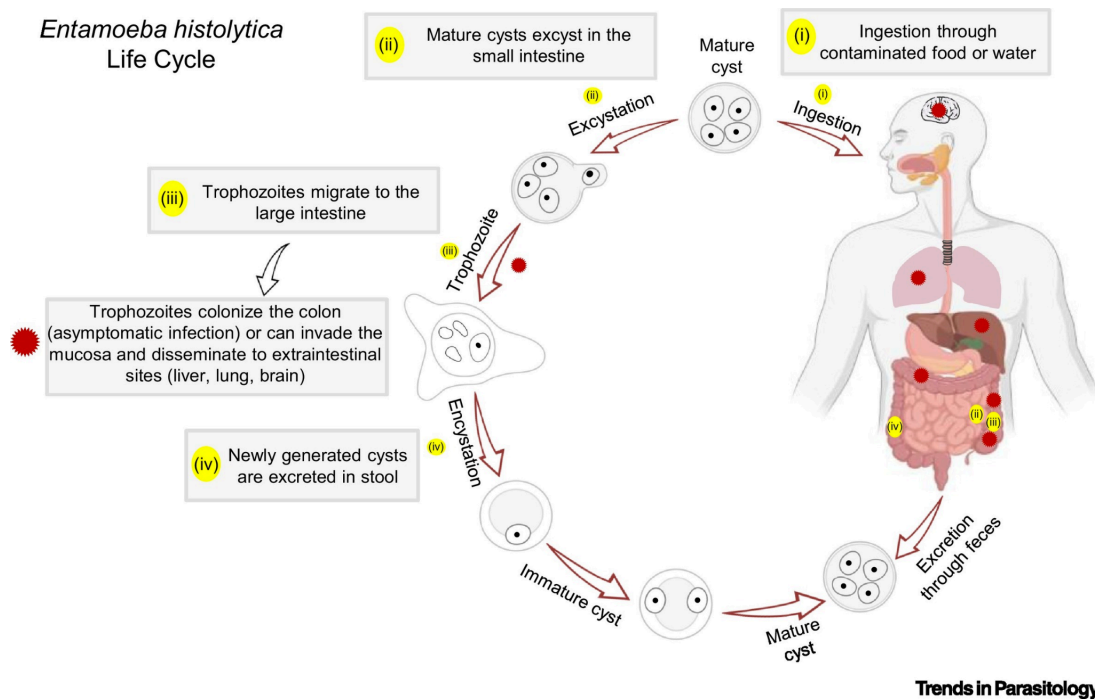
Asimismo, a través de ensayos de hibridación de RNA *in situ* y de inmunofluorescencia, se observó que los factores de *splicing*, el mRNA y sus *loci* genómicos colocalizan, lo cual sugiere que los mecanismos de transcripción y *splicing* están acoplados y que el *splicing* del RNA ocurre antes de que la transcripción finalice (Shukla y Oberdoerffer, 2012). Adicionalmente, por estudios de inmunofluorescencia, se observó que los factores de *splicing* se localizan en los sitios de transcripción o fábricas de transcripción en el núcleo, donde se localizan genes que contienen intrones y al inmunoprecipitar a factores de transcripción, estos a su vez se encuentran asociados a proteínas del spliceosoma (Rambout *et al.*, 2017). Por otro lado, en un estudio se demostró que los promotores tienen un efecto en la selección de los sitios de *splicing* y que esto puede ser consecuencia de las interacciones entre las maquinarias de transcripción y *splicing* (Schwartz *et al.*, 2009). Finalmente, la regulación

cotranscripcional se ha demostrado en diferentes organismos y ha mostrado un papel importante en la coordinación de la transcripción y el *splicing* constitutivo y alternativo (Merkhofer *et al.*, 2014).

### **1.5. *Entamoeba histolytica*: enfermedad, epidemiología, ciclo de vida**

El género *Entamoeba* contiene muchas especies, seis de las cuales ( *Entamoeba histolytica*, *Entamoeba dispar*, *Entamoeba moshkovskii*, *Entamoeba polecki*, *Entamoeba coli* y *Entamoeba hartmanni* ) residen en la luz intestinal humana (Fotedar *et al.*, 2007). *E. histolytica* está asociada con secuelas patológicas en humanos, es responsable de la amebiasis, puede causar diarrea severa y es la tercera causa de muerte por infecciones de parásitos con números que ascienden a 100,000 muertes anuales. Casi el 90 % de las infecciones de *E. histolytica* son asintomáticas, sin embargo, cerca de 50 millones de personas presentan síntomas (Shirley *et al.*, 2018).

*E. histolytica* no requiere de un vector para su transmisión y la infección inicia con la ingesta de quistes (forma infectiva no móvil y resistente) a través de agua o alimentos contaminados. Estos se desenquistan en el íleon terminal, en la forma invasiva o trofozoítos móviles. Los trofozoítos proliferan debido a que se dividen rápidamente por fisión binaria y se adhieren a la superficie de la mucosa e invaden el intestino grueso, causando diarrea y colitis. La mayoría de estas infecciones son enfermedades autolimitadas, y bajo ciertas condiciones las amebas se diferencian nuevamente a quistes que son expulsados con las heces, reiniciando así su ciclo y provocando una nueva infección en otros huéspedes (Carrero *et al.*, 2020; Watanabe y Petri, 2015). En algunos casos, los trofozoítos penetran y atraviesan el epitelio intestinal, donde pueden diseminarse a otros órganos, más comúnmente al hígado, donde forman abscesos hepáticos (Wilson *et al.*, 2012) (Figura 9).



**Figura 9. Ciclo de vida de *E. histolytica*.** i) ingestión de alimentos o agua contaminados, ii) desenquistamiento en el intestino delgado, iii) migración de los trofozoítos (colonización en el colon), iv) generación de quistes en y excreción en heces. Tomado de Begum *et al.*, 2021.

En un estudio realizado en una comunidad rural del estado de Morelos, México, se encontró una alta tasa de prevalencia de infección asintomática por *E. histolytica*, ya que la frecuencia de infección fue del 27.5 % y por *Entamoeba dispar* fue del 23.3 % (Ramos *et al.*, 2005). En 2015, se determinó el porcentaje de infección por *E. histolytica* en la población general de la zona rural de Durango, México, y se encontró que la seroprevalencia era del 4,5 % (Alvarado-Esquivel *et al.*, 2015). Recientemente, se determinó que la frecuencia de infección por *E. histolytica* y *E. dispar* en escolares de la comunidad de la Ciudad de México era del 2.3 % y del 7.7 % respectivamente (Rojas *et al.*, 2016).

Entre los factores de virulencia que son responsables de la patogenicidad de este parásito se encuentra la lectina Gal/GalNac que es un complejo proteico de múltiples subunidades implicado en la adhesión a las células huésped. Se han identificado 2 genes que codifican subunidades ligeras: *IgI2* y *IgI3*, y que se expresan diferencialmente durante la infección intestinal (Gilchrist *et al.*, 2006). Otras moléculas relacionadas con la unión con las células del huésped son las proteínas ricas en serina, treonina e isoleucina (EhSTIRP) (MacFarlane y Singh 2007). Dentro del huésped, los trofozoítos pueden ser expuestos al óxido nítrico y especies reactivas de oxígeno (ROS) liberadas por las células de la respuesta inmune. Estos pueden ser contrarrestados por peroxiredoxina, la cual reduce el

H<sub>2</sub>O<sub>2</sub> a H<sub>2</sub>O (Nandi *et al.*, 2010). *E. histolytica* contiene varios genes putativos de peroxiredoxina, además, en un estudio realizado por Davis y col. (2006) encontraron que la peroxiredoxina se expresa más en *E. histolytica* HM-1:IMSS (cepa virulenta) que en Rahman (cepa no virulenta) (Wilson *et al.*, 2012). Otros factores de virulencia importantes son las proteasas de cisteína que degradan la matriz extracelular del huésped. Por ejemplo, CP-A5 es una proteasa que degrada el polímero MUC2 (Lidell *et al.*, 2006) y CP112 desestabiliza las uniones entre célula y célula en el epitelio intestinal facilitando la invasión del parásito (Cuellar *et al.*, 2017). Otras proteasas con mayor expresión en HM-1:IMSS son CP-A4, CP-A6 y CP-B1 (Davis y Stanley, 2007). Finalmente, el trofozoito expresa también péptidos formadores de poros llamados amebaporos, capaces de lisar células diana al formar poros y desestabilizar las membranas celulares (Loftus *et al.*, 2005). En el intestino, *E. histolytica* puede utilizar glicanos del moco intestinal del huésped como fuente de carbohidratos, iniciando la fase invasiva. La expresión de AIG1 puede ser importante en el colon, ya que se cree que está implicada en la defensa antibacteriana (Naiyer *et al.*, 2019).

### 1.6. Genoma de *E. histolytica*

*E. histolytica* es un parásito unicelular cuyo genoma se secuenció, ensambló y analizó en 2005 (Loftus *et al.*, 2005). Sin embargo, el genoma se reanotó y reanalizó en 2010, reportando un tamaño de aproximadamente 20 Mb, con un contenido total de AT del 77.6 % y 24.2 % de GC.

El genoma de *E. histolytica* está compuesto por aproximadamente 38 cromosomas lineales que varían en tamaño de 0.3 a 1.8 Mbp y posiblemente una ploidía de 4, así como varias moléculas de DNA circulares similares a plásmidos de diferentes tamaños (Bhattacharya *et al.*, 2000; Kawano-Sugaya *et al.*, 2020). En el genoma se encuentran codificados ~8200 genes con un tamaño promedio de 1260 pb. La expresión de los genes es regulada comúnmente a través de una región promotora que se conforma de tres elementos conservados: la caja TATA (GTATTTAAA localizada a -30), un elemento GAAC (AATGAACT localizado entre la caja TATA y el INR) y el INR (AAAAAATTCA) (Singh U *et al.*, 1997; Lorenzi *et al.*, 2010). La transcripción de *E. histolytica* ocurre como en otros organismos superiores por medio de una RNAPII resistente a  $\alpha$ -amanitina (Lioutas y Tannich, 1995). El bloque de homología F de la subunidad más grande de la RNAPII (RPB1) ha sido identificado como el sitio de unión de  $\alpha$ -amanitina. El homólogo

RPB1 de *E. histolytica* también difiere del consenso en esta región. Se han identificado 10 subunidades de la RNA polimerasa en *E. histolytica* (Clark *et al.*, 2007). Por otro lado, la repetición heptapeptídica (TSPTSPS) común a otros RNA eucariotas y el dominio CTD de la RNAPII no están presentes en la proteína de *E. histolytica*, de hecho, el CTD de *E. histolytica* no es similar a ningún otro dominio de RNAPII. Sin embargo, el CTD de la enzima de *E. histolytica* permanece rica en prolina/serina (estos aminoácidos constituyen el 40 % de la secuencia CTD). El CTD de *E. histolytica* también presenta aminoácidos susceptibles de fosforilación: 9 serinas, 3 treoninas y 1 tirosina. A pesar de su divergencia, la modificación del extremo CTD por cinasas y fosfatasas podría modular las interacciones proteína-proteína como ocurre en otras RNA polimerasas (Yeo *et al.*, 2003).

En cuanto al contenido de intrones, el 24 % de los genes contienen intrones y el 6 % de estos genes contienen más de 2 intrones (Pearson y Singh, 2010). Interesantemente, se observó la presencia conservada de proteínas que conforman la maquinaria del *splicing* en *E. histolytica* y por ensayos funcionales, se confirmó su participación en este mecanismo (Valdés *et al.*, 2014).

**Tabla 2. Características del genoma de *E. histolytica*.**

<b>Genoma</b>	
Tamaño del genoma (Mpb)	20
Contenido de G+C (%)	24.2
Número de genes	8201
Tamaño promedio de los genes (pb)	1260.9
% de DNA codificante	49.2
% de genes con intrones	24.4
Tamaño promedio de las proteínas (aa)	389
Densidad génica (kb por gen)	1.9
Número de cromosomas	38
Número de proteínas	8734
<b>Exones</b>	
Número de exones	10,754
Promedio de exones por gen	1.3
Contenido de G+C (%)	28
Longitud promedio (pb)	962
Longitud total (Mpb)	10.34
<b>Intrones</b>	
Número de intrones	2,553
Promedio de número de intrones/gene	1.5
Contenido de G+C (%)	19.3
Longitud promedio (pb)	74.1
Longitud total (pb)	189,260
Tamaño promedio de intrones (pb)	102.1
<b>Regiones intergénicas</b>	
Contenido de G+C (%)	20.5

Longitud promedio (pb)	708.7
Promedio de distancia intergénica (kb)	0.8

Elaborada a partir de Lorenzi *et al.*, 2010, Kawano-Sugaya *et al.*, 2020

## 2. Antecedentes particulares

### 2.1. Factores de transcripción en *E. histolytica*

*E. histolytica* regula transcripcionalmente sus genes en respuesta a múltiples estímulos como el choque térmico, resistencia a drogas, estrés oxidativo e invasión al hospedero (Ehrenkaufer *et al.*, 2007). A través del análisis de transcriptomas en estas condiciones, se identificaron genes regulados diferencialmente que contienen dominios de unión a DNA y que sugieren su actividad como FT. Entre estos dominios se identificaron proteínas con un dominio Dopey (de la familia de FT con zipper de leucina), y proteínas con dominio MYB que en otros protozoarios como *Giardia lamblia* son particularmente importantes en el proceso de enquistamiento (Sun *et al.*, 2002; Huang *et al.*, 1996). Se han identificado y caracterizado varios FT en este parásito, incluida la proteína de unión a la caja TATA (EhTBP) (de Dios-Bravo *et al.*, 2005, Luna-Arias *et al.*, 1999) y proteínas relacionadas a TBP (EhTRF) (Castañon-Sanchez *et al.*, 2010), EhEBP1, EhEBP2 (Schaenman, *et al.* 2001), EhHMG (Abhyankar *et al.*, 2008), EhHRM-BP, EhURE-BP (Gilchrist *et al.*, 2001), EhMyb (Ehrenkaufer *et al.*, 2009), STAT (Marchat *et al.*, 2002), GATA (Huerta *et al.*, 2020), HSF (Iyer *et al.*, 2008) y la proteína similar a p53 (Mendoza *et al.*, 2003). Adicionalmente el factor EhPC4 fue identificado como un factor clave que afecta la ploidía y la estabilidad del genoma en *E. histolytica* (Cruz *et al.*, 2016).

Entre los FT caracterizados destacan:

- EhHSTF7; los factores de choque térmico, bajo estrés térmico se unen entre sí para formar polímeros y activar la expresión de proteínas de choque térmico al reconocer y unirse a secuencias de DNA conservadas, conocidas como motivos en la región promotora de estas proteínas (Bello *et al.*, 2022).
- URE3-BP (proteína de unión al elemento regulador); es un regulador sensible al calcio de dos genes de virulencia de *E. histolytica*, *hgl5* y *hgl4* (Gal/GalNAc lectin y ferredoxin), que participan en la actividad citopática y la fagocitosis. Un análisis genómico de su secuencia reguladora que actúa en *cis* ha identificado que la transcripción regulada por URE3-BP puede desempeñar un

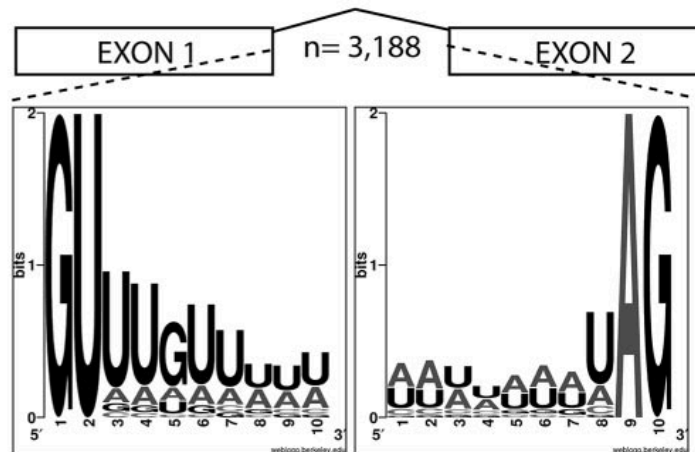
papel en la motilidad de los trofozoítos. Además, los genes que son regulados por el FT URE3-BP son genes que desempeñan un papel en la actividad citopática y la fagocitosis de los trofozoítos (Gilchrist *et al.*, 2008).

- TBP y TRF1; son proteínas de unión a la caja TATA que desempeñan un papel crucial en la regulación de la expresión génica. TBP y TRF1 se unen a la caja canónica [TATTTAAA], así como a diferentes variantes de TATA. Estos FT exhiben una expresión génica diferencial bajo condiciones de estrés, afectan la expresión génica en *E. histolytica* uniéndose a secuencias de DNA específicas y regulando la transcripción de genes en respuesta a diferentes estímulos y durante la interacción con células de mamíferos (Narayanasamy *et al.*, 2018).

## **2.2. Splicing en *E. histolytica***

Los intrones en los genes de *E. histolytica* poseen sitios de *splicing* (ss) 5' (GUUUGU) y 3' (AG) conservados (Figura 10), pero las secuencias de ramificación carecen de conservación. Estos sitios son necesarios para ensamblar el spliceosoma y conformar los sitios catalíticos responsables de la escisión de un intrón y la unión de los exones flanqueantes. En *E. histolytica* no se ha encontrado a U1 snRNP (Dávila *et al.*, 2008), sin embargo, mediante análisis bioinformáticos se encontró la presencia de tres factores U1 snRNP U1-A, U1-C y U1-70k lo que sugiere que la activación de los 5'ss podría deberse a la interacción directa de las proteínas snRNP o por el U6 snRNA-5'ss como en otros sistemas (Kandels-Lewis and Seraphin, 1993; Förch *et al.*, 2002; Rhode *et al.*, 2006; Huang *et al.*, 2012). Posteriormente, las subunidades pequeñas y grandes del factor auxiliar U2, U2AF, interactúan con los 3'ss formando el complejo E spliceosomal. En el siguiente paso, la snRNP U2 interactúa de manera estable con el sitio de ramificación del pre-mRNA, lo que lleva al complejo A o pre-spliceosoma. El complejo A interactúa con la partícula preformada U4/U6-U5 tri-snRNP dando lugar al complejo B precatalítico, que sufrirá reordenamientos de RNPs, incluida la pérdida de U1 y U4 snRNP, convirtiéndose así en un complejo B activado que llevará a cabo el primer ataque nucleofílico del *splicing* generando el intermediario intrón lariat exón 3' y el exón 5' escindido. Finalmente, se forma el complejo C que cataliza el segundo paso de corte y *splicing*, eliminando el intrón y ligando los exones 5' y 3', produciendo así un mRNA maduro listo para la traducción. Las interacciones RNA-RNA son principalmente responsables de la remodelación del spliceosoma y la conformación del núcleo catalítico, sin embargo, se requieren numerosas

proteínas para el ensamblaje, la catálisis y fidelidad del *splicing* (Valdés *et al.*, 2014).



**Figura 10. Conservación de sitios de *splicing* en *E. histolytica*.** Gráfico de secuencia del LOGO que ilustra la frecuencia relativa (bits) de uso de nucleótidos para los primeros 10 nucleótidos (sitio de *splicing* del extremo 5') y los últimos 10 nucleótidos (sitio de *splicing* del extremo 3') para cada uno de los intrones predichos. Imagen de Davis *et al.*, 2007.

Por ejemplo, se ha observado que algunas RNA helicasas están implicadas en la corrección de los pasos secuenciales del ensamblaje y catálisis del spliceosoma (Prp43, Prp22, Brr2, Prp16, etc.) en *E. histolytica* (Marchat *et al.*, 2008; Valdés *et al.*, 2014). Por otro lado, EhCFIm25 controla la selección del sitio de poli(A) distal durante la formación del extremo 3' del mRNA (Hon *et al.*, 2013). Evidencias por medio de RNAseq indican que la mayoría de los pre-mRNAs en Entamoeba son monointrónicos, lo que indica que la retención de intrones podrían ser producto de *splicing* alternativo (Davis *et al.*, 2007; Hon *et al.*, 2013).

En un estudio realizado por Torres-Cifuentes (2018) se observó que la expresión de EhSyf, ortólogo de Syf1, impacta directamente el *splicing*. Esto refleja su relación con la subunidad grande de U2AF que se une a los trectos de polipirimidina frente al 3'ss (sitio de *splicing* 3') y une el pre-mRNA/spliceosoma al CTD. De manera interesante, el silenciamiento de EhSyf, afecta negativamente la transcripción, incluso de genes sin intrones, lo que indica su papel en la elongación de la transcripción.

En un estudio realizado por González-Blanco y colaboradores, se identificó un dominio atípico denominado KH-QUA2 (84KQ) en el factor de *splicing* U2AF2 de *E. histolytica*, el cual está involucrado en la inhibición del *splicing* y en consecuencia provoca la RI de varios genes del parásito relacionados con

virulencia. Dicho dominio interactúa con SF1 y con otros factores, atenuando la virulencia del parásito al favorecer la RI (González-Blanco *et al.*, 2022).

### 2.3. Poliadenilación

En 1993 se reportó un motivo putativo de poliadenilación TAATT y 12 pirimidinas en el extremo 3'UTR (Bruchhaus *et al.*, 1993). Empleando un ensayo CLIP para estudiar a la proteína EhCstF64 fusionada a una etiqueta HA, se identificó que esta proteína forma parte de la maquinaria de poliadenilación y EhCFIm25 controla la selección del sitio poli(A) distal durante la formación del extremo 3' del mRNA (Valdés *et al.*, 2014).

Un análisis *in silico* reveló que el extremo 3' UTR de *E. histolytica* contiene tres motivos conservados: 1) la señal de poliadenilación consenso UA(A/U)UU o variantes (entre 10 y 30 nt río arriba del sitio poli(A)); 2) un tracto rico en U (entre 1 y 30 nt río arriba del sitio poli(A)), y 3) un elemento rico en U (entre 3 y 30 nt río abajo del sitio poli(A)) (López-Camarillo *et al.*, 2005). Curiosamente, los genes con sitios poli(A) alternativos pueden tener un gran impacto en la expresión génica global en *E. histolytica*, ya que la mayoría de ellos participan en la condensación y la unión del DNA, la traducción, el *splicing*, la unión del mRNA, el plegamiento y el transporte de proteínas; otros genes están relacionados con señalización, oxidación/reducción, unión de iones calcio, ciclo celular y transporte intracelular. *E. histolytica* contiene siete genes que codifican exosomas, incluidos Rrp41, Rrp43, Rrp46, Mtr3-Rrp42 y la subunidad catalítica Dis3, así como proteínas estabilizadoras accesorias Rrp4 y Rrp40 (López-Camarillo *et al.*, 2014 ).

EhCAF1 es la principal poli(A) ribonucleasa de *E. histolytica*. Además, la maquinaria de degradación del CAP contiene homólogos de DCP2 y XRN2, que son marcadores típicos de cuerpos P (cuerpos de procesamiento). Curiosamente, las proteínas EhXRN2 y EhDCP2 implicadas en la desintegración del mRNA de 5' a 3' se acumulan en focos citoplasmáticos discretos de una manera similar a la observada en eucariotas superiores (López-Rosas *et al.*, 2012).

En un estudio realizado por Hon y colaboradores (2013), se utilizaron datos de secuenciación de mRNA para cuantificar el ruido estocástico del *splicing* y la poliadenilación en *E. histolytica* y se encontró que el ruido estocástico del *splicing* y la poliadenilación era mayor en genes de baja expresión que en genes con niveles de expresión altos, de modo que los eventos de *splicing* y poliadenilación

alternativa son limitados y afectan una pequeña proporción de genes en este parásito. Finalmente, usando DREME, identificaron el motivo AAWUDA (AAUUA, AAUUUA, AAUUA, AAUUA, AAUUGA y AAUUGA) alrededor de 20 nt río arriba del sitio de poli (A), que se asemeja a la señal de poli (A) canónica AAUAAA en especies de mamíferos.

#### **2.4. Análisis Transcriptómicos en *E. histolytica***

El análisis transcriptómico es una herramienta poderosa para estudiar los patrones y la regulación de la expresión génica en *E. histolytica*. La publicación de la secuencia y los recursos del genoma de *E. histolytica* en 2005 (Loftus *et al.*, 2005) impulsó una serie de estudios de perfiles de expresión génica empleando microarreglos (Hackney *et al.*, 2007) y análisis de RNAseq en una variedad de condiciones para comprender la biología de este parásito (Naiyer *et al.*, 2019).

Por medio de estudios transcriptómicos se ha analizado la expresión génica comparando cepas virulentas y avirulentas en cultivos de laboratorio y después de la invasión de tejidos, células cultivadas en diferentes condiciones de estrés, respuesta a tratamientos con fármacos antiamebianos y cambios en la expresión genética durante el proceso de enquistamiento, etc. (González-Rivas *et al.*, 2020; Yanagawa *et al.*, 2022; Feng *et al.*, 2020).

En un estudio realizado por Naiyer y colaboradores (2019), se determinaron los patrones de expresión génica de *E. histolytica* y se identificaron genes con alto nivel de expresión y motivos promotores importantes para su expresión. A través de los datos de RNAseq de las tres condiciones (trofozoítos cultivados en condiciones normales de laboratorio, sometidos a estrés por falta de suero y después de la reposición de suero) se clasificaron a los genes de alta y baja expresión, especialmente en genes relacionados con la virulencia del parásito. A través de este estudio se encontraron nuevos motivos conservados río arriba y río abajo del codón de inicio de la transcripción necesarios para una transcripción eficiente en genes de virulencia de abundante expresión en *E. histolytica*, como ameboporos y cinasas. En los trofozoítos sin suero se observó que los FT de la familia Myb y genes de la familia BspA (relacionados con invasión de tejidos) están regulados positivamente.

Meyer y colaboradores (2019) compararon aislados patógenos y no patógenos de *E. histolytica* para identificar factores que se correlacionan con la invasión del

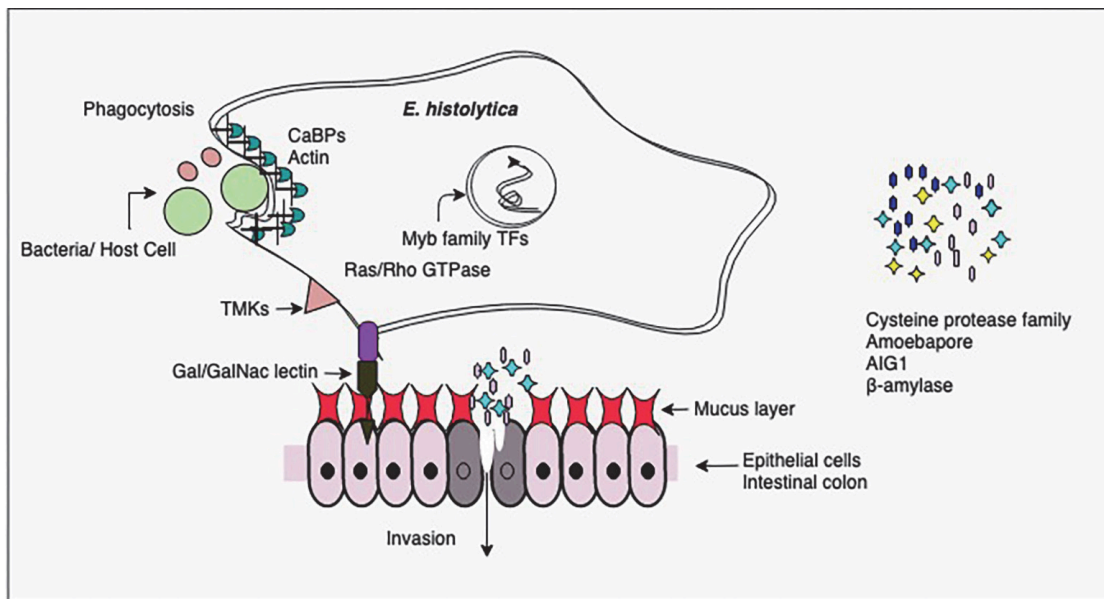
hígado en un modelo de jerbos. Estos investigadores determinaron que siete genes diferentes se sobreexpresan: genes que codifican para una metalopeptidasa, proteínas del dominio C2, alcohol deshidrogenasas y proteínas hipotéticas. Los genes expresados diferencialmente se analizaron en cepas patógenas y no patógenas de *E. histolytica*, esto permitió la identificación de nuevos factores de patogenicidad implicados en la formación de abscesos hepáticos.

Utilizando secuenciación de una sola célula (single-cell RNAseq), se encontró que los trofozoítos incubados con células CHO regulan positivamente genes de cinasas transmembrana (TMK). Algunas TMK solo se expresaron altamente en los trofozoítos del grupo de incubación en células CHO, lo que sugiere su probable participación en la percepción ambiental y la transmisión de señales. Algunas TMK se expresaron en casi cualquier condición, sugiriendo que participan en la supervivencia y proliferación celular. También se observó que pueden desempeñar funciones importantes en la transmisión de señales en condiciones de deficiencia de cisteína y pueden regular moléculas posteriores, como el metabolismo celular de los trofozoítos, con cierta regularidad. En ensayos con deficiencia de cisteína, se observó que los trofozoítos redujeron la expresión de Igl y que la deficiencia de nutrientes reduce la virulencia de *E. histolytica* (Feng *et al.*, 2020).

En un estudio realizado por Aguilar-Rojas y colaboradores (2020), tras dos horas de infección de *E. histolytica* en un modelo intestinal 3D, encontraron 201 proteínas las cuales están relacionadas con propiedades de unión a RNA, procesos celulares de óxido reducción, implicadas en metabolismo de carbohidratos o lípidos y relacionadas con el citoesqueleto. Se encontraron 67 proteínas secretadas exclusivamente en este modelo, dentro de las cuales se identificaron factores patogénicos como: antígeno B de superficie, ameboporos A y C, Grainina 1 y 2 que protegen a la ameba de la muerte celular y dos aminopeptidasas de la familia M24, así como la proteína de unión a cisteína proteasa 8.

Actualmente hay un registro en la base de datos AmoebaDB de varios estudios realizados sobre comparación entre cepas virulentas y avirulentas, o trofozoítos invasivos y células cultivadas en laboratorio, (Mattern *et al.*, 1978; Reuber y Ausubel, 1996; Bruchhaus *et al.*, 2002; Gilchrist *et al.*, 2006; Davis *et al.*, 2007;

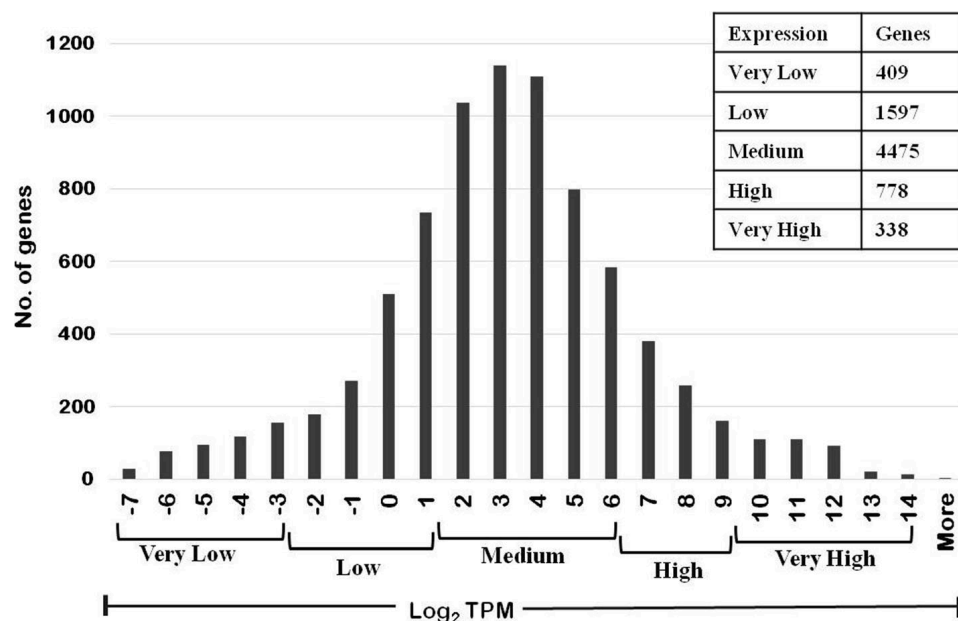
Biller *et al.*, 2009; Faust y Guillén, 2012; Thibeaux *et al.*, 2013; Meyer *et al.*, 2016; Weber *et al.*, 2016). La patogenicidad de *E. histolytica* depende de varias proteínas que se expresan diferencialmente. Entre los genes que más modifican su expresión en diferentes condiciones se encuentran: las vías de transducción de señales, el metabolismo de los carbohidratos, la actividad antibacteriana (AIG1), las funciones relacionadas con la virulencia (CP, lectina Gal/GalNAc) y los FT, en particular la familia Myb (Figura 11).



**Figura 11. Moléculas relacionadas con la virulencia reveladas a partir de estudios transcriptómicos.** Las proteasas de cisteína y la  $\beta$ -amilasa ayudan en la invasión de los tejidos degradando el moco intestinal. AIG1 funciona en defensa contra las bacterias intestinales y los amebaporos median la lisis celular. Se facilita el contacto de los ligandos de la célula huésped por lectina inhibible por Gal/GalNAc. Las cinasas transmembranales (TMK) y las GTPasas Ras/Rho participan en la transducción de señales. FT de la familia Myb regula la expresión de múltiples vías. La fagocitosis de bacterias/células huésped requiere la actividad de una gran cantidad de proteínas, incluida la unión de calcio (CaBP) que median en la reorganización de la actina. Imagen de Naiyer *et al.*, 2019.

El análisis de transcriptomas en diferentes condiciones permite identificar elementos o secuencias consenso de FT involucrados en la regulación génica. A través de bioinformática se han identificado elementos regulatorios putativos que están conservados en los promotores de genes coexpresados en *E. histolytica*, lo que ha permitido identificar secuencias consenso de FT a partir de la información de los transcriptomas. La identificación de motivos puede acoplarse además a la estadística Bayesiana para identificar interacciones potenciales entre los FT. Se ha observado que el cambio de expresión de genes ocurre a través de elementos o secuencias que controlan estos cambios transcripcionales y de los cuales se conoce poco. En un estudio realizado por Hackney y colaboradores (2007), utilizando los datos de expresión de microarreglos en dos condiciones (trofozoítos

en condiciones normales y trofozoítos sometidos a choque térmico), identificaron secuencias o motivos que se correlacionan con el nivel de expresión de los genes. Para clasificar los niveles de expresión ( $\log_2$  TPM) utilizando el concepto de clasificador Bayesiano, encontraron varios motivos en los promotores de genes con funciones similares o miembros de la familia de genes, probablemente indicando un enriquecimiento biológicamente significativo (Figura 12). Tres de estos motivos están relacionados a proteínas nucleares de *E. histolytica* y se encuentran principalmente en genes con altos niveles de expresión. Uno de los motivos encontrados ( $^A/TAAACCCT$ ), se encontró en varios genes relacionados a proteínas ribosomales y tRNA. En genes con bajos niveles de expresión se encontraron los motivos [GAATGAGT], [AACTATTTAAACAT<sup>C/T</sup>C] y [TGAACCTATAAACATC)], estos motivos se encontraron en promotores de genes homólogos a la familia Ehssp (stress sensitive protein).



**Figura 12. Clasificación de genes basada en TPM normalizado  $\log_2$ .** Los valores de TPM  $\log_2$  oscilaron entre  $-7$  y  $+14$ . Estos se dividieron arbitrariamente en clases de expresión, como se muestra. Tomada de Naiyer *et al.*, 2019.

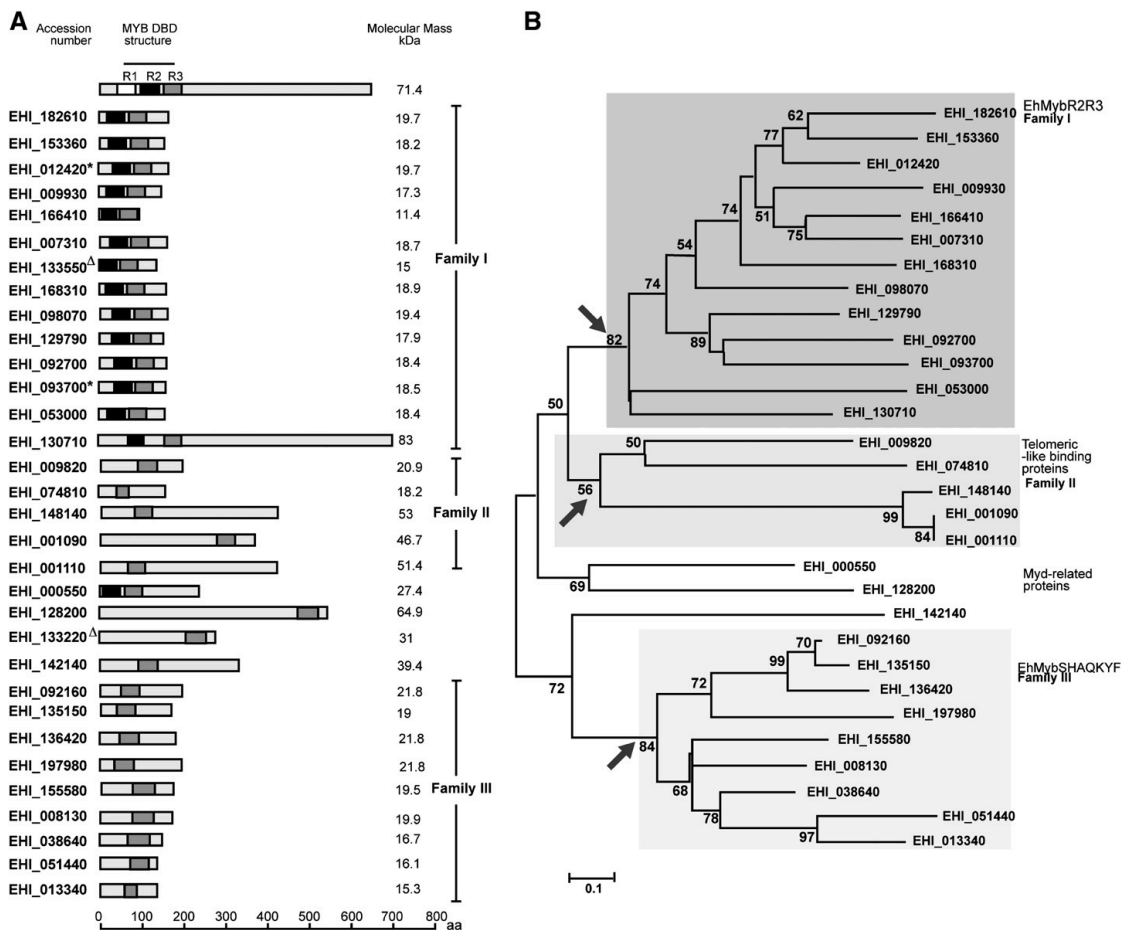
## 2.5. Proteínas Myb como factores de transcripción en *E. histolytica*

La familia de las proteínas Myb es amplia, diversa y representada en todas las células eucariotas. Las proteínas Myb desempeñan la función de FT que se unen al DNA a través de un dominio MYB (DBD-MYB) que consiste en una región de generalmente 3 repetidos (R1, R2, y R3). Dentro de cada repetido, hay tres triptófanos conservados que están separados por 18 o 19 aminoácidos y que

brindan una estructura secundaria hélice-vuelta-hélice. La tercera hélice de cada repetido es la que tiene contacto directo con el surco mayor del DNA. Durante este contacto, las dos hélices complementarias dan estabilidad a la interacción DNA-proteína (Ambawat *et al.*, 2013). El DBD-MYB está conservado en vertebrados, plantas y hongos. Este dominio media la actividad de unión a DNA en una secuencia específica y es usado también para la protección de regiones teloméricas (Dubos *et al.*, 2010).

En el genoma de *E. histolytica* se identificaron genes que codifican para proteínas que contienen el DBD-MYB, las cuales conforman una de las familias más grandes relacionadas con la regulación transcripcional (Loftus, 2005; Clark *et al.*, 2007).

Por otro lado, a través de diferentes análisis de transcriptomas se ha determinado que *E. histolytica* modifica la expresión de los genes que codifican para proteínas Myb en respuesta a diferentes condiciones (interacción con las células epiteliales de colon, durante la formación del absceso hepático, depleción de suero, entre otros) (Naiyer *et al.*, 2019), lo cual sugiere una regulación transcripcional de estos procesos. En un estudio previo se realizó un análisis filogenético en el genoma de *E. histolytica* y se encontraron 32 genes que codifican para proteínas que presentan 2 repetidos y 1 solo repetido Myb. Estas proteínas se clasificaron en tres diferentes familias: familia I, de 2 repetidos, denominada EhMybR2R3 con 15 miembros; familia II, proteínas tipo teloméricas de 1 repetido constituidas por 5 proteínas; y finalmente, la familia III conformada por 9 proteínas Myb de 1 repetido con dominio MybSHAQKYF (EhMybS) (Meneses *et al.*, 2010). Dentro de las proteínas identificadas 3 se agruparon como relacionadas a Myb (Figura 13).

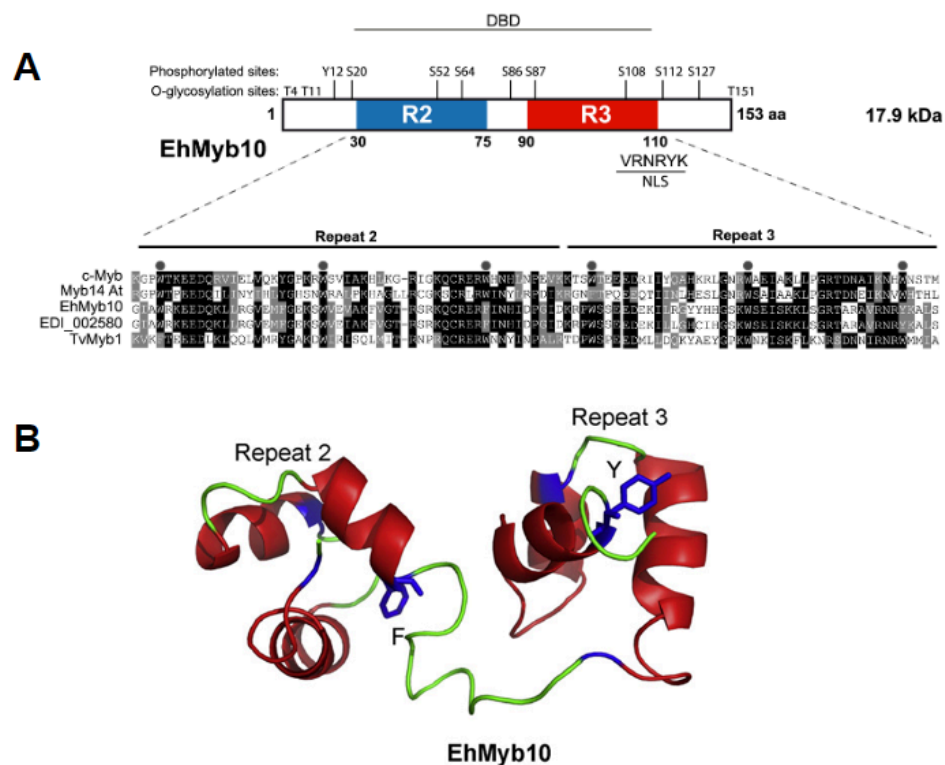


**Figura 13. Las proteínas que contienen DBD de *E. histolytica* MYB forman tres familias bien definidas.** (A) Diagrama esquemático de las proteínas que contienen c-Myb y proteínas de *E. histolytica* con MYB DBD. Se indica la organización (R1, blanco; R2, negro sólido; y R3, gris), los números de acceso de Pathema-Entamoeba y la masa molecular. \*Denota proteínas codificadas por genes duplicados, y  $\Delta$  denota proteínas codificadas por genes truncados. (B) Árbol filogenético que muestra las relaciones entre las proteínas que contienen DBD de MYB de *E. histolytica*. Familias específicas de *E. histolytica* están resaltadas por cajas de luz. Los números sobre los nodos del árbol indican el porcentaje de bootstrap (1000 réplicas) para un nodo del árbol en particular. Solo los bootstrap N50 % son exhibidos. Las flechas indican los valores de arranque que respaldan la clasificación de las proteínas que contienen DBD de *E. histolytica* MYB en familias. La barra de escala corresponde a 0,1 sustituciones estimadas de aminoácidos por sitio. Tomada de Meneses *et al.*, 2010.

## 2.6. EhMyb10

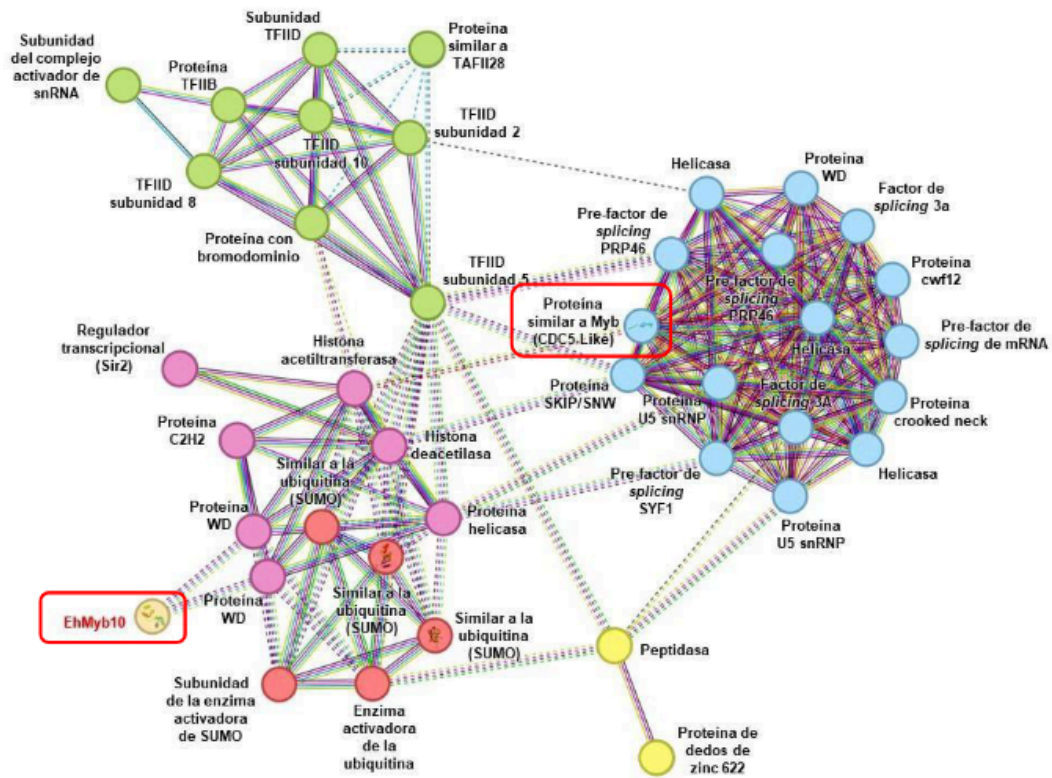
El gen *ehmyb10* (EHI\_129790) codifica para una proteína de 153 aminoácidos con un peso molecular de 17.9 kDa, con un punto isoeléctrico de 10.47 y es uno de los genes *myb* de mayor expresión (Meneses *et al.*, 2010). En un trabajo previo (Meneses *et al.*, 2010) se modeló la estructura terciaria teórica del DBD-MYB de esta proteína, mostrando que es similar a la del factor de transcripción c-Myb y que conserva una estructura de hélice-vuelta-hélice en los repetidos R2 y R3 (Figura 14). Adicionalmente, mediante análisis EMSA se reveló que la proteína EhMyb10 reconoce y se une al elemento de reconocimiento MYB (ERM) [TAACGG], incluso con una mayor afinidad que otras proteínas EhMyb

(Meneses *et al.*, 2010). Por otro lado, al analizar la región promotora de 5422 genes, se observó que 246 genes contienen este elemento y que participan en funciones como el transporte vesicular, la respuesta a choque térmico, así como factores de virulencia (Meneses *et al.*, 2010). Adicionalmente, mediante un ensayo de microarreglos se mostró que *ehmyb10* se sobreexpresa en trofozoítos al primer día de la interacción con enterocitos en epitelio de colon de ratón (Gilchrist *et al.*, 2006). Esto fortalece la idea de que EhMyb10 regula genes relacionados con la virulencia en este organismo.



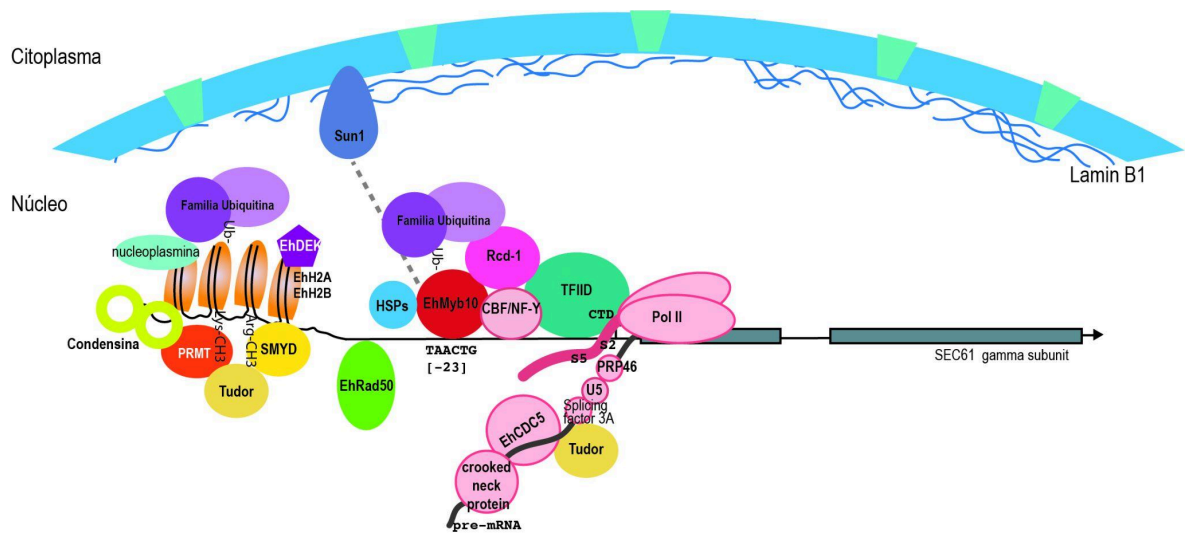
**Figura 14. EhMyb10 de *E. histolytica*.** A) Diagrama de la proteína EhMyb10 donde se muestran los repetidos R2 y R3 que conforman del DBD-Myb, la señal de localización nuclear y las posibles modificaciones post traducciónales, también se muestra un alineamiento del DBD-Myb en el alineamiento destacan los residuos de triptófano conservados en diferentes especies: c-Myb (*Homo sapiens*), Myb14 At (*A. thaliana*), EhMyb10 (*E. histolytica*), EDI\_002580 (*E. dispar*), TvMyb1 (*T. vaginalis*). B) Estructura tridimensional del DBD-Myb de EhMyb10. Los residuos de color azul corresponden a los triptofanos conservados presentes en los repetidos R2 y R3. Tomado de Meneses *et al.*, 2010 y Azuara-Liceaga *et al.*, 2015.

En un trabajo realizado por Danna Paola Ixyalit Aguirre Casimiro, se determinó que EhMyb10 interactúa con proteínas involucradas en eventos de organización de cromatina, ubiquitinación y de *splicing*. Entre las proteínas de *splicing* que interactúan con EhMyb10 se encuentra EhCDC5 (Figura 15), lo cual provee las primeras evidencias de su papel en la regulación cotranscripcional.



**Figura 15. Interactoma *in silico* de EhMyb10.** Interactoma obtenido mediante STRING, empleando un puntaje de interacción de confianza media (0.400) como valor de corte y un número máximo de no más de 10 interactuantes. Las esferas representan a las proteínas que interactúan con EhMyb10 (naranja) y que participan en diversos procesos: Transcripción (verde), *splicing* (azul), organización de la cromatina (morado), modificación postraduccional (rojo), reparación al daño del DNA (morado); y otras (amarillo). Número de nodos: 41, número de aristas: 265, valor p de enriquecimiento PPI <math>10e-16</math>. Tomado de Aguirre-Casimiro 2023.

También, como parte de este trabajo, se identificaron a las proteínas nucleares de *E. histolytica* que interactúan con la proteína EhMyb10 recombinante, mediante un ensayo de Pull-Down. Se observó que aproximadamente 93 proteínas se unen de manera específica a EhMyb10 y el 35.8 % desempeñan una función principal de unión a RNA. Estas proteínas comprenden principalmente aquellas que se enlazan al RNA ribosómico (rRNA), así como en menor medida al RNA mensajero (mRNA) en la región 5'-UTR y a los snRNA, lo que sugiere su implicación en la regulación de la expresión génica a nivel de transcripción, *splicing* y traducción. Además, un 10.8 % de las proteínas identificadas forman parte de los complejos de pre-ensamblaje de los ribosomas. Finalmente, EhMyb10 se une a proteínas que regulan la estructura de la cromatina como metiltransferasas. Estos hallazgos subrayan la complejidad y la versatilidad de las proteínas interactuantes de EhMyb10 (Figura 16) (Aguirre-Casimiro, 2023).



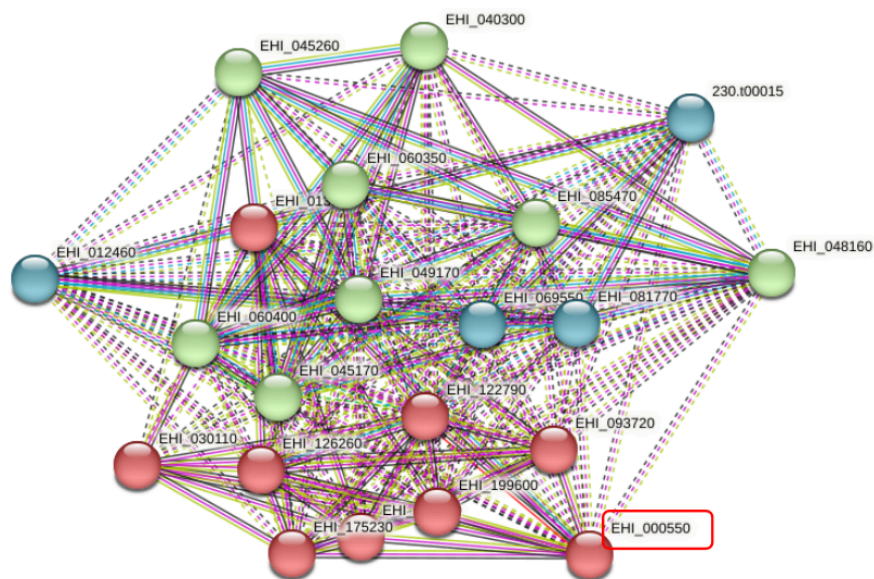
**Figura 16. Proteínas que interactúan con EhMyb10.** Proteínas identificadas mediante espectrometría de masas a partir de un ensayo de Pull-down empleando EhMyb recombinante y extractos nucleares de *E. histolytica*. Las proteínas en color rojo son las que se hizo la predicción en String. Tomado Aguirre-Casimiro 2023.

## 2.7. EhCDC5

El gen de *EhCDC5* de *E. histolytica* se conforma por 717 pb, no contiene intrones y codifica para una proteína de 238 aminoácidos con un peso molecular teórico de 27.4 kDa y un punto isoelectrico de 9.39. En la base de datos AmoebaDB se encuentra bajo el número de acceso EHI\_000550. Aunque presenta 2 repetidos, los análisis filogenéticos no mostraron asociación con la familia EhR2R3, por lo tanto se clasificó como una proteína relacionada con MYB (Meneses *et al.*, 2010). El análisis de su DBD-MYB mostró similitud con la proteína CDC5 (PTHR45885:SF1-PANTHER). Además de los dos repetidos MYB, presenta varias regiones desordenadas en el extremo carboxilo terminal.

Las proteínas de la familia CDC5 realizan una función esencial en el control del ciclo celular G2/M. De manera interesante, por medio de inmunoprecipitación, se demostró que la proteína CDC5 de *Homo sapiens* se asocia con las proteínas del spliceosoma y al utilizar anticuerpos contra la misma, se inhibe el proceso de *splicing*. Al utilizar un cultivo de *S. cerevisiae* deficientes en CEF1, una proteína ortóloga de CDC5, se observó que los mensajeros conservaban los intrones y eran incapaces de realizar el proceso de *splicing* (Burns *et al.*, 1999). Estas evidencias sugieren que CDC5 desempeña un papel importante en el evento de *splicing* y que las células eucariotas requieren de esta proteína durante el procesamiento del pre-mRNA.

En un análisis *in silico* para predecir interacciones proteína-proteína mediante la herramienta STRING, se predijo la interacción de EhCDC5 con EhMyb10 y factores de *splicing* (Figura 17).



**Figura 17. Interactoma de EhCDC5.** Las esferas representan a las proteínas que interactúan con EhCDC5 (EHI\_000550) y que participan en diversos procesos: Spliceosoma (verde), NTC (rojo), U2 snRNP (azul). Elaboración propia

Finalmente, en un trabajo realizado por Valdés y col. (2014), se identificaron por espectrometría de masas a las proteínas asociadas a la maquinaria de *splicing* en *E. histolytica* por medio de ensayos de inmunoprecipitación y entrecruzamiento (CLIP) (Figura 18). En este trabajo se identificaron cerca de 100 proteínas asociadas a *splicing*, algunas pertenecientes a los complejos A o B, asociadas a las RNPs U1, U2, U4, U6. Además, se identificaron las proteínas EhCDC5 y Prp19 que son componentes clave del NTC (Valdés *et al.*, 2014). Estas evidencias sugieren la participación de la EhCDC5 en el proceso de *splicing* de *E. histolytica*.

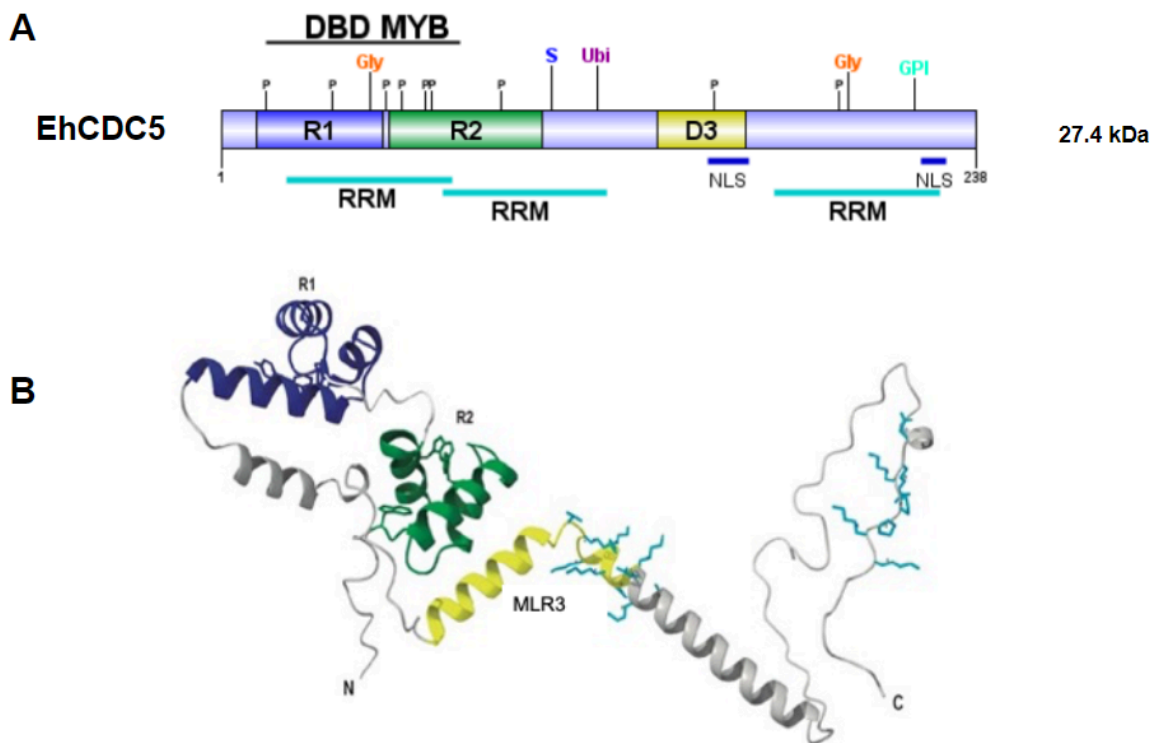
Table 1 – Components of <i>Entamoeba histolytica</i> pre-mRNA splicing complexes.														
Particle/complex/class	Protein name	Gene ID	Protein access	MW	Splicing factor	Number of peptides <sup>a</sup>								
						U1A	C1	C1+	X	X+	M	M+	HA	MS2
U4/U6.U5 tri-snRNP	Peptidyl-prolyl cis-trans isomerase	EHI_125840	67482289	18	GPR6	2			1				8	8
	Ubiquitin carboxyl-terminal hydrolase domain containing protein	EHI_152110	67484272	52	6SK/SAD1		2			1				5
Prp19/CDC5L complex	PRP38 family protein	EHI_000490	67474026	23	Prp38							1		15
	WD domain containing protein	EHI_130870	67478341	52	Prp19	3	5		1				5	6
Prp19/CDC5L complex-related	Regulator of nonsense transcripts	EHI_193520	67472499	108	KIAA0560 (fSAP164)		3						4	

<sup>a</sup> Number of peptides are listed for each sample: HA-IP (HA) using nuclear extracts from mock and HA-U1A (U1A) amoeba transformants, and MS2-IP (MS2) using nuclear extracts from untreated or UV crosslinked (+) HA-RabC1 (C1), HA-gRabX13 (X) and HA-gRabX13-MS2 (M) amoeba transformants.

<sup>b</sup> Percent of maximal coverage obtained for each protein in the different IP assays.

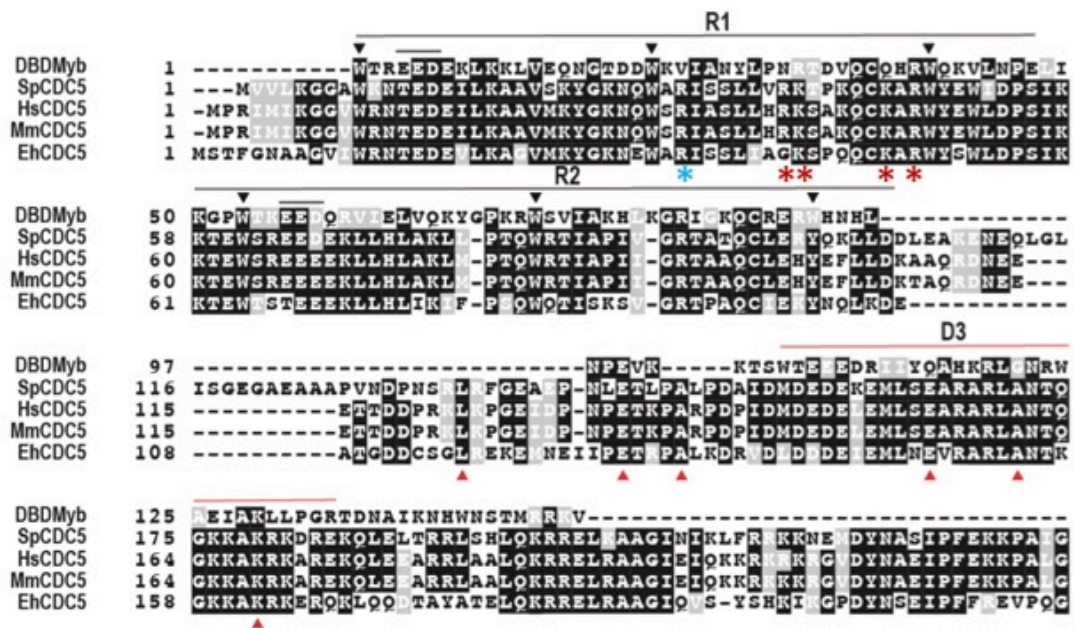
**Figura 18. Algunas proteínas del complejo de splicing identificadas mediante CLIP.** Dentro de las proteínas detectadas se encuentra el NTC donde el CDC5 es parte de este complejo. Tomado de Valdés *et al.*, 2014.

Para conocer más sobre la proteína EhCDC5 en *E. histolytica*, en este trabajo se realizó análisis *in silico* para predecir sus modificaciones post traduccionales utilizando diferentes software (NetPhos-3.1, big-PI Predictor, NetOGlyc - 4.0, JASSA v4 (Joined Advanced Sumoylation Site and Sim Analyser), CKSAAP\_UbSite Prediction, GPS-Sumo 2.0). Mediante este análisis se predice que EhCDC5 es susceptible a ubiquitinaciones, glicosilaciones, SUMOilaciones y fosforilaciones. Además, presenta un posible sitio de adición glicosil-fosfatidil-inositol en el extremo carboxilo terminal. También se predijeron mediante el software CatRapid signature, tres motivos de reconocimiento a RNA y dos señales de localización nuclear (NLStradamus) que fortalecen su posible participación en el proceso de *splicing*. Su estructura tridimensional muestra la característica hélice-vuelta-hélice de su DBD-MYB (Figura 19).



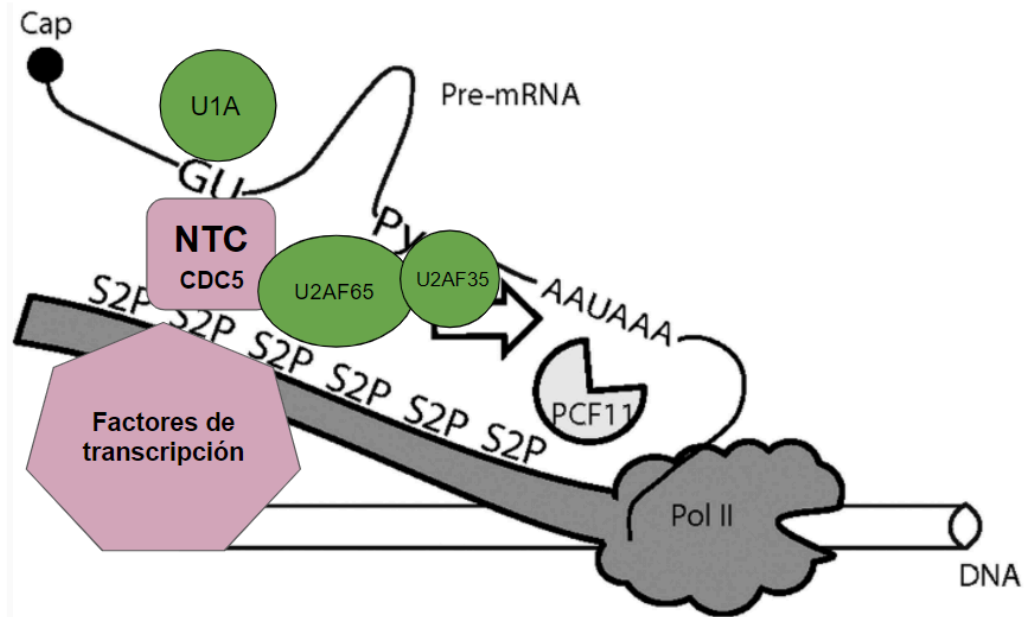
**Figura 19. Proteína EhCDC5 de *E. histolytica*.** A) Esquema de la proteína donde se muestran sus dominios Myb: el repetido R1 (azul) y R2 (verde) y D3 (amarillo). En la parte superior se indican sus posibles modificaciones postraduccionales: P fosforilación, S SUMOilación, Ubi ubiquitinación. RRM corresponde a los motivos de reconocimiento a RNA y NLS a la señal de localización nuclear. B) Modelo de la proteína EhCDC5 obtenido por AlphaFold y visualizado con el programa ChimeraX. Elaboración propia.

Al realizar un alineamiento de EhCDC5 con proteínas CDC5 de otros organismos, se observó la conservación de los triptófanos en los repetidos R1 y R2, además de un alto porcentaje de identidad con las otras proteínas CDC5, lo que puede dar un indicio de su función (Figura 20).



**Figura 20. Alineamiento de EhCDC5.** Alineamiento de secuencias de aminoácidos de EhCDC5 con otras proteínas CDC5 de otros organismos (*S. pombe*, *H. sapiens* y *Mus musculus*). En negro se observan regiones conservadas, con círculos rojos se indican los triptófanos conservados, las líneas indican los repetidos 1 y 2, además del tercer repetido denominado D3. También se muestra un porcentaje de identidad entre las secuencias del 56.2 %.

Un ensayo CLIP utilizando snRNP U1A con una etiqueta de HA (hemaglutinina), realizado en *E. histolytica* por Valdés y colaboradores (2014), permitió la identificación de numerosos factores relacionados con el *splicing*, entre ellos EhCDC5 (Figura 18), así como componentes de la transcripción (subunidad grande de RNAPII, y varios FT como EhMyb10 (comunicación personal Dr. Valdés) y lo que evidencia la complejidad de este proceso cotranscripcional (Figura 21).



**Figura 21. Procesamiento cotranscripcional de pre-mRNA en *E. histolytica*.** Durante la elongación de la transcripción por la RNAPII, los residuos de Ser2 de los repetidos del heptapéptido del extremo CTD son fosforilados y reclutan proteínas del spliceosoma y de poliadenilación. La subunidad grande del factor de *splicing* auxiliar U2 U2AF65 interactúa con el extremo CTD fosforilado (Ser2-CTD) con factores de *splicing* que conforman el NTC y con FT. Modificado de Valdés-Flore *et al.*, 2018; Gu *et al.*, 2013.

Los antecedentes anteriores sugieren que EhMyb10 es un TF que interacciona con diversas proteínas de unión a RNA y que puede interactuar con proteínas involucradas en el *splicing* como EhCDC5 que en otros organismos EhCDC5 forma parte del NTC. Por ello, en este proyecto es de interés conocer el comportamiento del transcriptoma en condiciones de sobreexpresión de estas dos proteínas de manera individual y de manera conjunta, para tener una mayor comprensión de la regulación cotranscripcional en este parásito por medio de análisis transcriptómicos.

### **3. Hipótesis**

Las proteínas EhMyb10 y EhCDC5 regulan cotranscripcionalmente la expresión génica de *E. histolytica* y al sobreexpresarse modifican el transcriptoma y los patrones de *splicing* en este microorganismo.

### **4. Justificación**

*Entamoeba histolytica* es un parásito de importancia médica y es responsable de la amebiasis que provoca cerca de 100,000 muertes anuales, por lo tanto, es importante comprender los eventos moleculares que controlan la expresión de los genes involucrados en la relación huésped-parásito. Asimismo, aunque se tiene conocimiento de los factores de transcripción y de los involucrados en el *splicing*, se conoce poco de los mecanismos cotranscripcionales involucrados en la regulación de la expresión génica en este protozoario.

Por tal motivo, en este proyecto se analizará el papel del factor de transcripción EhMyb10 y el factor de *splicing* EhCDC5 en los mecanismos cotranscripcionales, lo que ayudará a comprender los procesos de expresión génica de *E. histolytica*.

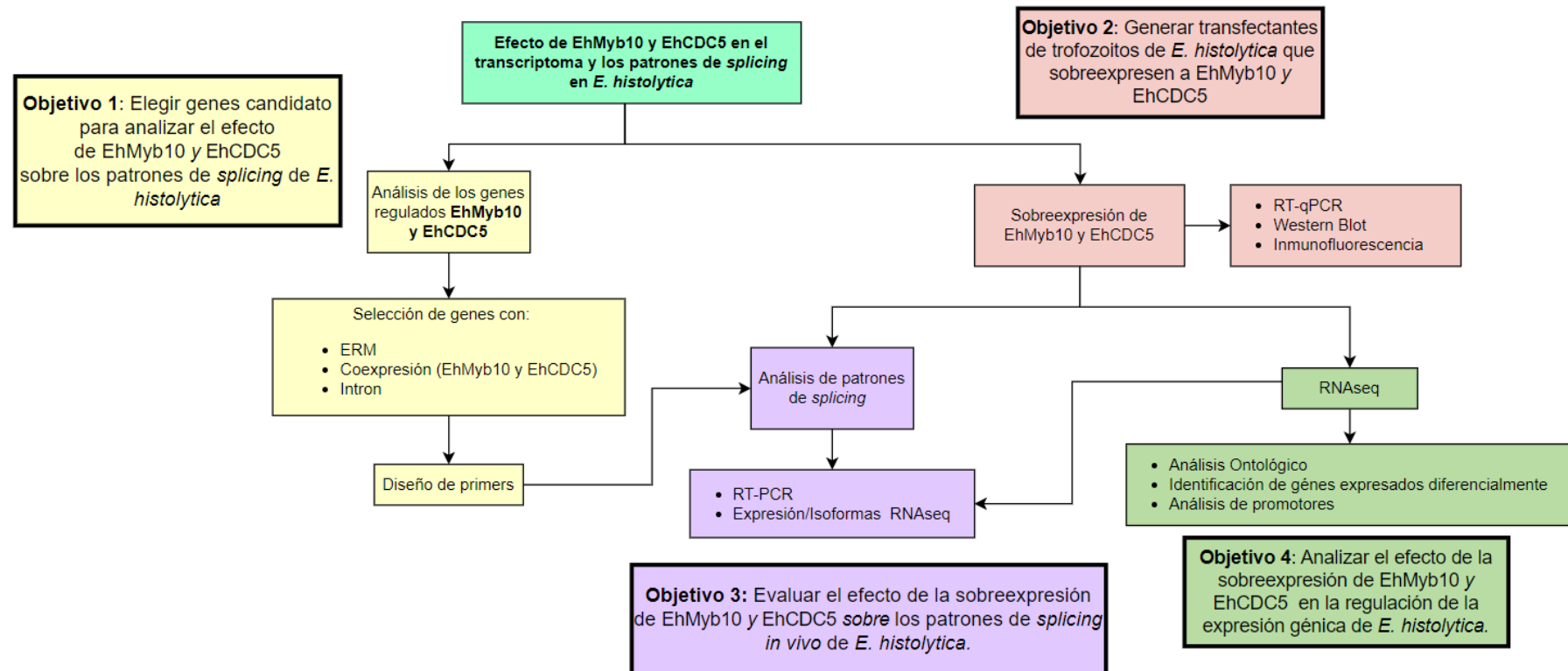
### **5. Objetivo general:**

Analizar el efecto sobre la regulación cotranscripcional de EhMyb10 y EhCDC5 en *E. histolytica*.

### **6. Objetivos particulares:**

1. Elegir genes candidatos para analizar el efecto de EhMyb10 y EhCDC5 sobre los patrones de *splicing* de *E. histolytica*.
2. Generar trofozoítos de *E. histolytica* que sobreexpresen a EhMyb10 y EhCDC5.
3. Evaluar el efecto de la sobreexpresión de EhMyb10 y EhCDC5 sobre los patrones de *splicing in vivo* de *E. histolytica*.
4. Analizar el efecto de la sobreexpresión de EhMyb10 y EhCDC5 en la regulación de la expresión génica de *E. histolytica*.

## 7. Estrategia experimental



## 8. Materiales y Métodos

### 8.1. Identificación *in silico* de los genes blanco que son regulados cotranscripcionalmente por EhMyb10 y EhCDC5 en *E. histolytica*

Para identificar los genes que tienen elementos de reconocimiento Myb (ERM) en sus regiones promotoras, se realizó una búsqueda del ERM canónico por medio de la expresión regular [C/T]AAC[G/T]G como se describió en Meneses *et al.*, (2010). Esta actividad se realizó con ayuda del M. en C. Helios Cárdenas Hernández. Posteriormente, se analizó el transcriptoma de trofozoítos en condiciones basales de cultivo publicado por Shabardina y col (2018) en la base de datos AmoebaDB (<https://amoebadb.org/amoeba/app>). Los valores de expresión basal en transcritos por millón (TPM) se procesaron de la siguiente forma: 1) Se aplicó la transformación  $\log_2$  de los TPM; 2) Se obtuvo la media ( $\bar{X}$ ) de los datos; 3) Cada dato se normalizó con la siguiente ecuación:  $x = \frac{(\log_2 TPM - \bar{x}^2)}{sd}$ . Una vez normalizados se clasificaron de acuerdo a sus niveles de expresión (expresión muy baja, baja, moderada, alta y muy alta) de acuerdo a la estrategia reportada por Naiyer *et al.* (2019). Para ello se contó con la ayuda del M. en C. Carlos Martínez, Profesor de la Academia de Matemáticas, Plantel Casa Libertad, UACM. Se seleccionaron aquellos genes con valores de expresión similares a *ehmyb10* y *ehcdc5* (expresión moderada) y sólo aquellos que contienen intrones (<https://amoebadb.org/amoeba/app>) y se seleccionaron 4 genes para analizar sus patrones de *splicing*.

### 8.2. Análisis ontológico de los genes

Para el análisis ontológico se utilizó Panther Classification System (<https://www.pantherdb.org>). La secuencia completa de los genes se obtuvo de la base de datos de AmoebaDB (<https://amoebadb.org/amoeba/app/>).

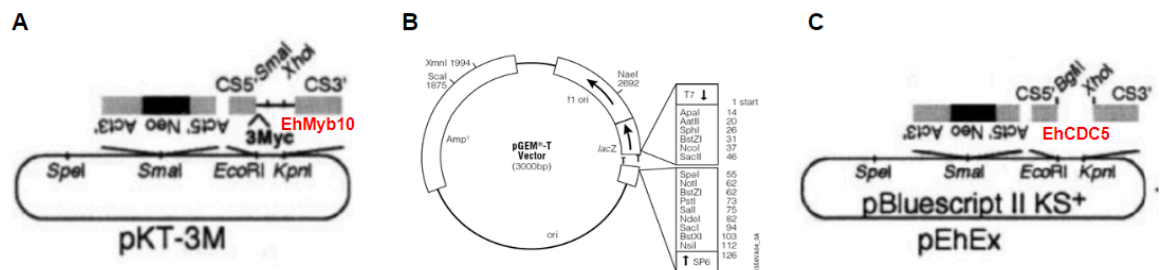
### 8.3. Cultivo de trofozoítos de *E. histolytica*

Los trofozoítos de la cepa HM1: IMSS se cultivaron axénicamente a 37 °C en tubos en medio TYI-S33 (Diamond *et al.*, 1978) suplementado con suero bovino (Biowest) al 15 % (inactivado por calor), 3 % de vitaminas de Diamond (Sigma-Aldrich), y 1x Pen-Strep (Sigma-Aldrich). En todos los experimentos, se

usaron trofozoítos en fase de crecimiento logarítmico y se cosecharon por centrifugación a 1800 rpm durante 5 minutos a 4 °C.

#### 8.4. Obtención de la construcción pEhEx-EhCDC5

Para analizar el efecto de la sobreexpresión de EhMyb10 y EhCDC5 se requería de clonar sus genes en vectores de expresión amebianos. Para ello se seleccionaron los plásmidos pKT3M (Saito-Nakano *et al.*, 2014) y pEhEx (Saito-Nakano *et al.*, 2014) (Figura 22). Para la construcción pEhEx-EhCDC5 se diseñaron oligonucleótidos que amplifican la secuencia completa del gen *ehcdc5* y que contienen los sitios de restricción *SmaI* y *XhoI* (Tabla 3). El gen amplificado a partir de DNAg se clonó en el vector pGEM-T-Easy (Promega) y se subclonó en el vector de transfección de *E. histolytica* pEhEx (Saito-Nakano *et al.*, 2004) en células *E. coli* DH5 $\alpha$  (figura 22). Las clonas obtenidas se analizaron por PCR colony y por ensayos de restricción con las enzimas *SmaI* (New England) y *XhoI* (New England). La identidad del gen *ehcdc5* en el vector de interés se determinó por medio de secuenciación en el equipo 3130 Genetic Analyzer. Bacterias *E. coli* DH5 $\alpha$  se transformaron con la construcción pEhEx-EhCDC5 para la obtención del plásmido. Los cultivos de bacterias se cosecharon y el plásmido se purificó con el kit Qiagen Maxi Plasmid Purification (Qiagen) siguiendo las indicaciones del fabricante y se cuantificó para transfectar trofozoítos de *E. histolytica*. Como control negativo se utilizó el vector vacío pEhEx.



**Figura 22. Vectores utilizados para la construcción de EhMyb10 y EhCDC5.** A) Vector pKT3M utilizado para sobreexpresar EhMyb10. B) Vector PGEM-T-easy utilizado para clonar el gen *ehcdc5*. C) Vector de subclonación pEhEx

#### 8.5. Extracción de DNA plasmídico para transfección

El DNA plasmídico se extrajo de bacterias transformadas empleando el Plasmid Maxi Kit (QIAGEN). Las colonias bacterianas seleccionadas se inocularon en 1.5 mL de medio LB suplementado con ampicilina (100  $\mu$ g/mL), después de 16 h de incubación a 37 °C, con agitación (360 rpm), se inocularon 500 mL de medio en

matraces con medio LB + ampicilina y se incubaron a 37 °C por 16 h con agitación a ~180 rpm. Para cosechar las bacterias, éstas se centrifugaron a 8,000 rpm por 15 min a 4 °C y se obtuvo la pastilla, la cual se resuspendió con 10 mL de buffer de resuspensión (P1). Posteriormente, se agregaron 10 mL de buffer de lisis (P2) y se mezcló por inversión 6 veces hasta obtener una consistencia viscosa y azulada (indica la presencia del vector). Después de 5 min de incubación a temperatura ambiente, se agregaron 10 mL de buffer de neutralización (P3). Inmediatamente, las muestras se mezclaron por inversión, se incubaron 20 min a 4 °C y se centrifugaron a 8,500 rpm por 30 min. El sobrenadante se pasó por una columna, previamente equilibrada con 10 mL de buffer de equilibrio (QBT). Posteriormente, la columna se lavó dos veces con 30 mL de buffer de lavado (QC) para eliminar contaminantes. El DNA se eluyó pasando por la columna 15 mL de buffer de elución (QF). El DNA se precipitó con 10.5 mL de isopropanol y se centrifugó a 8,500 rpm durante 30 min a 4 °C. La pastilla se lavó con 5 mL de etanol al 70 % y la pastilla seca se resuspendió en 150 µL de agua inyectable. El DNA plasmídico se cuantificó mediante Nanodrop (Thermo Fisher Scientific) y se almacenó a 4 °C hasta su uso.

#### **8.6. Transfección de trofozoítos de *E. histolytica* con las construcciones pEhEx-EhCDC5 y pKT3M-EhMyb10**

La transfección de la cepa HM1: IMSS de *E. histolytica* se realizó en un tubo de 6 mL con  $1 \times 10^6$  trofozoítos. Para ello, se mezclaron 5 µg/mL del plásmido correspondiente (plásmido pEhEx, pEhEx-EhCDC5 o/y pKT3M-EhMyb10), 150 µL de medio TYI-S-33 sin suero y 5 µL de SuperFect (Qiagen) (Saito-Nakano *et al.*, 2004). Transcurrido el periodo de incubación, se retiró el medio TYI-S-33 y se agregó lentamente la mezcla de transfección a cada tubo. Los tubos se incubaron de forma horizontal a 37 °C durante 3 h. Después, los trofozoítos se transfirieron a tubos de 12 mL y se incubaron por 24 h. Las amibas se cosecharon a las 24, 48 y 72 h postransfección (hpt). Para la transfección estable se inició el proceso de selección con el antibiótico G418 (Life technologies, un análogo de neomicina) a una concentración de 1 µg/mL a las 24 hpt aumentando gradualmente según la confluencia de los cultivos (80 %). También se generaron transfectantes estables de las 4 condiciones y se mantuvieron a una concentración de 20 µg/mL de G418.

### **8.7. Aislamiento de RNA total de los trofozoítos transfectados**

La extracción de RNA total se realizó a partir de los trofozoítos transfectados y cosechados a los diferentes tiempos. Las células se colectaron, centrifugaron a 1,800 rpm por 5 min y se adicionaron 0.5 mL de TRIzol (Thermofisher). Posteriormente, se extrajo el RNA total y las proteínas de acuerdo a las indicaciones del fabricante (Thermo Scientific). La integridad del RNA se verificó en un gel de agarosa al 1 % y por Qubit RNA IQ Assay (Thermofisher). La concentración del RNA se determinó con el kit Qubit RNA BR (Broad Range) (Thermofisher). El RNA total para secuenciación se almacenó a - 70 °C hasta su uso.

### **8.8. Extracción y cuantificación de DNA genómico**

El DNA genómico de los trofozoítos se extrajo empleando el kit Wizard ® Genomic DNA (Promega). Los trofozoítos se centrifugaron a 8,000 rpm durante 10 min. La pastilla resultante se lavó con 200 µL de PBS y se centrifugó a 8,000 rpm por 1 min para descartar el PBS. Posteriormente, la pastilla se resuspendió en 600 µL de solución de lisis de núcleos (NLS) hasta homogenizar la muestra. Después, se agregaron 3 µL de RNasa, se mezcló por inversión 2-5 veces y se incubó por 30 min a 37 °C, seguido de una incubación de 5 min a temperatura ambiente. Posteriormente se añadieron 200 µL de solución de precipitación de proteínas (PPS), se agitó con vórtex y se centrifugó a 8,000 rpm por 4 min. El sobrenadante, se transfirió a un tubo limpio de 1.5 mL y se agregaron 600 µL de isopropanol a TA, se mezcló suavemente por inversión y se centrifugó 1 min a 8,000 rpm a temperatura ambiente. La pastilla se lavó con 600 µL de etanol al 70 % y se centrifugó 1 min a 8,000 rpm a temperatura ambiente. La pastilla seca se resuspendió en 100 µL de agua inyectable y se cuantificó mediante Nanodrop (Thermo Fisher Scientific) y se almacenó a 4 °C hasta su uso.

### **8.9. Análisis de la expresión de *ehmyb10* y *ehcdc5* por qRT-PCR**

Para evaluar la sobreexpresión de *ehcdc5* y *ehmyb10*, 50 ng de RNA total obtenido de los diferentes cultivos se trataron con DNasa I y posteriormente se utilizaron para la síntesis de cDNA y amplificación de los genes *ehcdc5* y *ehmyb10* utilizando el kit OneStep RT-PCR (QIAGEN). Como control interno se amplificó el gen *s2* (40s) de la subunidad ribosomal 40s. Los oligonucleotidos

utilizados se muestran en la Tabla 3. La expresión relativa de *ehmyb10* y *ehcdc5* se calculó con base en la expresión del gen *s2* de la subunidad ribosomal 40s mediante el método de doble delta Ct ( $\Delta\Delta Ct$ ) de acuerdo a Livak y Schmittgen (2001).

### 8.10. Diseño de oligonucleótidos

Las secuencias completas de los genes seleccionados se obtuvieron de la base de datos AmoebaDB. Para el diseño de oligonucleótidos se utilizó Primer3 (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) y se analizaron por medio de la herramienta OligoAnalyzer Tool (IDT) para determinar la formación de estructuras secundarias y dímeros (Tabla 3).

**Tabla 3. Oligos empleados en este proyecto de investigación.**

ID	Nombre		Secuencia del oligonucleótido	Tm (°C)	Tamaño del amplicón (pb)
<b>Clonación del gen <i>EhCDC5</i></b>					
	<i>EhCDC5 SmaI</i>	S	TCCCCCGGGATGTCAACGTTTGAAATGCT	66.5	717
	<i>EhCDC5 XhoI</i>	AS	CCGTCGAGTTAATTCATTTCTTTACCAATAAAT	59.4	
<b>Detección por qRT-PCR</b>					
EHI_000550	<i>EhCDC5</i>	S	TTTCCATCACAATGGCAAAC	51.7	163
		AS	TTAAAGCAGGTCGCGTTTCT	54.8	
EHI_020280	<i>40S ribosomal protein S2 (s2 40s)</i>	S	GAGATCACAACGGTCATGCT	62	190
		AS	CTGACACGAACTGAACCACA	55	
EHI_129790	<i>EhMyb10</i>	S	AGACCATGGAGTTCAGAAGAAGATG	56.3	110
		AS	ACAGCACGAGCAGTTCTTCCA	59	
<b>Retención intrónica</b>					
EHI_139370	<i>Proteína con dominio HMG</i>	S	GCATTAGAATTATGGGAAACAGAAC	52.5	446
		AS	ATGGTGGAATGTCAGTAGAATG	52.4	
EHI_158190	<i>Proteína del complejo Swi5/Sfr1</i>	S	AAAAAATCTTCTCAAGGGGC	20	402
		AS	TCCCATTGAAAATTGATCACC	21	
EHI_129760	<i>Factor de splicing rico en serinas y argininas</i>	S	TGAGCGTATTGATATAAACTGG	24	424
		AS	TACTTCTGCTGTTTCGGTGAC	20	
EHI_153670	<i>U1 SNRNP 70</i>	S	ATTCATTTGCGCCAACCTTG	19	467
		AS	GCTCTTTCTGCTTCACTTCG	20	

### **8.11. Inmunofluorescencia indirecta (IFI)**

Para los ensayos de IFI, trofozoítos transfectados se colocaron y fijaron en cubreobjetos de 15 mm con paraformaldehído al 4 % y se permeabilizaron con tritón X-100 (0.3 %). Posteriormente, se lavaron 3 veces con PBS pH 7.4 filtrado y se bloquearon con albúmina de suero bovino (BSA) al 1 %. El anticuerpo primario se incubó a 4 °C toda la noche a la dilución recomendada por el fabricante (anti-HA, 1:100 (conejo), anti-Myc, 1:500 (ratón)) (Invitrogen). Las laminillas se lavaron 3 veces con PBS pH 7.4 y se incubaron con el anticuerpo secundario correspondiente (anti-conejo FITC (1:100) (SIGMA-Aldrich), anti-ratón Alexa 647 (1:50) (Invitrogen). Las laminillas se lavaron 3 veces con PBS pH 7.4, se montaron en portaobjetos con Vectashield con DAPI y se visualizaron en un microscopio confocal Zeiss (LS700), empleando el programa ZEN Light Edition.

### **8.12. Western Blot**

Para obtener los extractos de proteínas totales, trofozoítos transfectados a las 72 hpt de las diferentes condiciones se cosecharon y lavaron con PBS 1X pH 6.8. Posteriormente, se les adicionó amortiguador RIPA e inhibidores de proteasas Complete™ Protease inhibitor cocktail (Sigma-Aldrich) y se sonicaron por 30 s a 30 % de intensidad 2x. Los extractos obtenidos se cuantificaron por el método de Bradford y las proteínas se separaron en un gel SDS-PAGE al 10 % para analizar su integridad. Posteriormente se transfirieron a una membrana de nitrocelulosa y se incubaron con los siguientes anticuerpos primarios:  $\alpha$ -actina (1:1,000, Sigma-Aldrich),  $\alpha$ -HA acoplado a HRP (1:1500) (Invitrogen) para EhCDC5 y  $\alpha$ -EhMyb10 hecho en conejo (1:500). Como anticuerpo secundario se utilizó un  $\alpha$ -conejo acoplado a HRP (1:5000). La señal se reveló con HRP Western Luminata Forte (Merck Millipore) y se visualizó con Chemidoc MP de Bio-Rad.

### **8.13. Análisis *in vivo* de los patrones de *splicing* en los trofozoítos que sobreexpresan a las proteínas EhMyb10 y EhCDC5**

Para analizar la RI se diseñaron oligonucleótidos que flanquean un intrón en los 4 genes seleccionados (Tabla 3). Para estos genes se realizaron ensayos de PCR por triplicado con el cDNA obtenido de los trofozoítos transfectados en las diferentes condiciones. Los amplicones se cuantificaron por densitometría empleando el programa ImageJ y se calculó el índice de RI de acuerdo a la

siguiente fórmula:  $100x \frac{\text{intrón retenido}}{\text{intrón retenido} + \text{mRNA procesado}}$  y se comparó con la condición control (González-Blanco *et al.*, 2022; De la Mata *et al.*, 2010). Los resultados se analizaron mediante la prueba t de Student (GraphPad Prism v.5). Los amplicones del gen EHI\_153670 se secuenciaron para corroborar su identidad mediante electroforesis capilar en el equipo 3130 Genetic Analyzer.

#### **8.14. Construcción y secuenciación de bibliotecas y obtención de los transcriptomas**

El RNA extraído de los trofozoítos transfectados transitoriamente con las diferentes construcciones se analizó por fluorometría por Bioanalyzer 2100 para verificar la integridad de las muestras y posteriormente construir una biblioteca empleando el kit TruSeq Stranded (Illumina). Las bibliotecas de cDNA se prepararon usando tres réplicas biológicas de cada condición y se analizaron de forma pareada empleando RNAseq masiva. La secuencia se realizó en un equipo de secuenciación NextSeq 500/550 (Illumina) por medio de 150 ciclos para una secuencia de 2 x 75 ciclos, analizando  $10 \times 10^6$  lecturas pareadas. La calidad de los datos de secuenciación masiva se analizó con el software FastQC. Esto se realizó en la Unidad Universitaria de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la UNAM. Posteriormente, se hizo un mapeo contra el genoma de referencia (AmoebaDB). La cuantificación de la expresión de los transcritos y el análisis de expresión diferencial se realizó mediante RSEM y por medio de DESeq2, respectivamente, comparando cada muestra contra el control. Finalmente, la expresión e isoformas de los transcritos en cada condición de las transfectantes se analizaron por medio de Ballgown, utilizando RStudio para la visualización de la expresión de las isoformas en cada condición.

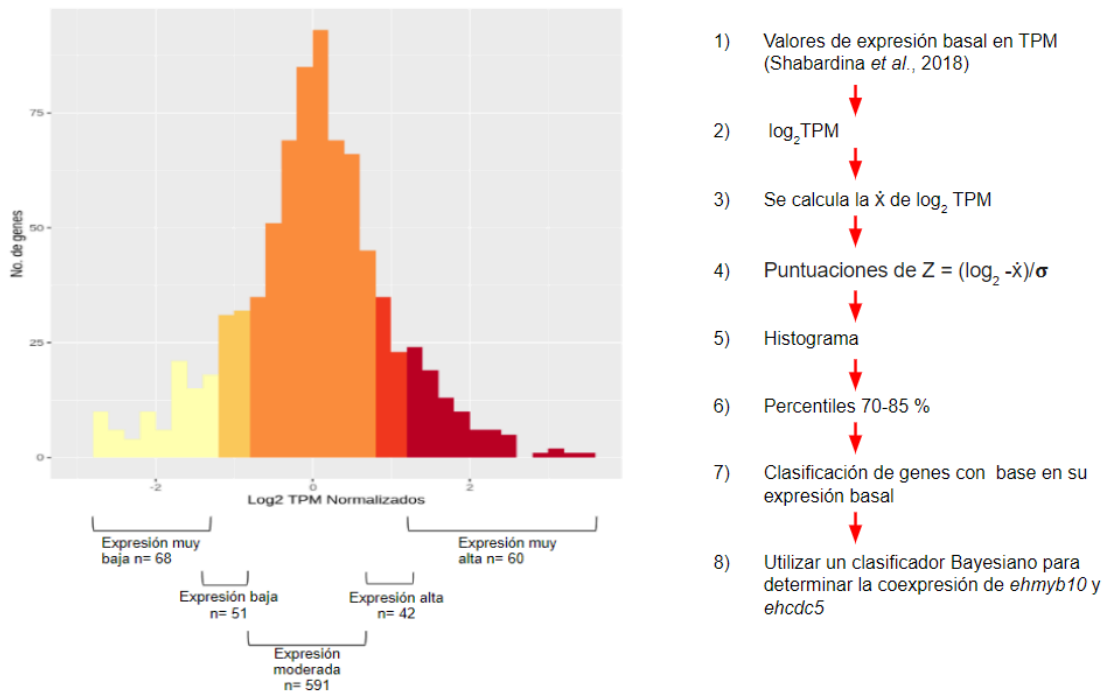
## 9. Resultados

### 9.1. Selección de genes candidatos de EhMyb10 y EhCDC5 sobre los patrones de *splicing* de *E. histolytica*

Para identificar los genes de este parásito que contienen el elemento de reconocimiento a Myb (ERM) ([C/T]AAC[G/T]G) en sus regiones reguladoras, se hizo un análisis *in silico* de acuerdo a la estrategia reportada por Meneses (2010). Para ello, se tomaron los 8,201 genes de *E. histolytica* y se analizaron 500 pb río arriba del ATG (región promotora) de cada gen. A partir de este análisis se identificaron 873 genes que tienen en su promotor el ERM. Posteriormente, se obtuvieron sus niveles de expresión basal en la base de datos AmoebaDB de acuerdo al transcriptoma obtenido por Shabardina *et al.*, 2018 a través de RNAseq. Posteriormente, se construyó un histograma, el cual mostró que la expresión de estos genes presenta una curva de distribución normal similar a la que han obtenido otros autores (Hackney *et al.*, 2007, Naiyer *et al.*, 2019). Los niveles de expresión se organizaron en 5 categorías mediante el cálculo del cuantil 75-85 % (Tabla 4). Las categorías obtenidas fueron: 1) expresión muy baja (68 genes), 2) baja (51 genes), 3) moderada (591 genes), 4) alta (42 genes) y 5) muy alta (60 genes) (Figura 23). Posteriormente, se determinó el grupo en el que los genes *ehmyb10* y *ehcdc5* pertenecían empleando el método estadístico de máxima verosimilitud. Se encontró que estos tienen una expresión moderada, por lo cual se analizaron los genes pertenecientes a este grupo. Dentro de este grupo conformado por 591 genes, 147 tuvieron al menos un intrón (84 %). El máximo número de intrones que contienen estos genes es de 4, con una longitud promedio de 50 nt.

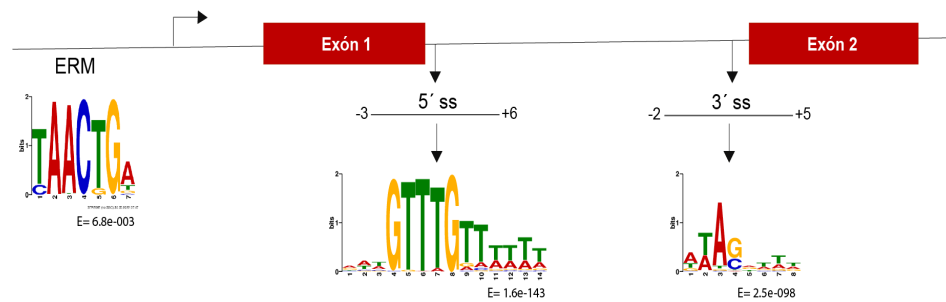
Tabla 4. Niveles de expresión de genes que contienen el ERM en sus promotores.

Categoría (nivel de expresión)	Número de genes	Rango ( $\log_2$ TPM)
Muy baja	68	-2.70948 a -1.4395315
Baja	51	-1.4395315 a -1.0364334
Moderada	591	-1.0364334 a 1.0364334
Alta	42	1.0364334 a 1.4395315
Muy Alta	60	1.4395315 a 3.43538



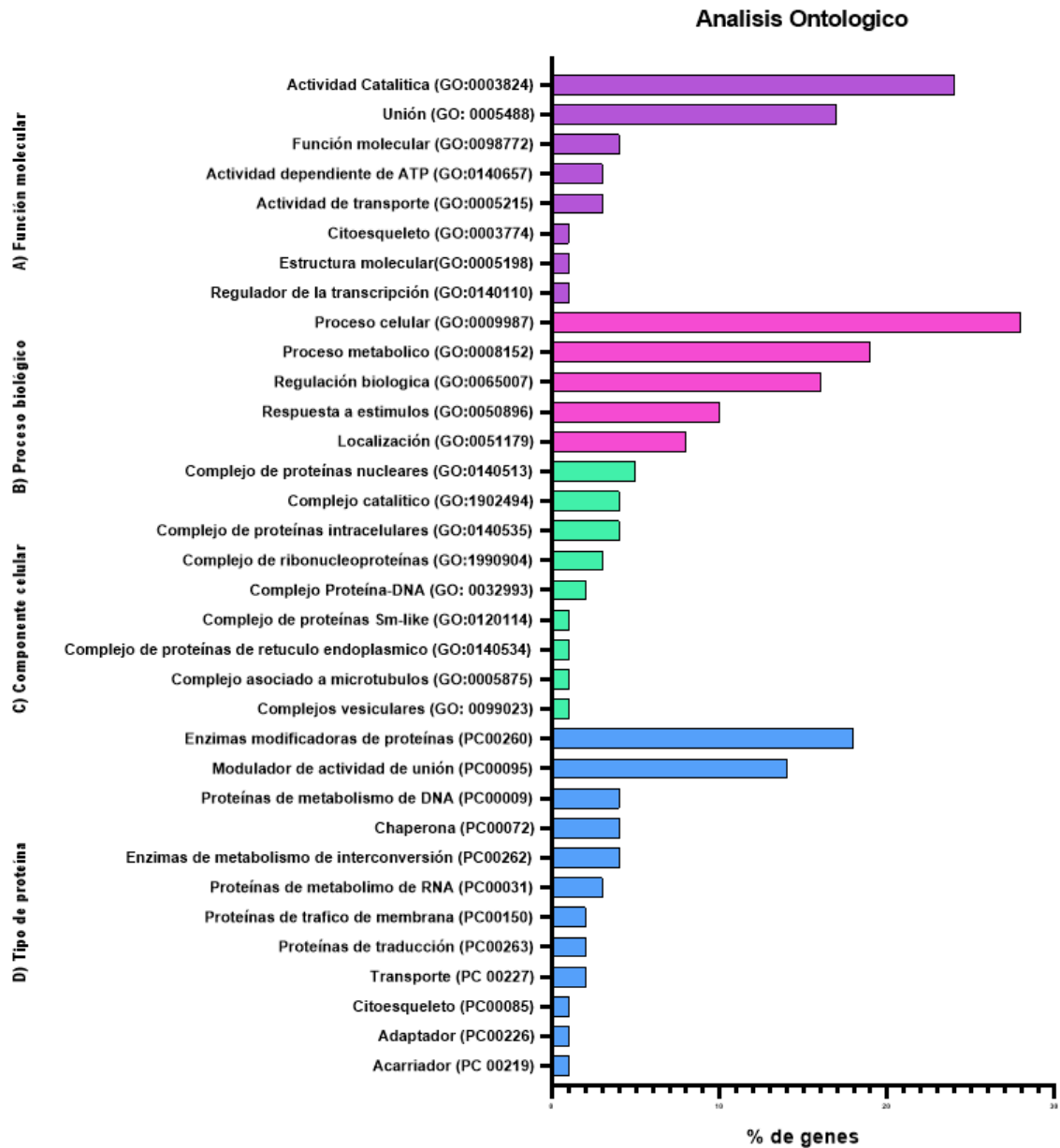
**Figura 23. Clasificación de la expresión de los genes que contienen ERM en sus promotores.** Histograma del  $\log_2$  de los TPM normalizados  $[(\log_2 - \bar{x}^2)/sd]$  provenientes de trofozoítos en condiciones basales de cultivo de acuerdo al transcriptoma de Shabardina y col (2018). La clasificación de los grupos se realizó de acuerdo con el cálculo del percentil 70 % (amarillo) y 85 % (rojos).

Una vez determinado que los genes *ehmyb10* y *ehcdc5* se coexpresan en condiciones basales y pertenecen al grupo de expresión moderada, se procedió a determinar los genes que podrían ser regulados por estos dos factores. Para ello, se tomó el grupo de genes de expresión moderada (591 genes) y se seleccionaron aquellos genes que contenían al menos un intrón y al elemento ERM en su región promotora (-500 pb río arriba del INR y +30 pb río abajo). Producto de este análisis, se encontraron 147 genes que cumplen con estos criterios. Por medio de la herramienta MEME-Suite se determinó la secuencia y logo que corresponde al ERM en la región promotora de estos genes y además se identificaron las secuencias GTTG (sitio 5' ss) y TAG (sitio 3' ss) necesarias para el proceso de *splicing* (Figura 24), lo cual coincide con lo previamente descrito para este organismo (Hon *et al.*, 2013).



**Figura 24. Presencia del ERM y secuencias de *splicing* en los posibles genes blanco de EhMyb10 y EhCDC5.** Esquema que muestra los motivos conservados relacionados con el ERM en sus promotores y sitios de *splicing* 5' ss y 3' ss de los 147 genes con intrón empleando la herramienta MEME-Suite.

Posteriormente, se realizó el análisis ontológico de los 147 genes, los cuales se clasificaron en cuatro categorías: función molecular (69), proceso biológico (125), componente celular (60) y tipo de proteína (69) (Figura 25). De los 69 genes agrupados en función molecular, la más frecuente corresponde a actividad catalítica (GO:0003824), seguido de unión (GO:0005488) y finalmente actividad reguladora de la transcripción (GO:0140110). Ejemplos de estos genes son los que codifican para cinasas, proteínas Rho y Ras y proteínas remodeladoras de cromatina SWI5. De los 125 genes relacionados con procesos biológicos, los más frecuentes fueron el proceso celular (GO:0009987), proceso metabólico (GO:0008152) de macromoléculas y regulación biológica (GO:0065007), el cual, por ejemplo, se refiere a la modulación del control de la expresión génica. En cuanto al tipo de proteína, los términos más abundantes fueron enzimas modificadoras (PC00260), modulador de actividad de unión (PC00095) como la GTPasa, y activador cinasa y proteínas involucradas en el metabolismo de RNA (PC00031) como factores de *splicing*, factores generales de transcripción y RNA polimerasa. Finalmente, de acuerdo al componente celular, el término más frecuente corresponde a complejos de proteínas nucleares (GO:0140513), complejo catalítico (GO:1902494) y complejo de ribonucleoproteínas (GO:1990904). De forma interesante, los tres términos se asocian con procesos de *splicing* por lo que refuerza la hipótesis de que EhMyb10 y EhCDC5 podrían interactuar para participar de manera cotranscripcional. Además, 13 de estos genes se modulan en condiciones de colonización intestinal y en clones virulentos de HM1:IMSS, como el gen que codifica para la ribonucleoproteína pequeña U1 (EHI\_153670), involucrada en el proceso de *splicing* y una proteína cinasa (EHI\_109970), lo cual estaría implicado en el control de la expresión de genes relacionados con la virulencia (Naiyer *et al.*, 2019).



**Figura 25. Análisis ontológico de los posibles genes blanco de EhMyb10 y EhCDC5.** Los 147 genes que tienen el ERM en sus promotores y que tienen intrones se analizaron mediante Panther Gene Ontology Functional Enrichment. A) Función molecular B) Proceso biológico C) Componente celular D) Tipo de proteína.

De los 147 genes blancos putativos que cumplieron con los criterios establecidos, se seleccionaron 4 genes (Tabla 5) que estuvieran relacionados con transcripción o *splicing* para analizar el efecto de la sobreexpresión de EhMyb10 y EhCDC5 en sus patrones de *splicing*.

**Tabla 5. Genes seleccionados (candidatos) para analizar el efecto de EhMyb10 y EhCDC5 sobre los patrones de *splicing* de *E. histolytica*.**

ID	Anotación*	Función	Intrones	Tamaño del intrón (pb)
EHI_139370	Proteína con dominio HMG	Proteína de unión a DNA e interacción proteína-proteína	1	67
EHI_158190	Proteína del complejo Swi5/Sfr1	Reparación de DNA y recombinación homóloga	1	48
EHI_129760	Factor de <i>splicing</i> rico en serinas y argininas	Factor de <i>splicing</i>	2	44 y 51
EHI_153670	U1 SNRNP 70	Factor de <i>splicing</i>	1	154

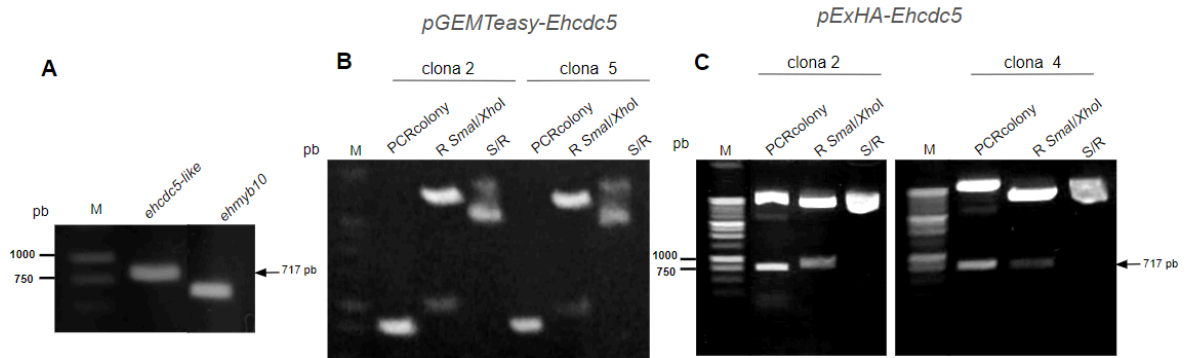
De acuerdo a la base de datos AmoebaDB

## 9.2. Sobreexpresión de EhMyb10 y EhCDC5 en trofozoítos de *E. histolytica*

Para la sobreexpresión de EhMyb10 se contaba en el laboratorio con la construcción pKT3M-EhMyb10 que permite la sobreexpresión de la proteína unida a una etiqueta de Myc (3x Myc). Por ello se clonó el gen *ehcdc5* en el plásmido pEhEx que contiene la secuencia de una etiqueta de hemaglutinina y así poder determinar la sobreexpresión de ambas proteínas. Esto permitió sobreexpresar a las proteínas fusionadas a los tags de Myc (EhMyb10) y HA (EhCDC5).

### 9.2.1. Clonación del gen *ehcdc5* en el vector pEhEx

Para clonar el gen completo *ehcdc5* se diseñaron oligonucleótidos específicos. Por medio de PCR a partir de DNA genómico de *E. histolytica* se amplificó una banda de 717 pb que corresponde al tamaño completo del gen *ehcdc5*. El plásmido pKT3M-EhMyb10 se utilizó como control positivo obteniendo un amplicón ~500 pb (Figura 26A). Posteriormente, el gen *ehcdc5* se purificó y ligó con el vector de clonación pGEM-T-Easy. Se obtuvieron clonas candidatas que se analizaron por PCR colony y por restricción con las enzimas *XhoI* y *SmaI* de las cuales, las clonas 2 y 5 liberaron el inserto que corresponde al tamaño del gen *ehcdc5* (717 pb) (Figura 26B). La clona 5 se seleccionó para restringir con las enzimas antes descritas para posteriormente purificar y subclonar el inserto en el vector de transfección pEhEx previamente linearizado con las mismas enzimas. Las colonias obtenidas se analizaron por PCR y se restringieron para verificar la correcta subclonación del gen. Las clonas 2 y 4 amplificaron y liberaron un producto del tamaño esperado (717 pb) (Figura 26C).



**Figura 26. Clonación y subclonación del gen *ehcdc5* de *E. histolytica*.** A) Amplificación de los genes *ehcdc5* y *ehmyb10* a partir de DNA genómico y DNA plasmídico pKT3M-EhMyb10 respectivamente. B) Clonación de *ehcdc5* en el vector pGEM T Easy. PCR colony y restricción con las enzimas *SmaI* y *XhoI* de las clonas 2 y 5. C) Subclonación de *ehcdc5* en el vector pExHA. PCR colony y restricción con las enzimas *SmaI* y *XhoI* de las clonas 2 y 4. Los geles de agarosa se realizaron al 1 %.

La identidad de *ehcdc5* de las clonas obtenidas se confirmó por medio de secuenciación mediante electroforesis capilar. Al alinear las secuencias obtenidas con el gen de referencia (EHI\_000550) se obtuvo la secuencia completa sin mutaciones y se localizó la etiqueta de hemaglutinina (Figura 27). Esta construcción nos permitió la sobreexpresión de EhCDC5 en trofozoítos de *E. histolytica*.

```

SC4          -----
SC4_Rcomp1  GTAATTATTTAAAGGTTGGTCAATGAAGTACTGACAAACACATTAAACAGATCTATCTATCCATATGATGTTCCAGATTATGCT 640
EHI_000550  -----ATGTCAACGTTTGGAAATGCTGC 23
SC4          TATCCATATGATGTTCCAGATTATGCTATCCAGATTATCCGGGATGCAACGTTTGGAAATGCTGC 720
SC4_Rcomp1  TATCCATATGATGTTCCAGATTATGCTATCCAGATTATCCGGGATGCAACGTTTGGAAATGCTGC 720
EHI_000550  TGGGGTAATATGGAGAAATACTGAAGATGAAGTACTTAAAGCAGGAGTAAATGAAATATGGTAAAAATGAATGGGCAAGAA 103
SC4          TGGGGTAATATGGAGAAATACTGAAGATGAAGTACTTAAAGCAGGAGTAAATGAAATATGGTAAAAATGAATGGGCAAGAA 800
SC4_Rcomp1  TGGGGTAATATGGAGAAATACTGAAGATGAAGTACTTAAAGCAGGAGTAAATGAAATATGGTAAAAATGAATGGGCAAGAA 800
EHI_000550  TTAGTAGCTTAATTCGTGTAATCTCCACACAGTGTAAAGCTCGGTGGTATAGTTGGCTAGACCCATCAATTTAAAAG 183
SC4          TTAGTAGCTTAATTCGTGTAATCTCCACACAGTGTAAAGCTCGGTGGTATAGTTGGCTAGACCCATCAATTTAAAAG 0
SC4_Rcomp1  TTAGTAGCTTAATTCGTGTAATCTCCACACAGTGTAAAGCTCGGTGGTATAGTTGGCTAGACCCATCAATTTAAAAG 880
EHI_000550  ACTGAATGGACTTCAACTGAAGAGAGAAATTAATTCATTAAATTAAGATATTTCCATCACAATGGCAACAATTTCTAA 263
SC4          ACTGAATGGACTTCAACTGAAGAGAGAAATTAATTCATTAAATTAAGATATTTCCATCACAATGGCAACAATTTCTAA 0
SC4_Rcomp1  ACTGAATGGACTTCAACTGAAGAGAGAAATTAATTCATTAAATTAAGATATTTCCATCACAATGGCAACAATTTCTAA 960
EHI_000550  AAGTGTGGAGAACTCCAGCACAATGTATTGAAAAATATAACCAACTTAAAGATGAAGCTACTGGAGATGATTGCTCAG 343
SC4          AAGTGTGGAGAACTCCAGCACAATGTATTGAAAAATATAACCAACTTAAAGATGAAGCTACTGGAGATGATTGCTCAG 67
SC4_Rcomp1  AAGTGTGGAGAACTCCAGCACAATGTATTGAAAAATATAACCAACTTAAAGATGAAGCTACTGGAGATGATTGCTCAG 1040
EHI_000550  GATTAAGAGAAAAAGAAATGAATGAATCATTCCAGAAACGGACCTGCTTTAAAGACAGAGTTGACCTTGATGATGAT 423
SC4          GATTAAGAGAAAAAGAAATGAATGAATCATTCCAGAAACGGACCTGCTTTAAAGACAGAGTTGACCTTGATGATGAT 147
SC4_Rcomp1  GATTAAGAG----- 1048
EHI_000550  GAGATTGAAATGTTAAATGAAGTTCGTGCTCGTCTAGCAAATACAAAAGGAAAGAAAGCTAAACGTAAGAACGTCAAAA 503
SC4          GAGATTGAAATGTTAAATGAAGTTCGTGCTCGTCTAGCAAATACAAAAGGAAAGAAAGCTAAACGTAAGAACGTCAAAA 227
SC4_Rcomp1  ----- 1048
EHI_000550  ATTACAACAGACACAGCCTATGCCACTGAACTTCAAAAAAGAGAGAAGCTTAGAGCAGCTGGTATTCAAGTTAGTTATT 583
SC4          ATTACAACAGACACAGCCTATGCCACTGAACTTCAAAAAAGAGAGAAGCTTAGAGCAGCTGGTATTCAAGTTAGTTATT 307
SC4_Rcomp1  ----- 1048
EHI_000550  CTCATAAAATAAAAGGACCAGATTATAATTAGAAATACCATTCTTAGAGAAGTACTCAAGGAAAATTTAATCCTCAA 663
SC4          CTCATAAAATAAAAGGACCAGATTATAATTAGAAATACCATTCTTAGAGAAGTACTCAAGGAAAATTTAATCCTCAA 387
SC4_Rcomp1  ----- 1048
EHI_000550  AAAGATCGTAAACCACCAAGAAACCATCATTATTGGTAAAGAAATGAATTAATCTGAGTTGAACTCTCTTTTAA 717
SC4          AAAGATCGTAAACCACCAAGAAACCATCATTATTGGTAAAGAAATGAATTAATCTGAGTTGAACTCTCTTTTAA 467
SC4_Rcomp1  ----- 1048

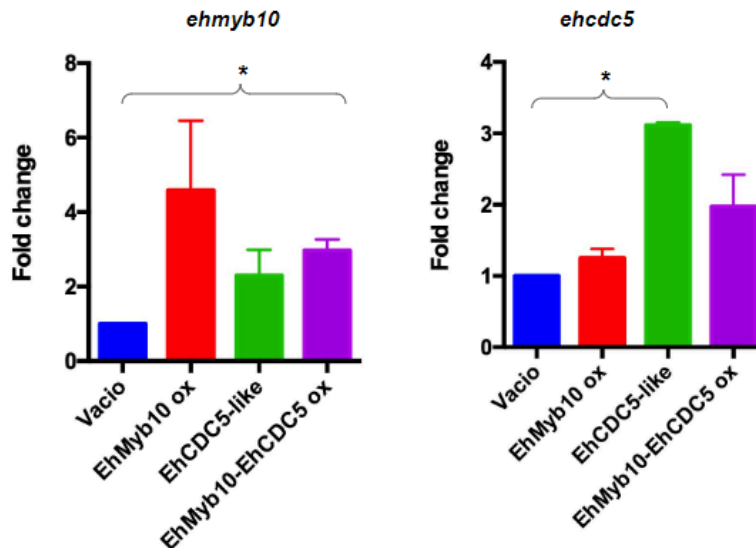
```

**Figura 27. Secuenciación de la construcción pEhEx-EhCDC5.** Alineamiento de la secuencia de DNA subclonada en el vector pEhEx y la secuencia de referencia de EhCDC5 (EHI\_000550 en AmoebaDB). La región sombreada corresponde al traslape de la lectura en sentido y antisentido. Recuadro azul, etiqueta de hemaglutinina; recuadro rojo, codón de inicio; recuadro verde, codón de paro.

Los plásmidos pEhEx, pEhEx-CDC5 y pKT3M-Myb10, se purificaron, se cuantificaron y se revisó su integridad en un gel de agarosa. Posteriormente, se realizó la transfección transitoria empleando 5 µg de cada plásmido en las siguientes condiciones: A) Control (pEhEx), B) EhMyb10ox (pKT3M-Myb10), C) EhCDC5ox (pEhEx-CDC5) y D) EhMyb10+EhCDC5ox (cotransfección pKT3M-Myb10 y pEhEx-CDC5).

### **9.2.2. Análisis de la expresión de los genes *ehmyb10* y *ehcdc5* por RT-PCR tiempo real**

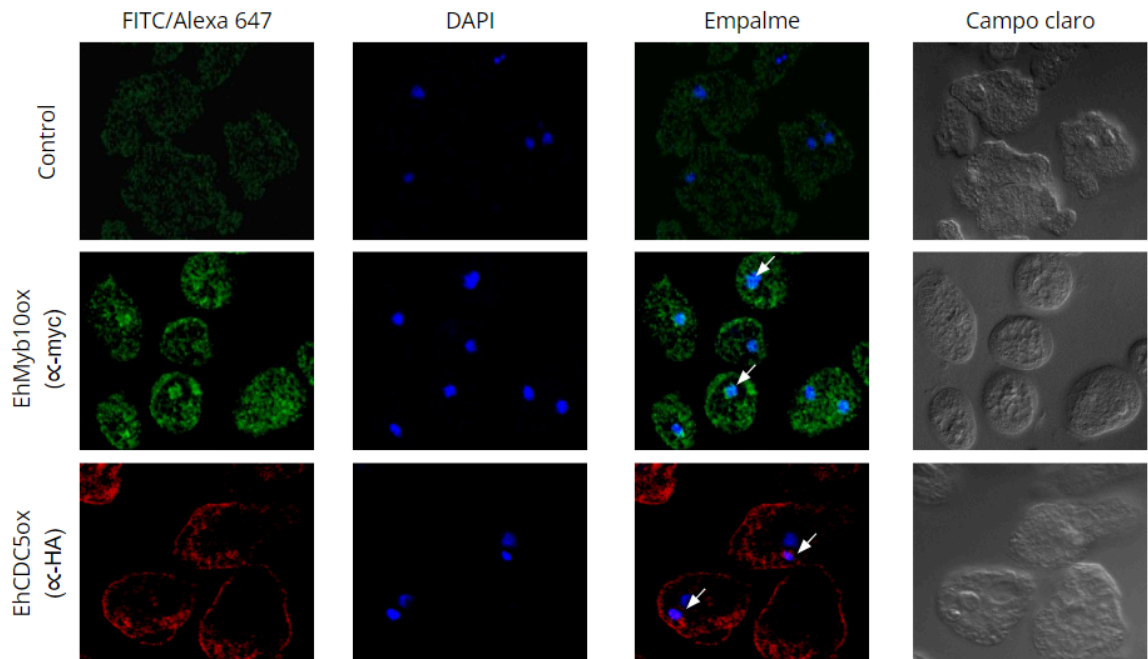
Para determinar el efecto de la sobreexpresión de las proteínas EhMyb10 y EhCDC5 los trofozoítos transfectados se cosecharon a las 24, 48 y 72 hpt para analizar la sobreexpresión por medio de RT-qPCR. Este análisis mostró que a las 72 hpt *ehmyb10* y *ehcdc5* tuvieron una mayor expresión respecto al control cuando se transfectan cada construcción por separado (Figura 28). El gen *ehmyb10* aumentó su expresión ~4 veces en la condición EhMyb10ox; sin embargo, en la condición EhCDC5ox y en la cotransfección se observa una sobreexpresión de aproximadamente 2 y 3 veces respectivamente con respecto al control. En el caso del gen *ehcdc5* se observó una sobreexpresión de ~3 y de 2 veces con respecto al control en la condición EhCDC5ox y la cotransfección, respectivamente. Sin embargo, en la condición EhMyb10ox no se sobreexpresa el gen *ehcdc5*. Podemos concluir que en las condiciones de transfección transitoria se sobreexpresaron los genes *ehmyb10* y *ehcdc5* con respecto al control en todas las condiciones analizadas.



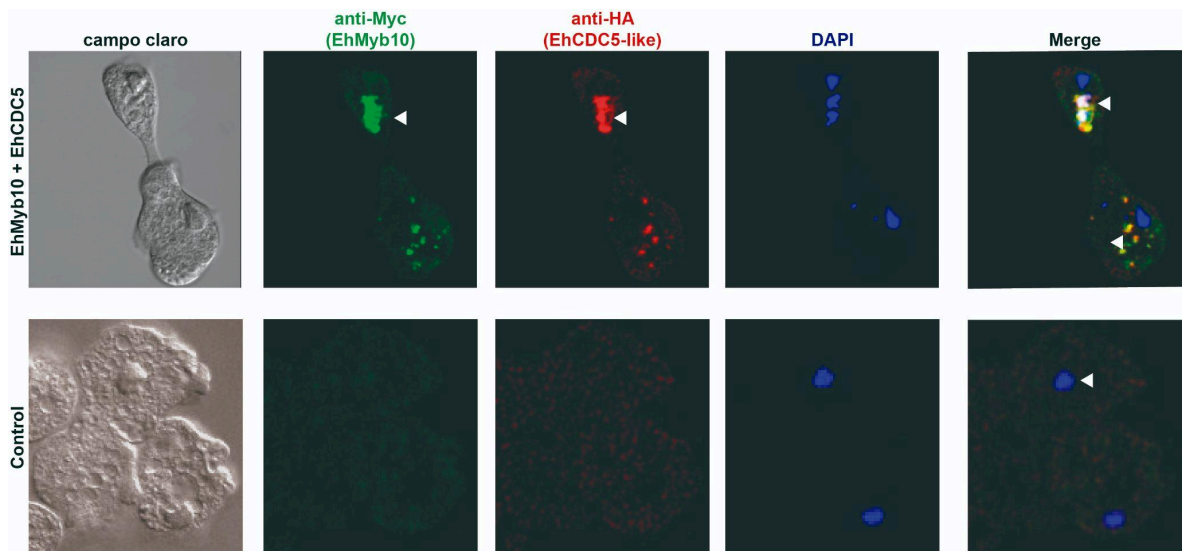
**Figura 28. Análisis de la expresión de los genes *ehmyb10* y *ehcdc5*.** A partir de 50 ng de RNA tratado DNase obtenido de los trofozoítos transfectados con los plásmidos pEhEx, pKT3M-*ehmyb10* y pEhEx-*ehcdc5* cosechados a las 72 hpt se sintetizó cDNA, el cual se empleó en ensayos de RT-qPCR. Se calculó el fold change empleando como control el vector vacío y el gen *s2 40s* empleando el método de doble delta Ct.

### 9.2.3. Expresión y localización de las proteínas de EhMyb10 y EhCDC5 en los trofozoítos transfectados de *E. histolytica*

Para determinar la presencia, localización subcelular y colocalización de las proteínas EhMyb10 y EhCDC5 en los trofozoítos transfectados en las diferentes condiciones, se realizaron ensayos de inmunofluorescencia. Para ello, los trofozoítos transfectados se cosecharon a las 24, 48 y 72 hpt. En estos ensayos se observa que la proteína EhMyb10 es abundante en los núcleos de los trofozoítos al detectarla con el anticuerpo que reconoce el *tag* de Myc, el cual reconoce a la proteína heteróloga. En el caso de la proteína EhCDC5 se reconoce al *tag* de HA (el cual reconoce también a la proteína heteróloga) se observó una tinción abundante en el citoplasma y en algunos núcleos (Figura 29). Sin embargo, sólo a las 72 hpt se colocalizaron a las proteínas EhMyb10 y EhCDC5 en *focis* en el núcleo de los trofozoítos cotransfectados (Figura 30).



**Figura 29. Localización subcelular de las proteínas EhMyb10 y EhCDC5 en trofozoítos transfectados.** Los trofozoítos se transfectaron con los plásmidos pEhEx (control), pKT3M-*myb10* y pEhEx-*ehcdc5* y se cosecharon a las 72 hpt. Posteriormente, los trofozoítos se colocaron y fijaron en cubreobjetos y se incubaron con los anticuerpos primarios: anti-myc y anti-HA seguido de los anticuerpos secundarios anti ratón-FITC o anti-conejo-ALEXA 647. Los núcleos se contrañieron con DAPI. Las preparaciones se analizaron por microscopía confocal.

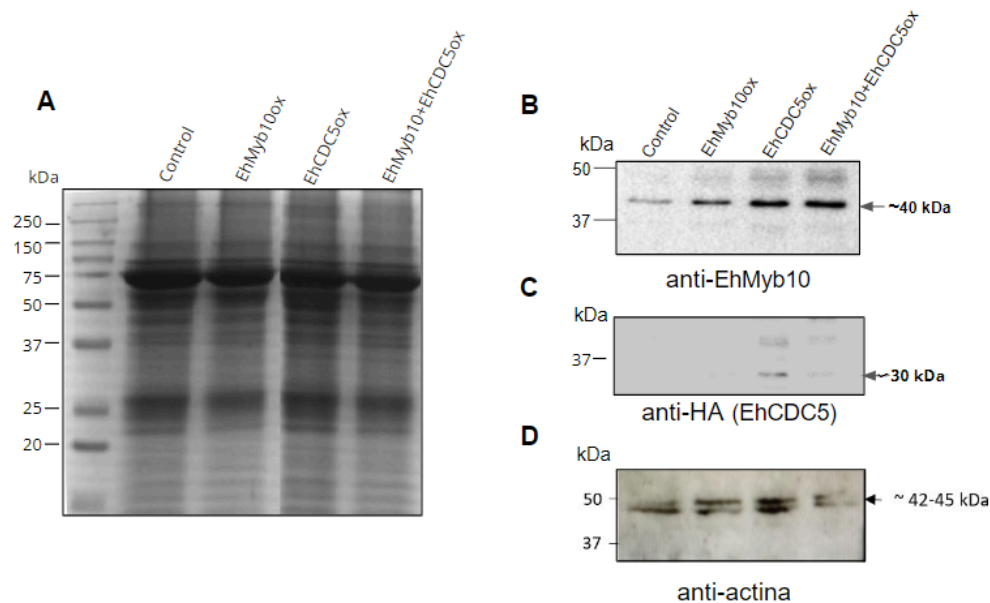


**Figura 30. Localización subcelular de las proteínas en la cotransfección (EhMyb10+EhCDC5) en trofozoítos 72 hpt.** Los trofozoítos se transfectaron con los plásmidos pKT3M-Myb10 y pEhEx-EhCDC5 y se cosecharon a las 72 hpt. Los trofozoítos se colocaron y fijaron en cubreobjetos y se incubaron con los anticuerpos primarios: anti-myc y anti-HA seguido de los anticuerpos secundarios anti ratón-FITC o anti-conejo-ALEXA 647. Los núcleos se contrañieron con DAPI. Las preparaciones se analizaron por microscopía confocal.

#### 9.2.4. Expresión de la proteína de EhCDC5 en los trofozoítos transfectados de *E. histolytica*

Una vez observada la localización y colocalización de las proteínas EhMyb10 y EhCDC5, se realizó un ensayo de inmunodetección. Para ello, se analizaron

extractos totales de trofozoítos a las 72 hpt en las diferentes condiciones (Control, EhMyb10ox, EhCDC5ox y cotransfección) ya que en los ensayos de inmunofluorescencia es donde se detectaba la mayor colocalización entre ambas proteínas. Las proteínas obtenidas se separaron en un gel SDS-PAGE y se tiñeron con azul de Coomassie, mostrando proteínas de alto y bajo peso molecular y una carga homogénea de los extractos en todas las condiciones (Figura 31A). Para detectar a las proteínas EhMyb10, EhCDC5 y actina se realizó una inmunodetección utilizando los anticuerpos anti-EhMyb10 (anticuerpo contra la proteína EhMyb10 recombinante obtenido en conejo), el anticuerpo anti-HA (el cual detecta la etiqueta HA fusionada a EhCDC5) y el anticuerpo anti-actina (control de carga). La proteína EhMyb10 se inmunodetectó en una banda de ~40 kDa presente en las muestras de todas; sin embargo, esta se observa con una mayor expresión en todas las condiciones cuando se compara con la condición del control (Figura 31B). La proteína EhCDC5 se inmunodetectó sólo en los trofozoítos de la condición EhCDC5ox y en la condición de cotransfección. EhCDC5 se inmunodetectó con un peso molecular de ~30 kDa lo cual corresponde al peso molecular teórico de esta. De manera interesante, también se inmunodetectó una banda de ~60 kDa que podría corresponder a un dímero de la misma proteína (Figura 31C). En el caso de la proteína actina, esta se inmunodetectó con un peso molecular de ~42-45 KDa presentando una abundancia similar en todas las condiciones. Esto nos corrobora que las diferencias observadas en las diferentes condiciones no son debido a la carga sino a la presencia o a la abundancia de las proteínas (Figura 31D).



**Figura 31. Inmunodetección de la proteína EhMyb10 y EhCDC5 en trofozoítos transfectados.** A) Extractos de trofozoítos transfectados con los plásmidos pEhEx (control), pKT3M-Myb10 y pEhEx-EhCDC5, después de 72 hpt separados en un gel SDS-PAGE al 10 % y teñidos con azul Coomassie. B) Inmunodetección de EhMyb10 (anti-EhMyb10), C) EhCDC5 (anticuerpo anti-HA-HRP), D) actina (anti-actina).

### 9.3. Efecto de la sobreexpresión de EhMyb10 y EhCDC5 sobre los patrones de *splicing in vivo* de *E. histolytica*

Para analizar el efecto de la sobreexpresión de las proteínas EhMyb10 y EhCDC5 en los patrones de *splicing*, se seleccionaron 4 posibles genes blancos a partir del análisis *in silico* previamente descrito. Los genes seleccionados fueron: 1) El gen *hmg* tiene un intrón de 67 pb y codifica para una proteína de unión a DNA, la cual participa en procesos relacionados con la cromatina, la transcripción y la replicación, recombinación, reparación del DNA y estabilidad genómica (Štros, 2010), 2) el gen *Swi5-Sfr* tiene un intrón de 48 pb y codifica para una proteína que es un activador de la recombinasa Rad51, que media el intercambio de cadenas de DNA en la recombinación homóloga. 3) un gen que codifica para una proteína “zinc knuckle-domain arginine/serine-rich” el cual tiene dos intrones uno de 44 pb y otro de 51 pb, que se une a RNA (Shepard y Hertel 2009). 4) El gen que codifica para una “U1 small nuclear ribonucleoprotein” el cual tiene un intrón de 154 pb, se une a la región exón-intrón en el extremo 5' del pre-mRNA, desempeña un papel crucial en una etapa temprana del *splicing* (Kattahy *et al.*, 2010).

Para ello, se diseñaron oligos que flanquean los intrones y así identificar las formas procesadas y no procesadas, de manera que se pueda determinar si el mRNA presenta RI. Para ello, se extrajo RNA total de los trofozoítos transfectados

en las diferentes condiciones a las 72 hpt, y posteriormente se realizó una RT-PCR semicuantitativa. Los productos de PCR, se separaron en geles de agarosa y se analizaron las diferentes bandas amplificadas. Las bandas de menor tamaño corresponden a los transcritos procesados predichos y las de mayor tamaño están relacionadas con transcritos no procesados que retienen el intrón (RI). Las bandas amplificadas se cuantificaron por densitometría para calcular el porcentaje de RI para cada uno de los genes (% RI) de acuerdo con de la Mata y cols. (2010).

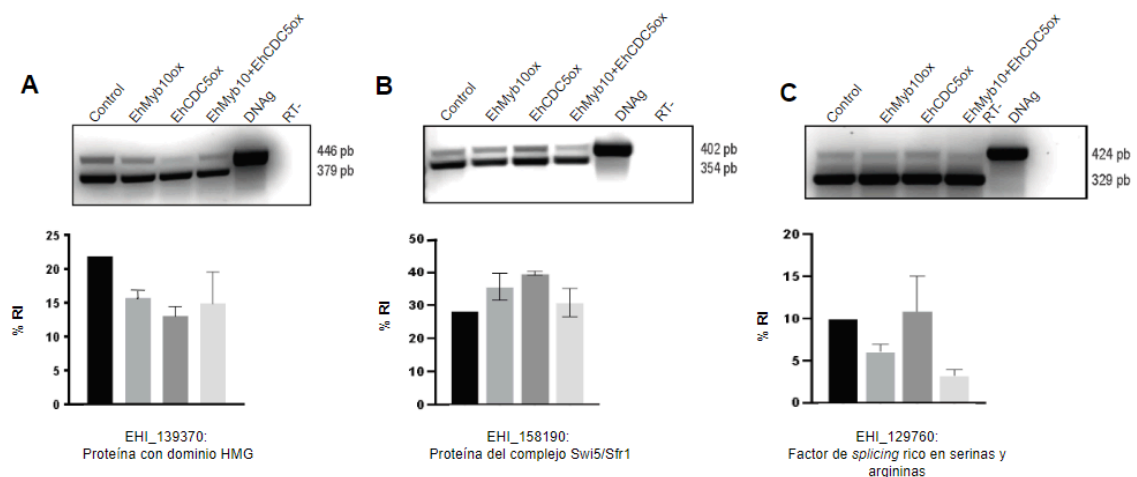
El gen EHI\_139370 (que codifica para una proteína con dominio HMG), en la condición control, la RI corresponde al 22 %. Sin embargo, aumenta su procesamiento, cuando se sobreexpresa EhMyb10, ya que su RI es de 15.7 %. Este efecto es más notorio en la condición EhCDC5ox ya que el porcentaje disminuye a 12 %. Finalmente, en la condición de cotransfección se observa un aumento en la RI ya que aumenta la intensidad de la banda que retiene el intrón. Estos resultados nos permiten proponer que EhMyb10 y EhCDC5 podrían participar en el procesamiento del gen EHI\_139370. Sin embargo, este efecto no se potencia cuando están presentes ambas proteínas (Figura 32A).

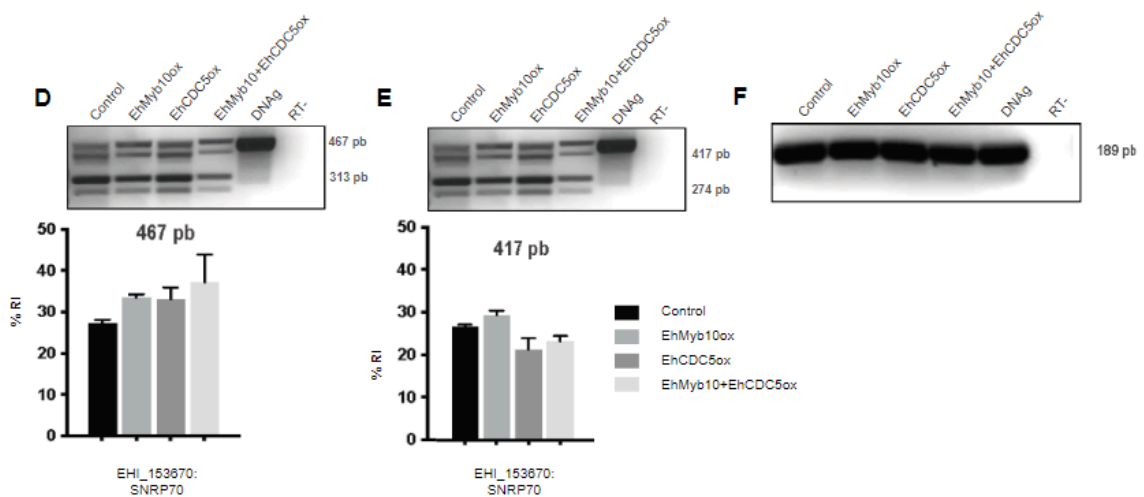
Cuando analizamos el procesamiento del gen EHI\_158190 (proteína del complejo Swi5-Sfr1) en la condición control, la retención intrónica corresponde al 28%. Sin embargo, en las condiciones EhMyb10ox y EhCDC5ox la RI aumentó 35.6 % y a 39.5 %, respectivamente. De manera interesante, en la condición de cotransfección aumentó el procesamiento del gen en comparación con EhMyb10ox y EhCDC5ox de manera independiente (RI 30 %). Este resultado nos sugiere que ambas proteínas podrían ser importantes para el procesamiento del gen EHI\_158190 (Figura 32B).

Para el gen EHI\_129760 (*zinc Knuckle-domain*/Factor de *splicing* rico en argininas y serinas) en la condición control la RI es del 10 %. En la condición de EhMyb10ox la RI disminuye a 6 %, en la condición de EhCDC5ox presentó un comportamiento similar al control con una RI del 10 %. Sin embargo, cuando se sobreexpresan a las dos proteínas se observa un mayor procesamiento del transcrito con respecto a todas las demás condiciones (RI 3 %). Este resultado nos sugiere que para el gen EHI\_129760 existe un efecto colaborativo entre ambas proteínas (Figura 32C).

De manera interesante, para el gen EHI\_153670 (SNRP70) esperábamos sólo la amplificación de dos bandas: una de 467 pb que corresponde a la forma no procesada (con intrón) y 313 que corresponde al mRNA maduro. Sin embargo, al realizar la PCR utilizando los oligos diseñados, obtuvimos 2 bandas adicionales de 417 y 274 pb en todas las condiciones. Para el análisis de RI se tomaron en cuenta solamente las bandas de 467 y 417 pb como formas no procesadas del mensajero del gen EHI\_153670. La banda de 274 pb no se consideró para el análisis, ya que al analizarla por secuenciación observamos que parte del exón también es procesado. Para la banda de 467 pb, esta incrementa su intensidad en todas las condiciones con respecto al control, esto sugiere una mayor RI (figura 32D). En cuanto a la banda de 417 pb, el procesamiento es muy semejante con respecto al control. La condición de EhMyb10ox es la que presenta una RI de 33 % RI. La condición de cotransfección presenta un menor porcentaje de RI (25 %) (figura 32E).

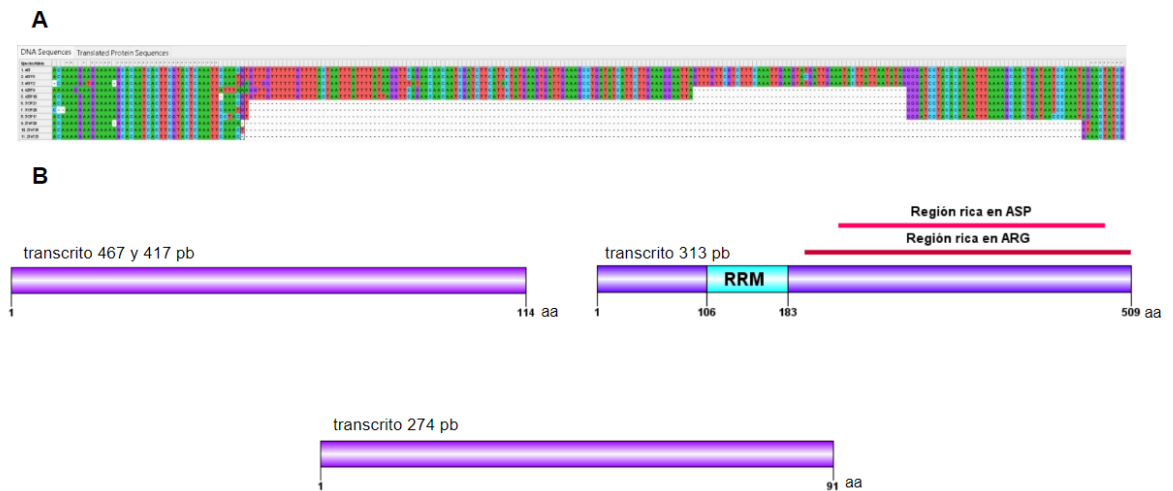
En todos los casos observamos que la sobreexpresión de ambas proteínas sobre el *splicing* es distinta al efecto de cada una por separado, indicando que hay un efecto correlativo (positivo o negativo) entre ambas proteínas sobre el *splicing* de los distintos transcritos probados.





**Figura 32. Análisis de la retención intrónica.** En la parte superior de cada figura se muestra el gel de agarosa al 2 % teñido con RedGel del análisis de RI y en la parte inferior se muestra el gráfico de % RI de cada condición para los genes seleccionados por medio de un análisis densitométrico utilizando ImageJ. A) Gen que codifica para la proteína con dominio HMG, B) Gen que codifica para la proteína del complejo Swi5/Sfr1, C) Gen que codifica para el factor de splicing rico en serinas y argininas, D) Gen que codifica para la proteína SNRP70 banda de 467 pb, E) Gen que codifica para la proteína SNRP70 banda de 417 pb.

Las 4 bandas obtenidas para el gen *snrp70* se clonaron en el vector pGEM-Teasy y se secuenciaron para corroborar la identidad con el gen. Las secuencias obtenidas se compararon con las contenidas en la base de datos de AmoebaDB, interesantemente encontramos que corresponden a diferentes isoformas de este gen (Figura 33A), de las cuales una de ellas no se encuentra anotada en esta base de datos. Posteriormente, hicimos una traducción *in silico* de los transcritos y determinamos el marco abierto de lectura de las proteínas que se sintetizarían de estas isoformas mediante Expasy translate (<https://web.expasy.org/translate>) y Prosite (<https://prosite.expasy.org>). La banda de 417 pb, retiene 105 pb del intrón y corresponde a la isoforma no anotada. De ella se sintetizaría una proteína de 114 aminoácidos, al igual que la isoforma que corresponde a la banda de 467 pb y que retiene todo el intrón de 154 pb. La banda de 270 pb corresponde a una isoforma en la que se remueven el intrón y parte del segundo exón y codificaría a una proteína de 91 aminoácidos. Finalmente, la banda 313 pb corresponde a la isoforma más abundante que codifica para una proteína de 509 aminoácidos que corresponde a la proteína SNRP70 completa (Figura 33B).



**Figura 33. Isoformas del gen que codifica para SNRP70.** A) Alineamiento de las secuencias del gen *snrp70* obtenidas mediante secuenciación y comparada con la secuencia obtenida de la base de datos AmoebaDB. B) Esquema de las isoformas de las proteínas a partir de los diferentes transcritos del gen *snrp70*. Azul el dominio de unión a RNA, línea roja región rica en arginina, línea rosa región rica en asparagina.

#### 9.4. Efecto de la sobreexpresión de EhMyb10 y EhCDC5 en la regulación de la expresión génica de *E. histolytica*

##### 9.4.1. Análisis de calidad e integridad del RNA total

Se realizó un análisis de calidad e integridad del RNA total mediante el mini Fluorómetro Qubit y posteriormente en el IBt-UNAM se analizó la integridad mediante Bioanalizador (Tabla 6) obteniendo el valor RIN (RNA Integrity Number) ~ 8 (TruSeq RNA Sample Preparation v2 Guide). Con estos resultados se determinó que las muestras tenían la integridad y calidad adecuadas para la generación de las genotecas para RNAseq.

**Tabla 6. Análisis de calidad de RNA total**

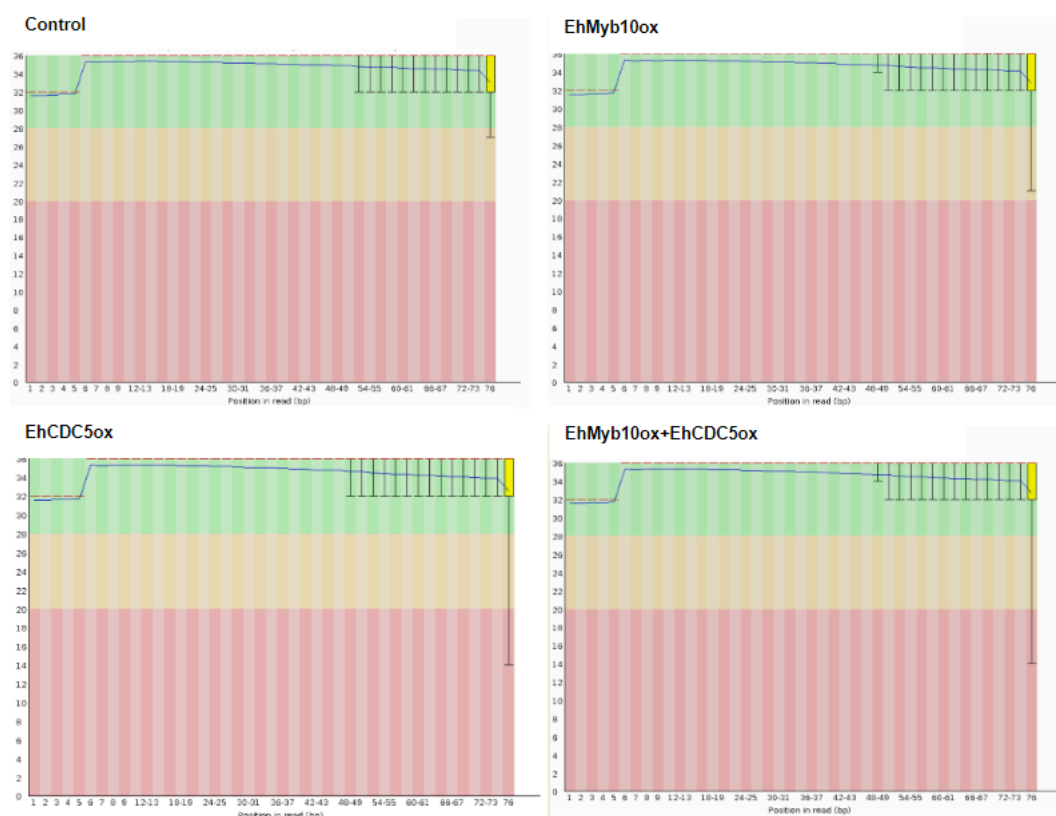
Transfectantes (72 hpt)	pEhEx	EhMyb10	EhCDC5	EhMyb10+EhCDC5
Valor RIN (Qubit)	8.9	8	7.5	6.3

Bioanalizador

#### 9.4.2. Análisis de calidad de las lecturas

El proceso de generación de las genotecas y secuenciación se llevó a cabo en el IBt-UNAM. Se obtuvieron lecturas pareadas de 2 x 75 con una calidad  $\sim Q \geq 30$  en la mayoría de las lecturas (Figura 34) Se llevó a cabo el mapeo al transcriptoma de referencia, obteniendo una cobertura de  $\sim 80\%$  en todas las condiciones (Tabla 7), este valor es común en células eucariotas (Sims *et al.*, 2014).



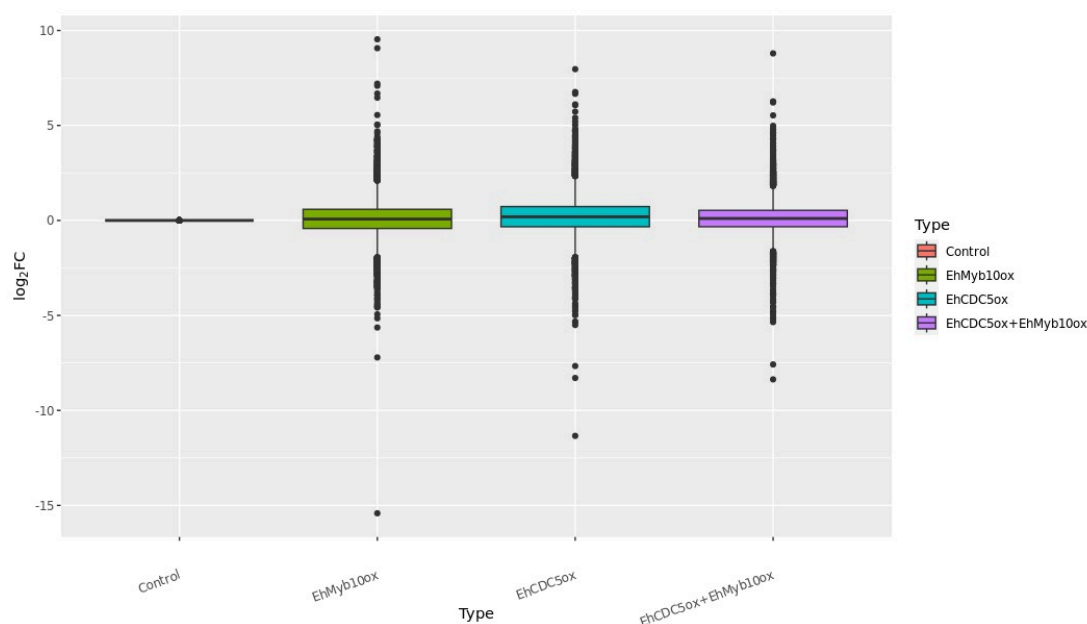
**Figura 34. Gráficas de calidad de las lecturas obtenidas de RNAseq.** Calidad de las lecturas obtenidas de RNAseq del equipo NextSeq 500 mediante el Software FastQC de algunas muestras (Control, EhMyb10, EhCDC5 y EhMyb10+EhCDC5).

**Tabla 7. Cobertura del transcriptoma de *E. histolytica*.**

Transflectante	% de cobertura (transcriptoma)
Control	87.91
EhMyb10ox	79.98
EhCDC5ox	83.85
EhMyb10+EhCDC5ox	86.85

#### 9.4.3. Análisis de la sobreexpresión de EhMyb10ox y EhCDC5ox

Un análisis preliminar de los 8201 transcritos de *E. histolytica* mediante un diagrama de cajas y bigotes permitió observar de manera general el efecto de la sobreexpresión de EhMyb10, EhCDC5 y la cotransfección (EhMyb10+EhCDC5), la mediana del transcriptoma de EhCDC5 se observa ligeramente arriba con respecto a las otras dos condiciones pero de manera general el transcriptoma en las tres condiciones pareciera tener el mismo comportamiento, un FC (Fold Change) cercano a 10 de los transcritos más expresados y un FC cercano a -.10 para los transcritos más reprimidos (Figura 35).



**Figura 35. Diagramas de caja y bigotes de las lecturas obtenidas de RNAseq.** Distribución de los transcritos de cada condición de acuerdo al FC con respecto al control.

Los análisis de RNAseq mediante DESeq2 mostraron que de los 8201 genes anotados en la base de datos AmoebaDB, alrededor de 4600 genes modificaron su expresión, sin diferencias significativas en el número de genes (4652-4693). Sin embargo, cuando se seleccionaron los genes que modificaron su expresión con un FC de 1.5 y con valor  $p < 0.05$  se observó que hubo una mayor y menor cantidad de genes diferencialmente expresados en los trofozoítos que sobreexpresan EhCDC5 y en la condición de cotransfección respectivamente (Tabla 8). De manera interesante y en general hubo mayor represión en todas las condiciones. Al realizar una correlación entre los genes que modificaron su expresión ya sea sobreexpresados o reprimidos con respecto a los genes con intrones tenemos una correlación de entre ~20-28 % en todos en todas las condiciones.

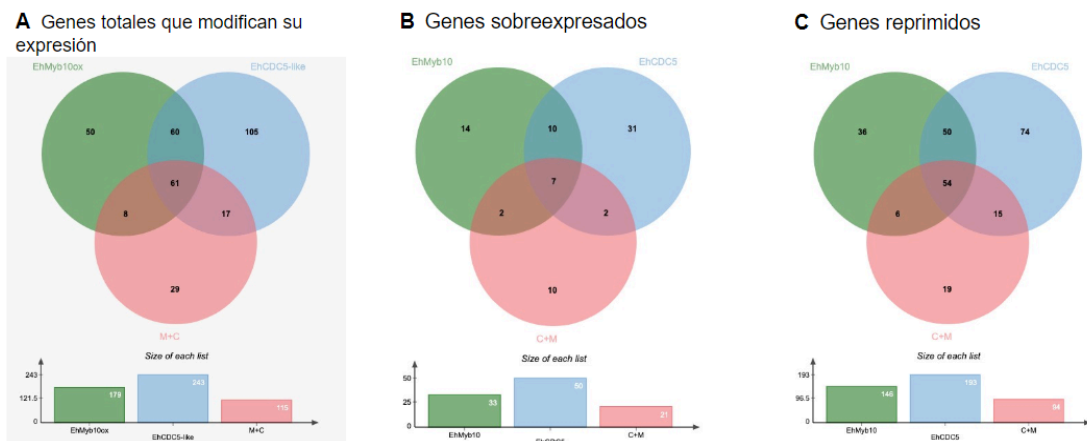
**Tabla 8. Genes que modifican su expresión en los trofozoítos que sobreexpresan a las proteínas EhMyb10, EhCDC5 y EhMyb10+EhCDC5.**

Condición	Genes que modificaron su expresión	*Genes que modificaron su expresión	*Genes sobreexpresión (con intrón)	Genes reprimidos (con intrón)
EhMyb10ox	4693	179	33 (8)	146 (39)
EhCDC5ox	4754	243	50 (10)	193 (52)
EhMyb10+EhCDC5ox	4652	115	21 (6)	94 (21)

\* Criterios de selección:  $p$ -value  $< 0.05$  Fold Change:  $> \pm 1.5$

Posteriormente, se analizaron los genes que son compartidos o son exclusivos en todas las condiciones analizadas. Mediante diagramas de Venn se pueden

visualizar los genes exclusivos de cada condición de sobreexpresión (Figura 36). La condición de sobreexpresión de EhCDC5 es la que tiene una mayor cantidad de genes que modifican su expresión (que se sobreexpresan y reprimen), seguido de la condición de EhMyb10 y la condición de cotransfección (EhMyb10+EhCDC5).



**Figura 36. Genes diferencialmente expresados en los trofozoítos que sobreexpresan las proteínas EhMyb10, EhCDC5 y EhMyb10+EhCDC5.** Diagramas de Venn de: genes totales que modificaron su expresión (A), genes que se sobreexpresaron (B) y genes reprimidos (C).

#### 9.4.4. Análisis de los transcritos que modifican su expresión en EhMyb10 y EhCDC5

Se realizó un análisis ontológico de los genes que modificaron su expresión en cada condición (Figura 37-39). Además, se muestran las tablas con los 10 genes con mayor y menor expresión de los genes que modificaron su expresión (Tabla 9-11).

Debido a que muchas proteínas de *E. histolytica* aún no están caracterizadas, en algunos casos se buscó la predicción de la función (anotación) por medio de ortología con ayuda del software HHpred y la base de datos AmoebaDB <https://toolkit.tuebingen.mpg.de/tools/hhpred>, <https://amoebadb.org/amoeba>.

#### 9.4.5. Condición de sobreexpresión de EhMyb10

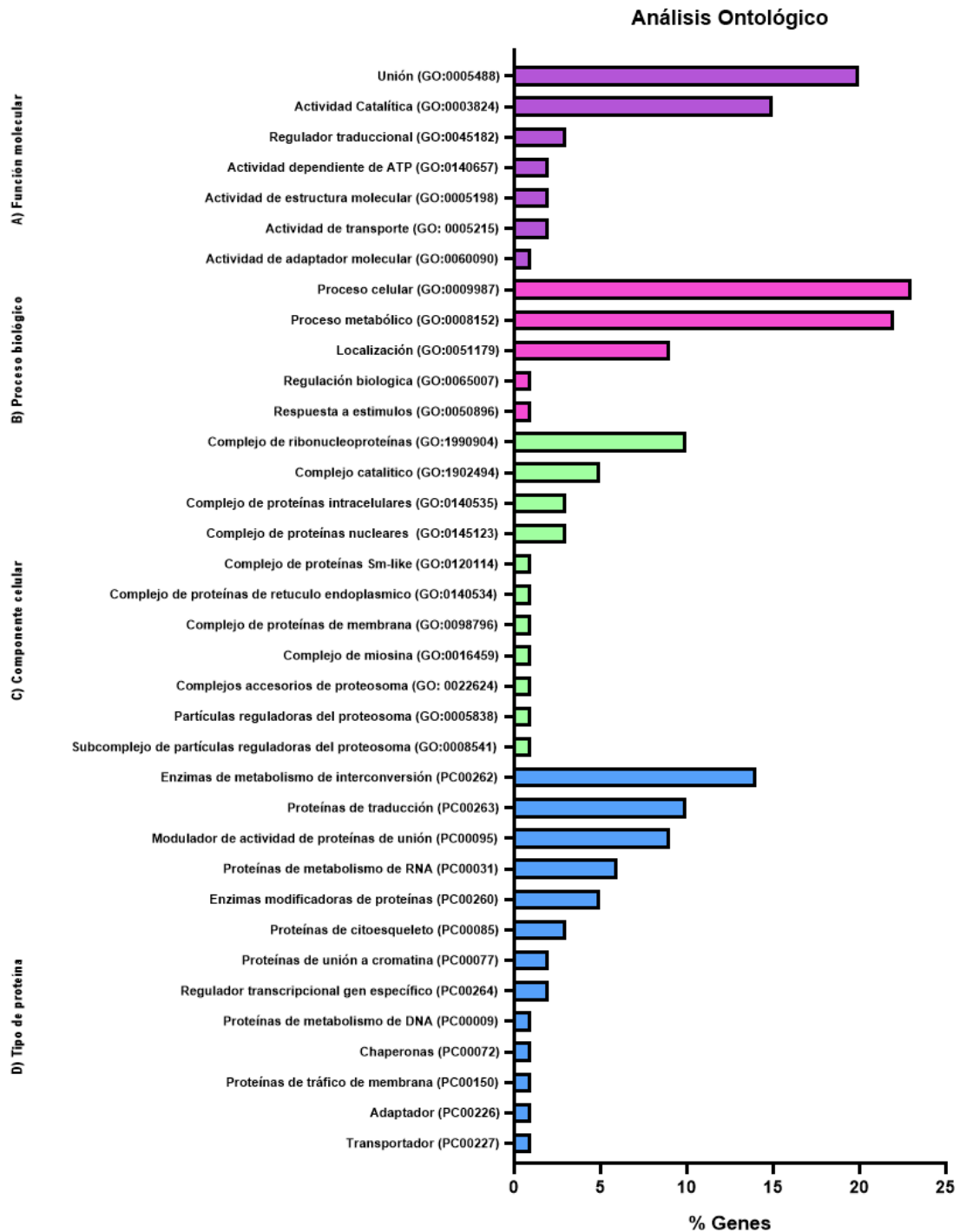
Cuando se sobreexpresó a la proteína EhMyb10 se modificó la expresión de 179 genes, de los cuales 33 genes se sobre-expresaron y 146 se reprimieron. De los 33 con sobreexpresión 8 contienen intrones y de los 146, 39 genes contienen intrón. Posteriormente, se hizo el análisis ontológico similar al que se realizó en el análisis *in silico*, eliminando las proteínas no clasificadas (hipotéticas), tomando

en cuenta las categorías de función molecular, proceso biológico, componente celular y tipo de proteína.

En el caso de la función molecular, tenemos 20 genes que participan en unión (GO:0005488), donde se encuentra la proteína ribosomal (60S ribosomal protein L18, putative)(FC: -14.57), el factor de *splicing* U2 (auxiliary factor, putative) y una chaperona (Chaperone proteinDNAJ, putative)(FC: - 8.42), un gen que codifica para una proteína con un motivo de reconocimiento a RNA. En actividad catalítica se encontraron 15 genes (GO:0003824) dentro de las cuales encontramos una aminoacil sintetasa ligase. En actividad de transporte (GO:0005215) se encontró una ATPasa Vacuolar. En proceso biológico se encontró un mayor número de genes, 23 genes involucrados en procesos celulares (GO:0009987) donde encontramos una proteína de metabolismo de RNA (DDB1- and CUL4-associated factor 13, protein SOF1)(FC: 5.96) y una nucleotidil transferasa (S-adenosylmethionine synthase (SAM)(FC: 12.82). En proceso metabólico (GO:0008152) tenemos 22 genes donde encontramos un factor de procesamiento de RNA, dos proteínas de metabolismo de RNA y dos factores de *splicing*. En el caso de componente celular, entramos a la subcategoría de complejos proteicos (GO:0032991), encontramos 10 genes en complejo de ribonucleoproteínas donde encontramos un factor de procesamiento de RNA y dos factores de *splicing*. Se encontraron dos proteínas de metabolismo de RNA ya antes mencionadas. En el caso de clase de proteínas, se encontraron 14 genes en enzimas de metabolismo de interconversión (PC00262), 10 genes para proteínas de traducción, 9 genes para proteínas de unión y 6 proteínas de metabolismo de RNA ya antes mencionados. En proteínas de citoesqueleto encontraron 3 Calmodulin expresadas con un FC -1.66, -2.23 y -1.96, respectivamente. En proteínas de unión a cromatina (PC00077) se encontraron histonas H2A reprimidas (FC: -1.8 y -2.1) y de regulación transcripcional del gen específico (PC00264) se encontraron dos genes que codifican para proteínas EhMyb reprimidas (FC: -1.6 y -1.9). En la clasificación de modulador de actividad de unión a proteína (PC00095) encontramos proteínas AIG, Rab family GTPase reprimidas.

En cuanto a los genes con mayor expresión, se encontraron 7, de estos anotados como hipotéticos, por lo cual se tuvo que buscar su anotación por ortología. De los 10 genes con mayor sobreexpresión, 4 fueron exclusivos de la condición (EhMyb10 ox) (BIG1, Yos1-like, SOF1 y Metaloproteasa), 3 genes fueron también

sobreexpresados en EhCDC5ox (SAM, Chaperona 4 y Fumarato hidratasa), 2 genes con la condición de cotransfección (EhMyb10+EhCDC5ox) (fosfopanteteína adenililtransferasa (PPAT) y una fosfatasa) y 1 gen fue sobreexpresado en las 3 condiciones (Tetraspanin family protein). En los genes reprimidos encontramos 4 exclusivos de EhMyb10ox una chaperona, una proteína con dominio Myb, un factor de *splicing* y una riboflavin cinasa, 4 genes compartidos con la condición de EhCDC5ox la proteína ribosomal L18, una hidratasa fumarato, una proteína de unión a C-Myc y una cortexilin II, y finalmente dos genes compartidos con EhCDC5ox y EhMyb10+EhCDC5ox dos proteínas relacionadas al complejo 3 (AP-3)(Tabla 9).



**Figura 37. Análisis ontológico de los genes con expresión diferencial en trofozoítos transfectados con EhMyb10ox.** A través de Panther Gene Ontology Functional Enrichment. A) Función molecular, B) Proceso biológico, C) Componente celular y D) Tipo de proteína.

**Tabla 9. Genes diferencialmente expresados en EhMyb10ox.**

Gene ID	Fold Change	p value	Anotación*	# Exones	Condición
<b>Genes sobreexpresados en EhMyb10</b>					
EHI_195110	12.82	1.32E-03	S-adenosylmethionine synthetase, putative	1	EhCDC5
EHI_006770	7.37	2.21E-05	hypothetical protein, phosphopantetheine adenyltransferase**	2	EhMyb10+EhCDC5
EHI_123020	7.28	2.01E-05	hypothetical protein, protein phosphatase domain-containing protein **	1	EhMyb10+EhCDC5
EHI_018630	7.09	6.58E-05	hypothetical protein, ER membrane protein, BIG1 family**	1	EhMyb10ox
EHI_178640	6.68	1.80E-04	Yos1-like domain-containing protein	2	EhMyb10
EHI_083380	5.96	2.26E-03	hypothetical protein, protein SOF1, putative **	1	EhMyb10
EHI_109930	3.95	2.46E-03	hypothetical protein, Proteasome assembly chaperone 4,**	1	EhCDC5
EHI_159450	3.73	6.82E-03	hypothetical protein, METALLOPROTEASE; HYDROLASE**	1	EhMyb10
EHI_103400	3.34	9.77E-05	hypothetical protein, Tetraspanin family protein **	2	EhCDC5 y EhMyb10+EhCDC5
EHI_117270	3.12	3.69E-04	fumarate hydratase class I, anaerobic, putative	1	EhCDC5
<b>Genes reprimidos en EhMyb10</b>					
EHI_098470	-14.57	3.36E-05	ribosomal protein L18, putative	1	EhCDC5
EHI_035060	-9.57	4.94E-10	fumarate hydratase class I, anaerobic, putative	1	EhCDC5
EHI_155900	-9.25	9.89E-09	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit**	1	EhCDC5 y EhMyb10+EhCDC5
EHI_049990	-8.42	1.00E-07	chaperone protein DNAJ	1	EhMyb10
EHI_004430	-8.36	1.26E-07	hypothetical protein, Myb-like domain-containing protein**	1	EhMyb10
EHI_089770	-7.64	2.73E-06	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit**	1	EhCDC5 y EhMyb10+EhCDC5
EHI_183200	-7.28	1.90E-05	hypothetical protein, C-Myc-binding protein; conserved hypothetical protein**	2	EhCDC5
EHI_179850	-6.32	7.23E-04	hypothetical protein, Pre-mRNA-splicing factor CWC25 homolog**	2	EhMyb10
EHI_083140	-6.31	6.39E-06	hypothetical protein, cortaxillin II, putative **	2	EhCDC5
EHI_033240	-6.24	8.19E-04	Riboflavin kinase / FAD synthetase	1	EhMyb10

\*AmoebaDB

\*\*Con base los grupos Ortologos (**OrthoMCL DB**) y HHpred

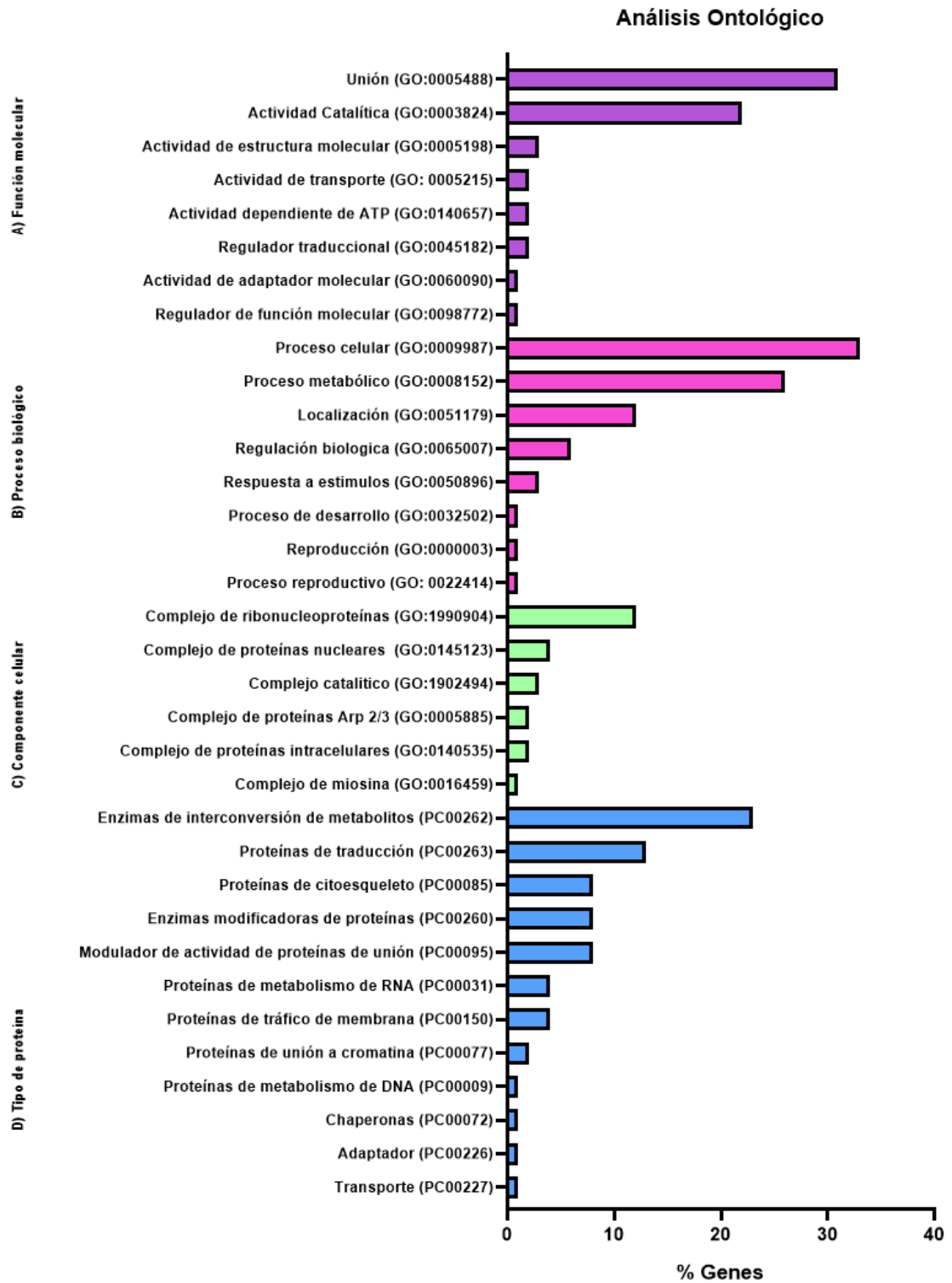
#### 9.4.6. Condición de sobreexpresión de EhCDC5

En la condición de sobreexpresión de EhCDC5 se encontraron 243 genes que modificaron su expresión, de estos, 50 se sobreexpresaron y 10 de estos contienen al menos un intrón, 193 genes fueron reprimidos y 52 de estos contienen al menos un intrón.

Al realizar el análisis ontológico para función molecular, en función de unión (GO:0005488), se encontraron 31 genes se encontraron 2 genes relacionados a unión a cromatina, Histone H2A con un FC -2.44 para ambas, 2 factores de *splicing* con un FC de -1.88 y -5.2 respectivamente. En actividad catalítica (GO:0003824) con 22 genes dentro de los cuales se encontró una SAM (FC: 12 .54) y varias GTPasas pequeñas reprimidas. En proceso biológico, con 33 genes en proceso celular (GO:0009987) y 26 genes en proceso metabólico (GO:0008152) y 12 en localización (GO:0051179). En proceso celular y proceso metabólico se identificó al gen que codifica para SAM, los factores de *splicing*, ya antes mencionados. En componente celular, en la subcategoría de complejos proteicos (GO:0032991), en complejos ribonucleoproteicos (GO:1990904) con 12 genes, se encontraron varias proteínas ribosomales, dos factores de *splicing*, ya antes mencionados, una proteína de metabolismo de RNA. En tipo de proteína la categoría donde se encontraron más genes fue en la de enzimas de interconversión (PC00262) con 23 genes, seguida de proteínas de traducción (PC00263) con 13 genes entre ellos varias proteínas ribosomales 40s y 60s expresadas negativamente, proteínas de citoesqueleto (PC00085) con 8 genes de igual manera hay varias proteínas relacionadas a actina que se expresan de manera negativa, proteínas de unión (PC00095) con varias GTPasas. En proteínas de unión a cromatina (PC00077) se encontró la sobreexpresión de un gen que codifica para una proteína EhMybSHAQKYF y un gen que codifica para una proteína de unión a nucleosoma.

Analizando los genes con mayor expresión en esta condición, se encontraron 6 genes exclusivos donde se encontró un gen que codifica para una ATPasa Vacuolar, una proteína con dominio LSM, proteína asociada a microtúbulos, un antígeno de superficie ariel1 y una proteína con dominio C2 y 4 genes compartidos con la condición de EhMyb10ox, SAM, una chaperona, una fumarato hidratasa y una P-glicoproteína. En los genes más reprimidos encontramos 4 genes compartidos con la condición de EhMyb10ox donde se encuentran genes

que codifican para una proteína ribosomal L18, una sintetasa, una hidratasa fumarato y una cortexilina, dos genes compartidos con la cotransfección, el gen que codifica para la proteína WD 68 y una proteína de la subunidad alfa de F-actina, dos genes compartidos con EhMyb10ox y la cotransfección, dos genes para AP-3 (Tabla 10). Dentro de genes reprimidos exclusivos encontramos dos genes que codifican para una proteína de filamina.



**Figura 38. Análisis ontológico de los genes con expresión diferencial en trofozoítos transfectados con EhCDC5ox a través de Panther Gene Ontology Functional Enrichment. A) Función molecular B) Proceso biológico C) Componente celular y D) Tipo de proteína.**

**Tabla 10. Genes diferencialmente expresados en EhCDC5ox.**

Gene ID	Fold Change	p value	Anotación	# Exones	Condición
<b>Genes sobreexpresados en EhCDC5</b>					
EHI_195110	12.54	9.81E-17	S-adenosylmethionine synthetase, putative	1	EhMyb10
EHI_118220	8.54	3.01E-07	vacuolar ATP synthase subunit B, putative	2	EhCDC5
EHI_025840	6.88	1.67E-04	LSM domain containing protein	2	EhCDC5
EHI_109930	4.42	3.76E-04	hypothetical protein, Proteasome assembly chaperone 3*	1	EhMyb10
EHI_117270	3.08	3.70E-02	fumarate hydratase class I, anaerobic, putative	1	EhMyb10
EHI_087650	2.82	1.37E-02	hypothetical protein, PROGRAMMED CELL DEATH 6-INTERACTING PROTEIN**	1	EhCDC5
EHI_050640	2.79	1.41E-03	P-glycoprotein 6, putative	1	EhMyb10
EHI_192420	2.72	2.31E-02	hypothetical protein, cilia, microtubule, dynein**	1	EhCDC5
EHI_101730	2.67	1.72E-02	surface antigen ariel1, putative	1	EhCDC5
EHI_197140	2.67	3.52E-02	hypothetical protein, C2 domain containing protein**	1	EhCDC5
EHI_103900	2.6	2.06E-11	nucleosome-binding protein 1, putative	1	EhCDC5 EhMyb10+EhCDC5
EHI_080100	2.53	7.7E-03	cylicin-2, putative	1	EhCDC5
EHI_068610	1.6	1.05E-04	hypothetical protein, pre-mRNA-splicing factor CWC25, putative*	1	EhMyb10+EhCDC5
<b>Genes reprimidos en EhCDC5</b>					
EHI_098470	-14.27	4.76E-05	ribosomal protein L18, putative	1	EhMyb10
EHI_094020	-10.42	6.42E-12	hypothetical protein, filamin/ABP280 repeat containing protein**	1	EhCDC5
EHI_000690	-9.97	1.11E-10	seryl-tRNA synthetase, putative	1	EhMyb10
EHI_035060	-9.27	1.23E-09	fumarate hydratase class I, anaerobic, putative	1	EhMyb10
EHI_155900	-8.96	1.73E-08	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit**	1	EhMyb10 y EhMyb10+EhCDC5
EHI_083140	-8.46	2.21E-03	hypothetical protein, cortexillin II, putative*	2	EhMyb10
EHI_140640	-7.43	1.10E-05	F-actin capping protein subunit alpha, putative	1	EhMyb10+EhCDC5
EHI_089770	-7.34	9.64E-06	hypothetical protein, AP-3 complex subunit delta-1, putative*	1	EhMyb10 y EhMyb10+EhCDC5
EHI_048300	-6.98	5.88E-06	WD repeat protein 68, putative	1	EhMyb10+EhCDC5
EHI_020990	-6.66	2.44E-04	hypothetical protein, Nucleoporin NSP1; nuclear pore**	1	EhCDC5

\*AmoebaDB

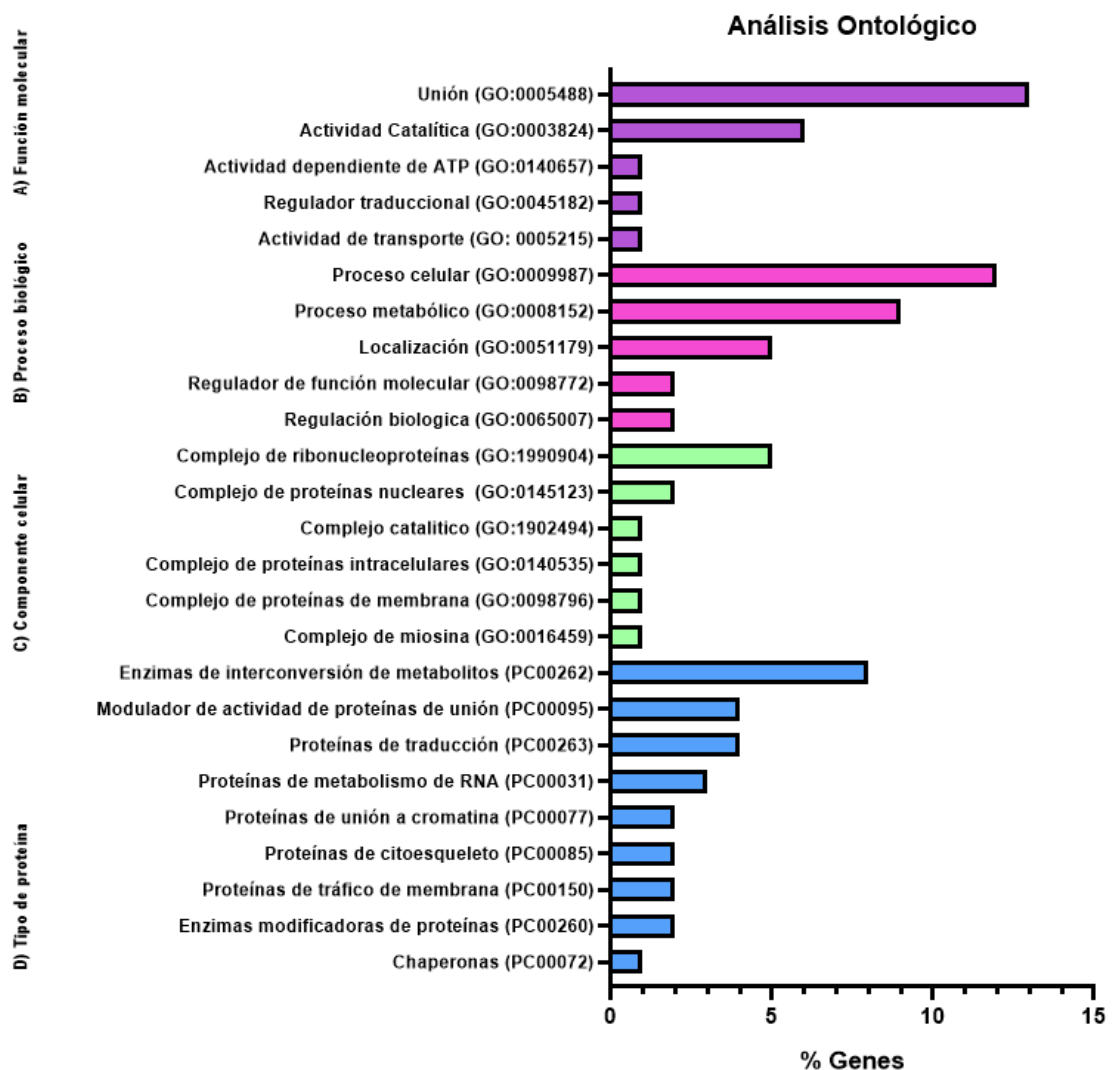
\*\*Con base los grupos Ortologos (**OrthoMCL DB**) y HHpred

#### 9.4.7. Condición de sobreexpresión de EhMyb10+EhCDC5

Para la condición que sobreexpresa las proteínas EhMyb10+EhCDC5 (cotransfección) 115 genes modificaron su expresión, 21 con sobreexpresión y 94 reprimidas. En el análisis ontológico en la categoría de función molecular, la mayoría de los genes tienen una función de unión (GO:0005488) con 14 genes, dentro de estos están dos Histone H2A (FC -1.56 y -1.62), el factor de *splicing* factor auxiliar u2 (FC: -2.22), Skp1 protein (FC: -2.31), seguido de actividad catalítica (GO:0003824) con 6 genes, principalmente dos GTPasas, una metil transferasa, una liasa y una ligasa. En actividad dependiente de ATP (GO:0140657) y actividad de transporte (GO:0005215) encontramos una Vacuolar ATP synthase subunit E (FC: -1.66). En proceso biológico encontramos 12 genes dentro de proceso celular (GO:0009987), entre ellos F-actin-capping protein subunit alpha (FC: -7.60) y U2 auxiliary factor (FC - 2.22), 9 genes involucrados en procesos metabólicos como una liasa methionine gama (Methionine gamma-lyase,MGL) (FC: 1.85). En la categoría de componente celular en la subcategoría complejos de proteínas (GO:0032991) se encontraron 5 genes en los complejos de ribonucleoproteínas (GO:1990904) donde encontramos nuevamente a U2 auxiliary factor y U3 small nucleolar ribonucleo protein (FC: -3.82), estas mismas fueron encontradas en complejos de proteínas nucleares (GO:0140513). En complejos de miosina (GO:0016459) encontramos el gen *calmodulin* (FC: -2.06). En clase de proteína encontramos 8 genes que codifican para enzimas de interconversión (PC00262), seguido con 4 genes que codifican para proteínas de unión (PC00095) donde encontramos proteínas AIG, Rab family GTPase reprimidas. En proteínas de metabolismo de RNA (PC00031) encontramos nuevamente U2 auxiliary factor y U3 small nucleolar ribonucleo protein. En proteínas de unión al citoesqueleto se encontró a la proteína F-actin capping protein subunit alpha con un FC -7.60.

De los genes que modificaron su expresión y que tuvieron una mayor expresión exclusivamente en la condición de cotransfección fueron 5 genes que codifican para diferentes proteínas, una proteína de unión a GTP, una metil transferasa, dos proteínas de antígeno de superficie ariel1 y una proteína mucin-like. Se encontraron dos genes compartidos con EhMyb10ox una adenil transferasa y una fosfatasa, dos genes compartidos con la condición de EhCDC5 una proteína de procesamiento de RNA y una proteína ribosomal 30S, finalmente una proteína

presente en las tres condiciones fue una tetraspanina putativa. En el caso de los genes con mayor represión exclusivos de la condición de cotransfección fueron 4 genes, un factor de inicio de la traducción una adenilil transferasa, un complejo AP-3 y finalmente una proteína PufX. Se encontraron dos genes compartidos con EhMyb10ox, una proteína con dominio WD y una proteína de la familia AIG1. Gen compartido con la condición de EhCDC5ox que codifica para la proteína de la subunidad alfa de F-actina. Se encontraron 3 genes en común en las tres condiciones, AP-3, prosap, isoform D y una proteína con dominio FMN reductasa (Tabla 11).



**Figura 39. Análisis ontológico de los genes con expresión diferencial en trofozoítos transfectados con EhMyb10+EhCDC5** a través de Panther Gene Ontology Functional Enrichment. A) Función molecular B) Proceso biológico C) Componente celular y D) Tipo de proteína.

**Tabla 11. Genes diferencialmente expresados en EhMyb10+EhCDC5.**

Gene ID	Fold Change	p value	Anotación	# Exones	Condición
<b>Genes sobreexpresados en EhMyb10+EhCDC5</b>					
EHI_006770	8.5823	5.68E-08	hypothetical protein, phosphopantetheine adenyltransferase *	2	EhMyb10
EHI_075040	7.7356	1.74E-11	GTP-binding protein	1	EhMyb10+EhCDC5
EHI_123020	7.4236	1.16E-05	hypothetical protein, protein phosphatase domain-containing protein *	1	EhMyb10
EHI_147990	3.9037	3.36E-03	hypothetical protein, Ribosomal RNA large subunit methyltransferase H **	1	EhMyb10+EhCDC5
EHI_109010	2.8553	2.49E-02	hypothetical protein, surface antigen ariel1, putative*	2	EhMyb10+EhCDC5
EHI_103400	2.7643	9.01E-03	hypothetical protein, Tetraspanin family protein**	2	EhMyb10 y EhCDC5
EHI_066970	2.5220	2.59E-02	mucin-like protein 1 precursor, putative	1	EhMyb10+EhCDC5
EHI_114000	2.2800	4.33E-02	hypothetical protein, Pecanex C-terminal domain-containing protein*	2	EhMyb10+EhCDC5
EHI_166860	2.2235	3.31E-02	hypothetical protein, RNaseIII, non canonical Dicer, RNA procesing **	1	EhCDC5
EHI_023400	2.1896	1.41E-04	Ribosomal protein S30, putative	1	EhCDC5
EHI_103900	2.1	1.047E-03	nucleosome-binding protein 1, putative	1	EhCDC5
<b>Genes reprimidos en EhMyb10+EhCDC5</b>					
EHI_155900	-9.1351	1.49E-08	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit **	1	EhMyb10 y EhCDC5
EHI_036630	-8.4061	1.20E-07	hypothetical protein, eukaryotic translation initiation factor 3 subunit H **	1	EhMyb10+EhCDC5
EHI_006680	-8.2397	1.95E-07	phosphopantetheine adenyltransferase, putative	2	EhMyb10+EhCDC5
EHI_140640	-7.6057	6.09E-06	F-actin capping protein subunit alpha, putative	1	EhCDC5
EHI_089770	-7.5219	4.22E-06	hypothetical protein, AP-3 complex subunit delta-1*	1	EhMyb10+EhCDC5
EHI_116040	-6.8317	7.34E-05	hypothetical protein, dab2-interacting protein*	1	EhMyb10 y EhCDC5
EHI_174150	-6.2802	6.54E-04	hypothetical protein, Photosynthetic reaction center PufX protein**	2	EhMyb10+EhCDC5
EHI_068380	-4.6311	1.16E-02	WD domain containing protein	1	EhMyb10
EHI_032670	-3.7174	1.57E-03	hypothetical protein, NADPH-dependent FMN reductase domain containing protein*	2	EhMyb10 y EhCDC5

EHI_180390	-3.5446	4.23E-02	AIG1 family protein, putative	1	EhMyb10
------------	---------	----------	-------------------------------	---	---------

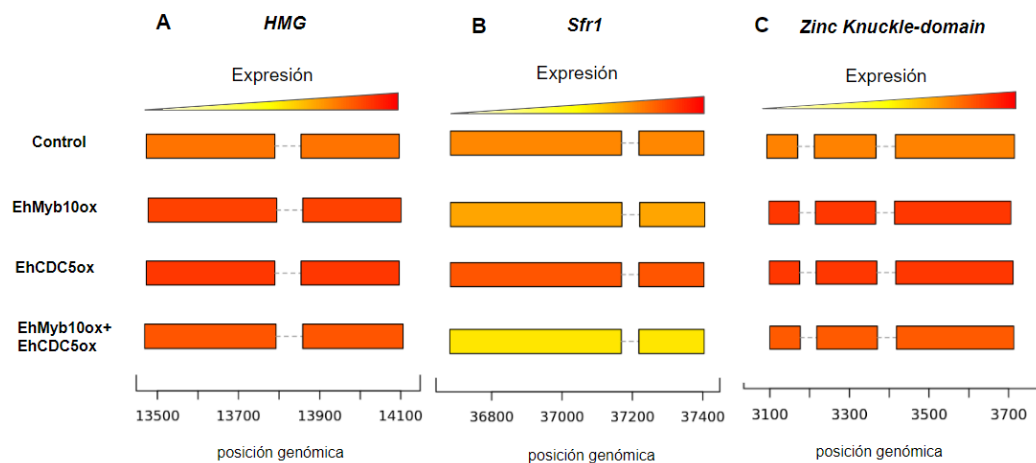
\*AmoebaDB

\*\*Con base los grupos Ortologos (**OrthoMCL DB**) y HHpred

#### 9.4.8. Análisis de expresión de los genes seleccionados por medio de análisis Ballgown

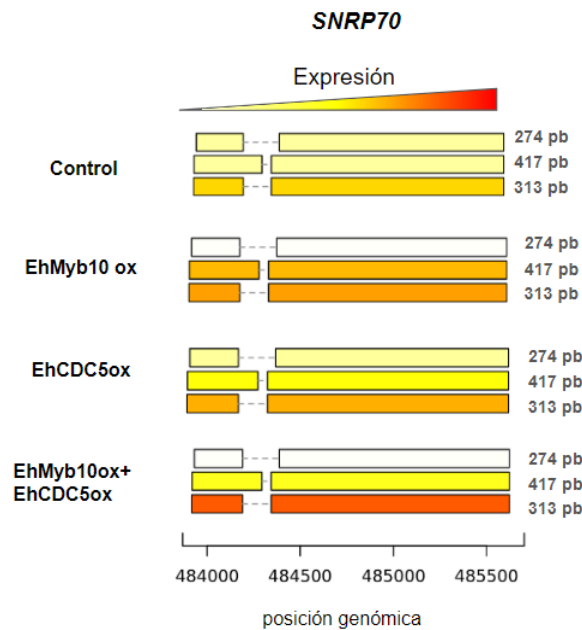
Una vez que se obtuvieron los datos del análisis de RNAseq y ya que los genes seleccionados en el análisis *in silico* inicial no se encontraron en las listas de los genes diferencialmente expresados. Procedimos a realizar un análisis de sus patrones de expresión empleando la herramienta de Ballgown. Esta herramienta permite analizar el ensamblaje del transcriptoma, visualizar la organización de los transcritos y las abundancias específicas de sus isoformas.

Primeramente, se analizó el gen *hmg* en las tres condiciones observándose un incremento en la cantidad de transcrito cuando se sobreexpresaron las proteínas EhMyb10 y EhCDC5 por separado, sin embargo, se observó una ligera disminución cuando las proteínas se coexpresan, sugiriendo una competencia entre ambas proteínas (Figura 40A). En el caso del gen *zinc knuckle-domain* se observó una mayor expresión cuando se sobreexpresa a la proteína EhCDC5, de manera interesante cuando las dos proteínas se sobreexpresaron se observó una menor cantidad del transcrito, lo cual apoya el efecto competitivo de las dos proteínas en este gen (Figura 40B). En el caso del gen *sfr1* se observó que hay un aumento en la expresión cuando se expresan las proteínas EhMyb10 y EhCDC5 por separado sin embargo en la sobreexpresión se observan una pequeña disminución de la expresión, lo cual podría deberse también a una competencia entre factores, aunque en este gen en menor proporción (Figura 40C). De manera interesante el tamaño de los intrones de estos genes se encuentra en el promedio del tamaño de los intrones para este organismo (~75 pb).



**Figura 40. Niveles de expresión de los transcritos de los genes *hmg*, *sfr1* y *zinc knuckle-domain*, procesados en cada condición.** Se realizó un análisis de los datos obtenidos de RNAseq mediante la herramienta Ballgown. Se muestran las diferentes isoformas de los transcritos y su nivel de expresión en cada condición (menor expresión (amarillo), mayor expresión (rojo)).

Finalmente, se analizó el gen *snrp70* el cual tiene un intrón de 154 pb (se considera un intrón grande), cuando se hizo el análisis por medio de RT-PCR se observaron 4 transcritos de diferente tamaño que corresponden a isoformas diferentes obtenidas a partir de sitios 5' s y 3' ss alternativos. Sin embargo, en el análisis de Ballgown solo se observan tres transcritos diferentes (Figura 32). La primera isoforma, el cual se genera debido a la presencia de sitios 5'y 3'ss alternativos, con la cual se obtiene un transcrito de 274 pb, esta isoforma disminuye cuando se sobreexpresa EhMyb10, indicado un posible efecto negativo de esta proteína. Cuando se sobreexpresa sólo a EhCDC5 no se ve un efecto sobre esta isoforma. Sin embargo, cuando se analiza la condición donde se sobreexpresan las dos proteínas el efecto negativo de EhMyb10 se mantiene. En el caso del transcrito de 417 pb, el cual se genera debido a la presencia de un sitio 5'ss alternativo, se observa un aumento de la expresión sólo en presencia de EhMy10. Finalmente, en el caso del transcrito de 313 pb, es el resultado de la eliminación del intrón completo se puede observar que hay un incremento en todas las condiciones analizadas, siendo el mayor efecto cuando ambas proteínas están presentes (cotransfección) (Figura 41). Lo cual nos podría sugerir que ambas proteínas tienen un efecto cotranscripcional con genes con intrones grandes y sitios de *splicing* alternativos.


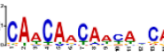






**Figura 41. Niveles de expresión de los transcritos del gen *snrp70* procesados en cada condición.** Se realizó un análisis de los datos obtenidos de RNAseq mediante la herramienta Ballgown. Se muestran las diferentes isoformas de los transcritos y su nivel de expresión en cada condición (menor expresión (amarillo), mayor expresión (rojo)).

#### 9.4.9. Motivos encontrados en los promotores de los genes que modificaron su expresión al sobreexpresar a las proteínas EhMyb10 y EhCDC5

Finalmente, se buscaron los motivos conservados en los promotores de los genes que modificaron su expresión al sobreexpresar a las proteínas EhMyb10 y EhCDC5 mediante la herramienta STREME de la suite de MEME. Para ello, se obtuvieron las secuencias de los promotores (-500 y +50 pb río arriba del INR) de los genes que modificaron su expresión en las diferentes condiciones de la base de datos AmoebaDB. Este análisis identificó secuencias que conservan un core AAC, el cual se ha demostrado que es importante para el reconocimiento de las proteínas Myb al ERM (Tanikawa *et al.*, 1993) (Tabla 12). De manera importante, se identificó una secuencia que se encuentra presente en todas las condiciones analizadas, la cual corresponde al motivo: CAACAACAAC.

**Tabla 12. Motivos conservados en los genes diferencialmente expresados.** Motivos encontrados en los promotores de los genes que modificaron su expresión en los trofozoítos que sobreexpresión a EhMyb10, EhCDC5 y EhMyb10+EhCDC5.

Secuencia	Logo	Significancia (p-value)	Tamaño (nt)	# Genes donde aparece
<b>EhMyb10</b>				
AAAGAAATACCA		3.6e -003	12	81
CAACAACAACATCA		3.3 e -001	15	77
<b>EhCDC5</b>				
AGAAGAAGAA		3.0 e -004	10	88
CAACAACAACA		2.4 e -002	11	130
AATAACAAAAAAA		4.1 e -002	13	237
<b>EhMyb10+EhCDC5</b>				
AACAACAAC		2.5 e -001	10	33

## 10. Discusión

La regulación cotranscripcional se refiere a los eventos coordinados entre la transcripción y el procesamiento del mRNA (*splicing*) y cómo estos procesos se influyen entre sí (Merkhofer *et al.*, 2014). En *E. histolytica* se han encontrado varios factores de transcripción, entre ellos los factores Myb que juegan un papel importante en la regulación transcripcional de genes durante el ciclo de vida del parásito (Clark *et al.*, 2007; Meneses *et al.*, 2010; Ehrenkauffer *et al.*, 2009). En cuanto al *splicing*, en este parásito también se han identificado varios genes relacionados con este proceso, entre los cuales U2F84 (Valdés *et al.*, 2014; Gonzalez-Blanco *et al.*, 2022) es el mejor caracterizado en *E. histolytica*. Sin embargo, a la fecha no se han estudiado los eventos cotranscripcionales, es por ello que en este trabajo nos propusimos evaluar el papel de las proteínas EhMyb10 y EhCDC5 en la regulación cotranscripcional, las cuales podrían tener un papel relevante como factores de transcripción y de *splicing* respectivamente. EhMyb10 es un factor de transcripción que reconoce al ERM canónico y se ha visto que se sobreexpresa en condiciones de virulencia (Meneses *et al.*, 2010; Gilchrist *et al.*, 2006). Sin embargo, la proteína EhCDC5 aún no se ha caracterizado en *E. histolytica*. Al realizar un análisis *in silico* se encontró una identidad de 56.2 % con la proteína CDC5 de *H. sapiens* y al realizar un alineamiento observamos que presenta los dominios característicos de las proteínas de la familia CDC5 además en la proteína EhCDC5 se encontraron tres regiones RRM (del inglés RNA Recognition Motif) que podrían estar implicados en el reconocimiento del mRNA en el proceso de *splicing*. Adicionalmente, por medio de ensayos de CLIP se identificaron a las proteínas EhMyb10 y EhCDC5 en el transcrito de RabC1 sugiriendo la presencia de un mecanismo de regulación cotranscripcional en este protozooario (Comunicación personal con el Dr. Jesús Valdés Flores).

Debido a ello, por medio de herramientas bioinformáticas seleccionamos 147 genes que tenían el ERM río arriba del INR. Además, estos genes contienen intrones y tienen un nivel de expresión similar a EhMyb10 y EhCDC5, lo que los hace candidatos de regulación cotranscripcional. Estos genes codifican para proteínas de citoesqueleto, involucradas en el procesamiento de RNA, proteínas de tráfico de membrana, entre otras, de manera interesante, muchos de ellos aumentan su expresión en condiciones de virulencia. Estos 4 genes están

relacionados con los procesos de transcripción, factores generales de transcripción, factores de procesamiento de RNA y *splicing* debido a la correulación que pudiera darse por EhMyb10 y EhCDC5. Para corroborar que estos pudieran ser blancos putativos de ambas proteínas, se eligió una estrategia de sobreexpresión en los trofozoítos de *E. histolytica*. Para ello, se clonaron los genes en vectores de expresión amebianos que permitieran la sobreexpresión de las proteínas EhMyb10 y EhCDC5. Estrategias similares se han utilizado para caracterizar factores de transcripción en *E. histolytica* como EhMybdr y HMG los cuales se sobreexpresaron de manera estable con un FC de ~40 y 100 veces respectivamente, lo cual les permitió identificar a los genes blanco de estos factores de transcripción (Gilchrist *et al.*, 2012; Abhyankar *et al.*, 2008). En este trabajo no se pudo obtener trofozoítos que expresaran las proteínas de manera estable, debido al tiempo que toma el proceso de selección de los trofozoítos. Por ello, se usó un modelo de expresión transitoria, obteniendo valores de sobreexpresión de ~3 para ambos genes a las 72 horas de transfección. Aunque esta diferencia en expresión pudo haber limitado los resultados, la transfección transitoria nos permitió contar con un modelo viable para sobreexpresar a estas proteínas y analizar sus efectos. Cabe mencionar que los modelos de transfección transitoria se han utilizado por otros investigadores en otras líneas celulares, observando que el número de transcritos expresados diferencialmente aumenta con respecto al tiempo de transfección (Jacobsen *et al.*, 2009).

Adicionalmente, inmunodetectamos a las proteínas por WB para determinar si la sobreexpresión a nivel de mensajero genera en consecuencia más proteína. De manera interesante, EhMyb10 es detectada como una banda de ~40 kDa en todas las condiciones que se enriquece en los trofozoítos que sobreexpresan EhMyb10 y EhCDC5, debido a que el anticuerpo detecta tanto a la proteína heteróloga como a la endógena; esto contrasta con el peso teórico de EhMyb10 que es de 17.9 kDa. Esta variación podría deberse a modificaciones postraduccionales como SUMOilaciones, ubiquitinaciones y glicosilaciones. La SUMOilación afecta a una amplia gama de procesos nucleares, como la transcripción, la remodelación de la cromatina, el empalme del pre-RNAm, el ensamblaje de los ribosomas, la reparación del DNA, el mantenimiento de la integridad del genoma, el transporte nuclear y la transducción de señales. Se ha demostrado que la SUMOilación de c-Myb aumenta la estabilidad de la proteína al unirse a su dominio de regulación negativa (Hendriks *et al.*, 2014). La

O-glicosilación se ha reportado principalmente en proteínas con función nuclear como la RNAPII y sus factores de transcripción asociados, así como nucleoporinas, etc (Kamemura y Hart, 2003). Aunque esta modificación no se ha reportado en c-Myb, por medio de análisis *in silico* se ha detectado que EhMyb10 presenta residuos susceptibles a glicosilación. Estas modificaciones postraduccionales podrían explicar el peso molecular detectado. En el caso de EhCDC5, se detectaron 2 bandas de ~30 y 60 KDa sólo en aquellas transfectantes con EhCDC5 y la cotransfección (EhMyb10+EhCDC5). La banda de ~30 KDa coincide con el peso teórico de EhCDC5 y la banda de 60 KDa podría corresponder a una formación de dímeros como se ha observado en la proteína ortóloga de humanos (Gräub *et al.*, 2018).

Por ensayos de inmunofluorescencia, observamos su localización en el citoplasma y en el núcleo del parásito. La localización nuclear tanto de EhMyb10 como de EhCDC5 es congruente con su posible función en la regulación transcripcional y *splicing* respectivamente. Por otro lado, observamos que en la condición de cotransfección (trofozoítos transfectados con ambas construcciones) las proteínas EhMyb10 y EhCDC5 colocalizan en el núcleo, lo que sugiere un posible papel en un evento cotranscripcional.

En *E. histolytica* se ha observado con anterioridad el proceso de la RI y su posible papel en la regulación génica (González-Blanco *et al.*, 2022; Hon *et al.*, 2013). En este trabajo se analizó el efecto de EhMyb10 y EhCDC5 en la RI de 4 genes monitores en las diferentes condiciones de transfección mediante RT-PCR semicuantitativo. Nuestros resultados mostraron que la sobreexpresión de EhMyb10 y EhCDC5 modificaron los patrones de *splicing* en los genes analizados. Las diferencias que observamos, podrían sugerir un efecto gen específico. También observamos que en la condición de cotransfección un comportamiento distinto al efecto de la sobreexpresión de cada proteína por separado, lo que sugiere una correlación ya sea negativa o positiva sobre el *splicing* de los genes analizados.

De manera interesante, en los genes con intrones pequeños (< 75 pb), no se observan ss 5' y 3' alternativos en comparación con el gen *snrp70*, el cual tiene un intrón de 154 pb el cual mostró sitios alternativos de *splicing*, lo cual pudimos confirmar por medio de secuenciación de las 4 isoformas transcritas, lo cual podría dar lugar a proteínas con diferentes funciones. La proteína completa de

SNRP70 está constituida por 509 aminoácidos, contiene un elemento de reconocimiento a RNA y una región rica en argininas y asparaginas. las otras dos isoformas estarían constituidas por 91 y 114 aminoácidos y no presentan el elemento de reconocimiento a RNA, ni la región rica en argininas y asparaginas, por lo que estas proteínas podrían no ser funcionales.

Por otro lado, en el trabajo realizado por González-Blanco y col. (2022) se analizó la RI en 7 genes relacionados con virulencia observando un efecto inversamente proporcional al tamaño del intrón ya que al sobreexpresar EhU2AF84 aumentó la RI cuando el tamaño del intrón era pequeño. En contraste, en este trabajo se observó que el efecto de EhMyb10 y EhCDC5 sobre la RI es gen específico, sin embargo, Monteuis y col. (2019) determinan que la RI no puede ser exclusivamente atribuida al tamaño del intrón sino a otras características como % de GC en los intrones, sitios de *splicing* débiles, motivos de unión de los factores de *splicing*, entre otros. La RI puede afectar el proceso de *splicing*, mediante la acumulación de mRNA sin procesar o parcialmente procesado. Esto podría interrumpir la regulación de la expresión génica y afectar la composición del proteoma, además de la generación de proteínas no funcionales o truncas (Ner-Gaon *et al.*, 2004). En el caso de *E.histolytica*, se ha observado que al inhibir el *splicing* aumenta la RI y a su vez baja la virulencia en este parásito (González-Blanco *et al.*, 2022).

Se realizó un análisis del transcriptoma e indica que en las tres condiciones hay un mayor número de genes con expresión negativa. Esto podría deberse a que se ha observado que la sobreexpresión afecta la estequiometría del ambiente celular o interacciones promiscuas, causando un efecto de represión en la transcripción (Prelich, 2012; Moriya, 2015). Lo cual podría explicar el porqué vemos un mayor número de genes reprimidos.

En la condición de EhMyb10, 179 genes modificaron su expresión. Uno de los genes que se encontró con mayor expresión fue el gen que codifica para S-adenosilmetionina sintetasa (SAM) con un FC de 12.8. Esta enzima participa, en reacciones de transmetilación, es necesaria para procesos celulares del DNA, RNA y la metilación de histonas, que son importantes para los eventos de la transcripción. En el parásito de la malaria *Plasmodium falciparum*, esta enzima está codificada por un solo gen y se ha sugerido que es crucial para la patogénesis y transmisión de la malaria (Lopez-Rubio *et al.*, 2007; Perez-Toledo

*et al.*, 2009). Otro gen que modificó su expresión es el que codifica para la proteína Yos1-like, (FC: 6.6), la cual se encuentra implicada en el metabolismo del RNA (Bycroft *et al.*, 1997). Por otro lado, el gen que codifica para la proteína similar a una metaloproteasa aumentó su expresión (FC: 3.7). Las metaloproteasas juegan un papel importante en la patogénesis de infecciones parasitarias. Por ejemplo, Teixeira y colaboradores en 2012 encontraron mediante el silenciamiento de EhMSP-1, que esta metaloproteasa está involucrada en la regulación de adherencia con efectos en la motilidad, destrucción de la monocapa celular y la fagocitosis. Finalmente, se encontró sobreexpresado un gen que codifica para una proteína similar a tetraspanina (FC: 3.3), estas son proteínas transmembranales que juegan un papel importante en el sistema inmune, modulan la señalización celular, la adhesión, la morfología y la motilidad (Hemler, 2008; Tomii *et al.*, 2019). Estas dos últimas proteínas mencionadas podrían jugar un papel importante en la patogenicidad de *E. histolytica*. El gen que codifica para la proteína ribosomal L18 reprimió su expresión con un FC de -14.5. Esta proteína actúa en respuesta al estrés, atenuando la síntesis global de proteínas, pero permitiendo la traducción eficiente de proteínas de choque térmico (HSP) (Zhang *et al.*, 2015). El número de genes que modificó su expresión al sobreexpresar a EhMyb10 (33 genes sobreexpresados y 146 genes reprimidos) es similar al número de genes que se obtuvieron al sobreexpresar EhMybdr (117 genes sobreexpresados y 88 genes reprimidos) (Ehrenkauf *et al.*, 2009), sin embargo, en la condición de EhMyb10 hay un mayor número de genes reprimidos. Además, encontramos que existen dos genes que se regulan de manera común, el gen que codifica para la proteína Ras family GTPase y el gen que codifica una proteína similar al factor de *splicing* CWC25.

Cuando se sobreexpresó a EhCDC5, 243 genes modificaron su expresión de los cuales 50 se sobreexpresaron y 193 se reprimieron. Al igual que en EhMyb10, se identificó al gen que codifica para SAM (FC: 12.54), además de un gen que codifica para ATP sintasa vacuolar subunidad B (FC: 8.54). Esta enzima está compuesta por múltiples subunidades y es responsable de acidificar los organelos intracelulares bombeando protones a través de las membranas. En *E. histolytica* se localizó a esta proteína en vesículas, fagosomas y EhkOs. Es de destacar que la subunidad B se acumuló en la boca fagocítica de trofozoítos que fagocitaron glóbulos rojos (Meléndez-Hernández *et al.*, 2018). Otro gen de interés que se sobreexpresó es el que codifica para una proteína de alta movilidad de unión a

nucleosoma (HMGN1, del inglés high-mobility group protein 1)(FC: 2.6). Estas proteínas se unen a los nucleosomas y regulan su arquitectura favoreciendo la transcripción (Bianchi y Agresti, 2005). Por ejemplo, HsHMGN5 se une preferentemente a la eucromatina modulando la transcripción al evitar el efecto de la cromatina compactada (González-Romero., *et al* 2015). Otro gen que modificó su expresión fue el que codifica para la proteína ciclina 2 (FC: 2.53) la cual forma parte del citoesqueleto, desempeñando un papel en la motilidad de los espermatozoides (Schneider *et al.*, 2023). Otro gen es el que codifica para el factor de *splicing* CWC25 (FC: 1.6), proteína requerida para la primera reacción catalítica durante el *splicing* después del paso de Prp2 dependiente de ATP (Chiu *et al.*, 2009). De manera general, en esta condición se encontraron varios genes sobreexpresados relacionados con la maquinaria de *splicing* (*U1 zinc finger protein, prp4, cwc25*), genes relacionados a la transcripción (*SAM, TBP, Myb, LSM, Swi5, hmgn1*), a reorganización del citoesqueleto (*Cilicina 2, protein Sojo citoesqueleto, filamina/ABP281*) y a virulencia (*Tetraspanina, antígeno ariel 1*) y resistencia a fármacos (*P-glicoproteína*).

Cuando se sobreexpresaron a EhMyb10 y EhCDC5, se identificaron 115 genes que modificaron su expresión, de los cuales 21 genes se sobreexpresaron y 94 se reprimieron. Dentro de los genes que se sobreexpresaron fueron el gen que codifica para una proteína similar a tetraspanina (FC: 2.7) la cual se sobreexpresa en las tres condiciones. Otro gen es el que codifica para la proteína similar a la mucina 1 (FC: 2.5) la cual está relacionada a virulencia ya que intervienen en la interacción célula-célula, ayudando así al parásito a establecer la infección (Maneesh *et al.*, 2001). Se identificó un gen que codifica para la proteína HMG1, al igual que en la condición de EhCDC5ox. De manera general, esta fue la condición con menor número de genes que modificaron su expresión. Los genes sobreexpresados están relacionados con virulencia y patogenicidad (*antígeno ariel1, tetraspanina, mucina, proteínas de unión a GTP, proteína de unión a inmunoglobulina*) y algunos genes de respuesta a estrés (*hsp70 y hsp101*). Se encontraron varios genes reprimidos relacionados con traducción y genes relacionados a ribosomas entre otros.

Finalmente, se buscaron motivos en los promotores de los genes que modificaron su expresión en cada condición con la idea de buscar ERM conservados en estos genes. De manera interesante, se encontró una secuencia en común en todos los

promotores de los genes que modificaron su expresión en todas las condiciones. Esta secuencia es CAACAACAAC, además en esta secuencia se conserva el core AAC el cual es relevante para la unión de los factores de transcripción Myb ya que la tercera alfa hélice del R3 interacciona con los nucleótidos AAC (Tanikawa *et al.*, 1993).

Este es el primer reporte en el que se analiza el papel funcional de dos proteínas Myb en la regulación cotranscripcional, encontrando un efecto cotranscripcional sobre la RI de tres genes analizados, además de los 61 genes regulados en común en las condiciones analizadas por medio de RNAseq. En la condición de cotransfección se encontraron sobreexpresados principalmente genes relacionados con la patogenicidad y virulencia.

## 11. Conclusiones

1. Las proteínas EhMyb10 y EhCDC5 se localizan en el núcleo de los trofozoítos transfectados transitoriamente y se colocan en la condición de cotransfección, sugiriendo su papel en los mecanismos cotranscripcionales.
2. Al sobreexpresar EhMyb10 y EhCDC5 se modificó el transcriptoma y los patrones de *splicing* en los genes seleccionados, encontrando efectos positivos o negativos sobre la retención intrónica dependiendo del gen.
3. La sobreexpresión de EhMyb10 y EhCDC5 tuvo un efecto a nivel transcripcional, ya que se encontraron 61 genes compartidos en todas las condiciones que modificaron su expresión, los cuales podrían ser sus genes blanco.
4. Los genes que modificaron su expresión al sobreexpresar a las proteínas EhMyb10 y EhCDC5 están relacionados con procesos de transcripción y *splicing* y además asociados a la virulencia.
5. El gen TSPAN-like (EHI\_103400) podría regularse cotranscripcionalmente, ya que tiene en la región promotora un elemento ERM, presenta un intrón y se sobreexpresó en las tres condiciones.
6. Mediante el análisis *in silico*, se predijeron 147 genes que pueden ser candidatos a ser regulados cotranscripcionalmente por EhMyb10 y EhCDC5 en *E. histolytica*; entre los cuales hay genes involucrados en la regulación transcripcional y el *splicing*. Se seleccionaron 4 de estos genes, los cuales tienen el ERM una expresión moderada y al menos un intrón.
7. En este trabajo se generaron transfectantes de trofozoítos que sobreexpresan a los FT y de *splicing* EhMyb10 y EhCDC5 lo cual se corroboró por medio de inmunodetección e inmunofluorescencia. Este último ensayo nos permitió observar la colonización nuclear de EhMyb10 y EhCDC5 en los trofozoítos de *E. histolytica* transfectados y su colocación en los cotransfectados, lo que nos podría indicar su participación en la regulación cotranscripcional.
8. Mediante el análisis *in vivo* al sobreexpresar a las proteínas EhMyb10 y EhCDC5, se modifica el patrón de *splicing* de los genes analizados. Para algunos genes se potencia la retención intrónica y para otros se inhibe. El efecto observado de la sobreexpresión de ambas proteínas sobre la

retención intrónica es distinto a su efecto individual, sugiriendo un efecto sinérgico sobre el *splicing*.

## 12. Perspectivas

1. Corroborar mediante RT-qPCR los cambios de expresión de genes seleccionados de las tres condiciones (EhMyb10ox, EhCDC5ox y EhMyb10-EhCDC5ox).
2. Realizar un ensayo de Inmunoprecipitación de la cromatina (CHIP) para las tres condiciones de los genes seleccionados para corroborar si son blancos EhMyb10 y EhCDC5.
3. Realizar un ensayo de CLIP para EhCDC5 con el transcrito de SPAN-like y extractos nucleares de *E. histolytica*, para analizar su papel en el *splicing*.
4. Analizar retención intrónica de los genes mayormente expresados en la condición de cotransfección de los datos obtenidos de RNAseq mediante alguna herramienta bioinformática (Ballgown, iRED, IRFinder-S, MATS).
5. Realizar ensayos funcionales (eritrofagocitosis, citotoxicidad, etc.) para las tres condiciones analizadas para caracterizar su fenotipo y correlacionar con los genes que modificaron su expresión.

### 13. Bibliografía

1. Abhyankar, M. M., Hochreiter, A. E., Hershey, J., Evans, C., Zhang, Y., Crasta, O., ... & Gilchrist, C. A. (2008). Characterization of an *Entamoeba histolytica* high-mobility-group box protein induced during intestinal infection. *Eukaryotic Cell*, 7(9), 1565-1572.
2. Adams, D. J., Van Der Weyden, L., Mayeda, A., Stamm, S., Morris, B. J., & Rasko, J. E. (2001). ZNF265—a novel spliceosomal protein able to induce alternative splicing. *The Journal of Cell Biology*, 154(1), 25-32.
3. Agresti, A., & Bianchi, M. E. (2003). HMGB proteins and gene expression. *Current opinion in genetics & development*, 13(2), 170-178.
4. Aguilar-Rojas, A., Castellanos-Castro, S., Matondo, M., Gianetto, Q. G., Varet, H., Sismeiro, O., ... & Guillen, N. (2020). Insights into amebiasis using a human 3D-intestinal model. *Cellular Microbiology*, 22(8), e13203.
5. Alvarado-Esquivel, C., Hernandez-Tinoco, J., & Sanchez-Anguiano, L. F. (2015). Seroepidemiology of *Entamoeba histolytica* infection in general population in rural Durango, Mexico. *Journal of clinical medicine research*, 7(6), 435.
6. Arts, R. J., Huang, P. K., Yang, D., Joosten, L. A., van der Meer, J. W., Oppenheim, J. J., ... & Cheng, S. C. (2018). High-mobility group nucleosome-binding protein 1 as endogenous ligand induces innate immune tolerance in a TLR4-Sirtuin-1 dependent manner in human blood peripheral mononuclear cells. *Frontiers in Immunology*, 9, 526.
7. Bakatselou, C., Beste, D., Kadri, A. O., Somanath, S., & Clark, C. G. (2003). Analysis of genes of mitochondrial origin in the genus *Entamoeba*. *Journal of Eukaryotic Microbiology*, 50(3), 210-214.
8. Begum, S., Gorman, H., Chadha, A., & Chadee, K. (2021). *Entamoeba histolytica*. *Trends in Parasitology*, 37(7), 676-677.
9. Bello, F., Orozco, E., Benítez-Cardoza, C. G., Zamorano-Carrillo, A., Reyes-López, C. A., Pérez-Ishiwara, D. G., & Gómez-García, C. (2022). The novel EhHSTF7 transcription factor displays an oligomer state and recognizes a heat shock element in the *Entamoeba histolytica* parasite. *Microbial Pathogenesis*, 162, 105349.
10. Bhattacharya, A., Satish, S., Bagchi, A., & Bhattacharya, S. (2000). The genome of *Entamoeba histolytica*. *International journal for parasitology*, 30(4), 401-410.
11. Bianchi, M. E., & Agresti, A. (2005). HMG proteins: dynamic players in gene regulation and differentiation. *Current opinion in genetics & development*, 15(5), 496-506.
12. Burns, C. G., Ohi, R., Krainer, A. R., & Gould, K. L. (1999). Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proceedings of the National Academy of Sciences*, 96(24), 13789-13794.
13. Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell*, 88(2), 235-242.
14. Castañón-Sánchez, C. A., Luna-Arias, J. P., de Dios-Bravo, M. G., Herrera-Aguirre, M. E., Olivares-Trejo, J. J., Orozco, E., & Hernandez, J. M. (2010). *Entamoeba histolytica*: a unicellular organism containing two active genes encoding for members of the TBP family. *Protein expression and purification*, 70(1), 48-59.
15. Chan, S. P., Kao, D. I., Tsai, W. Y., & Cheng, S. C. (2003). The Prp19p-associated complex in spliceosome activation. *Science*, 302(5643), 279-282.

16. Chanarat, S., Seizl, M., & Sträßer, K. (2011). The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes & development*, 25(11), 1147-1158.
17. Chanarat, S., & Sträßer, K. (2013). Splicing and beyond: the many faces of the Prp19 complex. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1833(10), 2126-2134.
18. Chiu, Y. F., Liu, Y. C., Chiang, T. W., Yeh, T. C., Tseng, C. K., Wu, N. Y., & Cheng, S. C. (2009). Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Molecular and cellular biology*, 29(21), 5671-5678.
19. Clark, C. G., Alsmark, U. C. M., Tazreiter, M., Saito-Nakano, Y., Ali, V., Marion, S., ... & Hall, N. (2007). Structure and content of the *Entamoeba histolytica* genome. *Advances in parasitology*, 65, 51-190.
20. Collier, S. E., Voehler, M., Peng, D., Ohi, R., Gould, K. L., Reiter, N. J., & Ohi, M. D. (2014). Structural and functional insights into the N-terminus of *Schizosaccharomyces pombe* Cdc5. *Biochemistry*, 53(41), 6439-6451.
21. Cramer, P., Pesce, C. G., Baralle, F. E., & Kornblihtt, A. R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proceedings of the National Academy of Sciences*, 94(21), 11456-11460.
22. Cruz, O. H. D. L., Marchat, L. A., Guillén, N., Weber, C., Rosas, I. L., Díaz-Chávez, J., ... & López-Camarillo, C. (2016). Multinucleation and polykaryon formation is promoted by the EhPC4 transcription factor in *Entamoeba histolytica*. *Scientific reports*, 6(1), 19611.
23. Davis, P. H., Zhang, X., Guo, J., Townsend, R. R., & Stanley Jr, S. L. (2006). Comparative proteomic analysis of two *Entamoeba histolytica* strains with different virulence phenotypes identifies peroxiredoxin as an important component of amoebic virulence. *Molecular microbiology*, 61(6), 1523-1532.
24. Davis, C. A., Brown, M. P., & Singh, U. (2007). Functional characterization of spliceosomal introns and identification of U2, U4, and U5 snRNAs in the deep-branching eukaryote *Entamoeba histolytica*. *Eukaryotic Cell*, 6(6), 940-948.
25. Davis, P. H., Schulze, J., & Stanley Jr, S. L. (2007). Transcriptomic comparison of two *Entamoeba histolytica* strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. *Molecular and biochemical parasitology*, 151(1), 118-128.
26. de Dios-Bravo, G., Luna-Arias, J. P., Riveron, A. M., Olivares-Trejo, J. J., Lopez-Camarillo, C., & Orozco, E. (2005, July). *Entamoeba histolytica* TATA-box binding protein has the ability to bind to different TATA variants in vitro. In *FEBS JOURNAL* (Vol. 272, pp. 469-469). 9600 GARSINGTON RD, OXFORD OX4 2DQ, OXON, ENGLAND: BLACKWELL PUBLISHING.
27. De la Mata, M., Lafaille, C., & Kornblihtt, A. R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *Rna*, 16(5), 904-912.
28. Deutschman, C. S. (2005). Transcription. *Critical care medicine*, 33(12), S400-S403.
29. Fornander, L. H., Renodon-Corniere, A., Kuwabara, N., Ito, K., Tsutsui, Y., Shimizu, T., ... & Takahashi, M. (2014). Swi5-Sfr1 protein stimulates Rad51-mediated DNA strand exchange reaction through organization of DNA bases in the presynaptic filament. *Nucleic acids research*, 42(4), 2358-2365.

30. Fotedar, R., Stark, D., Beebe, N., Marriott, D., Ellis, J., & Harkness, J. (2007). Laboratory diagnostic techniques for *Entamoeba* species. *Clinical microbiology reviews*, 20(3), 511-532.
31. Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews Genetics*, 7(9), 703-713.
32. Ge, Y. y Porse, BT (2014). Las consecuencias funcionales de la retención de intrones: splicing alternativo acoplado a NMD como regulador de la expresión génica. *Bioensayos* 36, 236–243.
33. Gilchrist, C. A., Holm, C. F., Hughes, M. A., Schaenman, J. M., Mann, B. J., & Petri, W. A. (2001). Identification and Characterization of an *Entamoeba histolytica* Upstream Regulatory Element 3 Sequence-specific DNA-binding Protein Containing EF-hand Motifs. *Journal of Biological Chemistry*, 276(15), 11838-11843.
34. Gilchrist, C. A., Houpt, E., Trapaidze, N., Fei, Z., Crasta, O., Asgharpour, A., ... & Petri Jr, W. A. (2006). Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Molecular and Biochemical Parasitology*, 147(2), 163-176.
35. Gilchrist, C. A., Baba, D. J., Zhang, Y., Crasta, O., Evans, C., Caler, E., ... & Petri Jr, W. A. (2008). Targets of the *Entamoeba histolytica* transcription factor URE3-BP. *PLoS Neglected Tropical Diseases*, 2(8), e282.
36. González-Arranz, S., Acosta, I., Carballo, J. A., Santos, B., & San-Segundo, P. A. (2021). The N-Terminal Region of the Polo Kinase Cdc5 Is Required for Downregulation of the Meiotic Recombination Checkpoint. *Cells*, 10(10), 2561.
37. González-Blanco, G., García-Rivera, G., Talmás-Rohana, P., Orozco, E., Galindo-Rosales, J. M., Vélez, C., ... & Valdés, J. (2022). An Unusual U2AF2 Inhibits Splicing and Attenuates the Virulence of the Human Protozoan Parasite *Entamoeba histolytica*. *Frontiers in Cellular and Infection Microbiology*, 12, 888428.
38. González-Rivas, E., Nieves-Ramírez, M., Magaña, U., Morán, P., Rojas-Velázquez, L., Hernández, E., ... & Ximénez, C. (2020). Differential pathogenic gene expression of *E. histolytica* in patients with different clinical forms of amoebiasis. *Microorganisms*, 8(10), 1556.
39. González-Romero, R., Eirín-López, J. M., & Ausió, J. (2015). Evolution of high mobility group nucleosome-binding proteins and its implications for vertebrate chromatin specialization. *Molecular Biology and Evolution*, 32(1), 121-131.
40. Grabski, D. F., Broseus, L., Kumari, B., Rekosh, D., Hammarskjold, M. L., & Ritchie, W. (2020). Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. *Wiley Interdisciplinary Reviews: RNA*, 12(1), e1631.
41. Gräub, R., Lancero, H., Pedersen, A., Chu, M., Padmanabhan, K., Xu, X. Q., ... & Bernstein, H. S. (2008). Cell cycle-dependent phosphorylation of human CDC5 regulates RNA processing. *Cell Cycle*, 7(12), 1795-1803.
42. Grote, M., Wolf, E., Will, C. L., Lemm, I., Agafonov, D. E., Schomburg, A., ... & Lührmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. *Molecular and cellular biology*, 30(9), 2105-2119.
43. Hemler, M. E. (2008). Targeting of tetraspanin proteins—potential benefits and strategies. *Nature reviews Drug discovery*, 7(9), 747-758.
44. Hendriks, I. A., D'souza, R. C., Yang, B., Verlaan-de Vries, M., Mann, M., & Vertegaal, A. C. (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. *Nature structural & molecular biology*, 21(10), 927-936.
45. Hirayama, T., & Shinozaki, K. (1996). A cdc5+ homolog of a higher plant, *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 93(23), 13371-13376.

46. Hon, C. C., Weber, C., Sismeiro, O., Proux, C., Koutero, M., Deloger, M., ... & Guillen, N. (2013). Quantification of stochastic noise of splicing and polyadenylation in *Entamoeba histolytica*. *Nucleic acids research*, 41(3), 1936-1952.
47. Hu, J. U. N., Lutz, C. S., Wilusz, J., & Tian, B. I. N. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *Rna*, 11(10), 1485-1493.
48. Huang, Y., Li, W., Yao, X., Lin, Q. J., Yin, J. W., Liang, Y., ... & Wang, G. (2012). Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Molecular cell*, 45(4), 459-469.
49. Huerta M, Reyes L, García-Rivera G, Bañuelos C, Betanzos A, Díaz-Hernández M, Galindo A, Bolaños J, Cárdenas H, Azuara-Liceaga E, Chávez-Munguía B, Orozco E. A noncanonical GATA transcription factor of *Entamoeba histolytica* modulates genes involved in phagocytosis. *Mol Microbiol*. 2020 Dec;114(6):1019-1037. doi: 10.1111/mmi.14592. Epub 2020 Sep 10. PMID: 32808689.
50. Idrissou, M., & Maréchal, A. (2022). The PRP19 ubiquitin ligase, standing at the cross-roads of mRNA processing and genome stability. *Cancers*, 14(4), 878.
51. Illumina (2014). TruSeq® RNA Sample Preparation v2 Guide (Rev F).
52. Iyer, L. M., Anantharaman, V., Wolf, M. Y., & Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *International journal for parasitology*, 38(1), 1-31.
53. Jacob, A. G., & Smith, C. W. (2017). Intron retention as a component of regulated gene expression programs. *Human genetics*, 136(9), 1043-1057.
54. Jacobsen, L. B., Calvin, S. A., & Lobenhofer, E. K. (2009). Transcriptional effects of transfection: the potential for misinterpretation of gene expression data generated from transiently transfected cells. *Biotechniques*, 47(1), 617-624.
55. Kamemura, K., & Hart, G. W. (2003). Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription. *Progress in nucleic acid research and molecular biology*, 73, 107-136.
56. Kattah, N. H., Kattah, M. G., & Utz, P. J. (2010). The U1-snRNP complex: structural properties relating to autoimmune pathogenesis in rheumatic diseases. *Immunological reviews*, 233(1), 126-145.
57. Koncz, C., Dejong, F., Villacorta, N., Szakonyi, D., & Koncz, Z. (2012). The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. *Frontiers in plant science*, 3, 9.
58. Krivega, I., & Dean, A. (2012). Enhancer and promoter interactions—long distance calls. *Current opinion in genetics & development*, 22(2), 79-85.
59. Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., ... & Weirauch, M. T. (2018). The human transcription factors. *Cell*, 172(4), 650-665.
60. Lidell, M. E., Moncada, D. M., Chadee, K., & Hansson, G. C. (2006). *Entamoeba histolytica* cysteine proteases cleave the MUC2 mucin in its C-terminal domain and dissolve the protective colonic mucus gel. *Proceedings of the National Academy of Sciences*, 103(24), 9298-9303.
61. Liu, X., Bushnell, D. A., & Kornberg, R. D. (2013). RNA polymerase II transcription: structure and mechanism. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1829(1), 2-8.
62. Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup> ΔΔCT method. *methods*, 25(4), 402-408.

63. López-Camarillo, C., Orozco, E., & Marchat, L. A. (2005). Entamoeba histolytica: Comparative genomics of the pre-mRNA 3' end processing machinery. *Experimental parasitology*, 110(3), 184-190.
64. Lopez-Rubio, J. J., Riviere, L., & Scherf, A. (2007). Shared epigenetic mechanisms control virulence factors in protozoan parasites. *Current opinion in microbiology*, 10(6), 560-568.
65. Luna-Arias, J. P., Hernandez-Rivas, R., de Dios-Bravo, G., Garcia, J., Mendoza, L., & Orozco, E. (1999). The TATA-box binding protein of Entamoeba histolytica: cloning of the gene and location of the protein by immunofluorescence and confocal microscopy. *Microbiology*, 145(1), 33-40.
66. MacFarlane, R. C., & Singh, U. (2007). Identification of an Entamoeba histolytica serine-, threonine-, and isoleucine-rich protein with roles in adhesion and cytotoxicity. *Eukaryotic Cell*, 6(11), 2139-2146.
67. Mancini, E. (2016). *Identificación y caracterización de eventos de splicing alternativo en Arabidopsis thaliana utilizando herramientas de transcriptómica de alto rendimiento* (Doctoral dissertation, Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales).
68. Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., & Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome research*, 25(7), 995-1007.
69. Meléndez-Hernández, M. G., Barrios, M. L., Orozco, E., & Luna-Arias, J. P. (2008). The Vacuolar ATPase from Entamoeba histolytica: Molecular cloning of the gene encoding for the B subunit and subcellular localization of the protein. *BMC microbiology*, 8, 1-14.
70. Mendoza, L., Orozco, E., Rodriguez, M. A., Garcia-Rivera, G., Sanchez, T., Garcia, E., & Gariglio, P. (2003). Ehp53, an Entamoeba histolytica protein, ancestor of the mammalian tumour suppressor p53. *Microbiology*, 149(4), 885-893.
71. Monteuuis, G., Wong, J. J., Bailey, C. G., Schmitz, U., & Rasko, J. E. (2019). The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic acids research*, 47(22), 11497-11513.
72. Moriya, H. (2015). Quantitative nature of overexpression experiments. *Molecular biology of the cell*, 26(22), 3932-3939.
73. Nandi, N., Sen, A., Banerjee, R., Kumar, S., Kumar, V., Ghosh, A. N., & Das, P. (2010). Hydrogen peroxide induces apoptosis-like death in Entamoeba histolytica trophozoites. *Microbiology*, 156(7), 1926-1941.
74. Narayanasamy, R. K., Castañón-Sánchez, C. A., Luna-Arias, J. P., García-Rivera, G., Avendaño-Borromeo, B., Labra-Barrios, M. L., ... & Orozco, E. (2018). The Entamoeba histolytica TBP and TRF1 transcription factors are GAAC-box binding proteins, which display differential gene expression under different stress stimuli and during the interaction with mammalian cells. *Parasites & vectors*, 11, 1-16.
75. Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., & Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *The Plant Journal*, 39(6), 877-885.
76. Noe Gonzalez, M., Blears, D., & Svejstrup, J. Q. (2021). Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nature reviews Molecular cell biology*, 22(1), 3-21.
77. Oesterreich, F. C., Preibisch, S., & Neugebauer, K. M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Molecular cell*, 40(4), 571-581.
78. Ohi, R., Feoktistova, A., McCann, S., Valentine, V., Look, A. T., Lipsick, J. S., & Gould, K. L. (1998). Myb-related Schizosaccharomyces pombe cdc5p is

- structurally and functionally conserved in eukaryotes. *Molecular and Cellular Biology*, 18(7), 4097-4108.
79. Palma, K., Zhao, Q., Cheng, Y. T., Bi, D., Monaghan, J., Cheng, W., ... & Li, X. (2007). Regulation of plant innate immunity by three proteins in a complex conserved across the plant and animal kingdoms. *Genes & development*, 21(12), 1484.
  80. Patel, A. A., & Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. *Nature reviews Molecular cell biology*, 4(12), 960-970.
  81. Perez-Toledo, K., Rojas-Meza, A. P., Mancio-Silva, L., Hernandez-Cuevas, N. A., Delgadillo, D. M., Vargas, M., ... & Hernandez-Rivas, R. (2009). Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic acids research*, 37(8), 2596-2606.
  82. Phatnani, H. P., & Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development*, 20(21), 2922-2936.
  83. Prelich, G. (2012). Gene overexpression: uses, mechanisms, and interpretation. *Genetics*, 190(3), 841-854.
  84. Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 352(6291), aad9926.
  85. Ramos, F., Morán, P., Gonzalez, E., García, G., Ramiro, M., Gómez, A., ... & Ximenez, C. (2005). High prevalence rate of *Entamoeba histolytica* asymptomatic infection in a rural Mexican community. *The American journal of tropical medicine and hygiene*, 73(1), 87-91.
  86. Rojas, L., Morán, P., Valadez, A., Gómez, A., González, E., Hernández, E., ... & Ximénez, C. (2016). *Entamoeba histolytica* and *Entamoeba dispar* infection in Mexican school children: genotyping and phylogenetic relationship. *BMC infectious diseases*, 16(1), 1-12.
  87. Saito-Nakano, Y., Yasuda, T., Nakada-Tsukui, K., Leippe, M., & Nozaki, T. (2004). Rab5-associated vacuoles play a unique role in phagocytosis of the enteric protozoan parasite *Entamoeba histolytica*. *Journal of Biological Chemistry*, 279(47), 49497-49507.
  88. Schaenman, J. M., Gilchrist, C. A., Mann, B. J., & Petri, W. A. (2001). Identification of Two *Entamoeba histolytica* Sequence-specific URE4 Enhancer-binding Proteins with Homology to the RNA-binding Motif RRM. *Journal of Biological Chemistry*, 276(2), 1602-1609.
  89. Schier, A. C., & Taatjes, D. J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes & development*, 34(7-8), 465-488.
  90. Schmitz, U., Pinello, N., Jia, F., Alasmari, S., Ritchie, W., Keightley, M. C., ... & Rasko, J. E. (2017). Intron retention enhances gene regulatory complexity in vertebrates. *Genome biology*, 18(1), 1-15.
  91. Schneider, S., Kovacevic, A., Mayer, M., Dicke, A. K., Arévalo, L., Koser, S. A., ... & Schorle, H. (2023). Cylicins are a structural component of the sperm calyx being indispensable for male fertility in mice and human. *Elife*, 12, RP86100.
  92. Shenasa, H., & Bentley, D. L. (2023). Pre-mRNA splicing and its cotranscriptional connections. *Trends in Genetics*.
  93. Shepard, P. J., & Hertel, K. J. (2009). The SR protein family. *Genome biology*, 10(10), 1-9.
  94. Sims, D., Sudbery, I., Iltott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121-132.

95. Štros, M. (2010). HMGB proteins: interactions with DNA and chromatin. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1799(1-2), 101-113.
96. Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., & Sarai, A. (1993). Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. *Proceedings of the National Academy of Sciences*, 90(20), 9320-9324.
97. Teixeira, J. E., Sateriale, A., Bessoff, K. E., & Huston, C. D. (2012). Control of *Entamoeba histolytica* adherence involves metallosurface protease 1, an M8 family surface metalloprotease with homology to leishmanolysin. *Infection and immunity*, 80(6), 2165-2176.
98. Torres-Cifuentes, D. M., Galindo-Rosales, J. M., Saucedo-Cárdenas, O., & Valdés, J. (2018). The *Entamoeba histolytica* Syf1 homolog is involved in the splicing of AG-dependent and AG-independent transcripts. *Frontiers in Cellular and Infection Microbiology*, 8, 229.
99. Valdés-Flores, J., López-Rosas, I., López-Camarillo, C., Ramírez-Moreno, E., Ospina-Villa, J. D., & Marchat, L. A. (2018). Life and Death of mRNA Molecules in *Entamoeba histolytica*. *Frontiers in Cellular and Infection Microbiology*, 8, 199.
100. Wilson, I. W., Weedall, G. D., & Hall, N. (2012). Host–Parasite interactions in *Entamoeba histolytica* and *Entamoeba dispar*: what have we learned from their genomes?. *Parasite Immunology*, 34(2-3), 90-99.
101. Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, 3(7), a003707.
102. Yanagawa, Y., Izumiyama, S., Saito-Nakano, Y., Nakada-Tsukui, K., Kobayashi, S., Yoshida, N. and Watanabe, K. (2022). Gene expression of axenically-isolated clinical *Entamoeba histolytica* strains and its impact on disease severity of amebiasis. *PLoS Pathogens*, 18(9), e1010880.
103. Zhang, S., Xie, M., Ren, G., & Yu, B. (2013). CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proceedings of the National Academy of Sciences*, 110(43), 17588-17593.
104. Zhang, X., Gao, X., Coots, R. A., Conn, C. S., Liu, B., & Qian, S. B. (2015). Translational control of the cytosolic stress response by mitochondrial ribosomal protein L18. *Nature structural & molecular biology*, 22(5), 404-410.
105. Zhao, P., Javed, S., Shi, X., Wu, B., Zhang, D., Xu, S., & Wang, X. (2020). Varying architecture of heat shock elements contributes to distinct magnitudes of target gene expression and diverged biological pathways in heat stress response of bread wheat. *Frontiers in Genetics*, 11, 30.
106. Zhu, D., Mao, F., Tian, Y., Lin, X., Gu, L., Gu, H., ... & Wu, Z. (2020). The features and regulation of co-transcriptional splicing in *Arabidopsis*. *Molecular plant*, 13(2), 278-294.

## 14. Anexos

Tabla . Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhMyb10 y EhCDC5-like.

Gene ID	FC	p Value	Product Description	# Exons in Gene
<b>Genes sobreexpresados</b>				
EHI_147990	3.9	3.00E-03	hypothetical protein/TLDC domain-containing protein	1
EHI_103400*	2.8	9.01E-03	hypothetical protein/Tetraspanin family protein	2
EHI_023400	2.2	1.41E-04	Ribosomal protein S30, putative	1
<b>EHI_103900</b>	<b>2.1</b>	<b>1.05E-03</b>	<b>nucleosome-binding protein 1, putative</b>	1
<b>EHI_073540</b>	<b>2.0</b>	<b>3.79E-03</b>	<b>heat shock protein 101, putative</b>	1
EHI_107170	2.0	3.73E-03	hypothetical protein/surface antigen ariel1, putative	1
EHI_166800	1.9	1.19E-02	ubiquitin, putative	1
<b>EHI_112590</b>	<b>1.9</b>	<b>1.50E-02</b>	<b>heat shock protein 70 family</b>	1
EHI_079930	1.6	1.85E-02	hypothetical protein	1
EHI_166370	1.6	2.93E-03	hypothetical protein/surface antigen ariel1, putative	1
EHI_194730	1.6	7.65E-03	mucin-like protein 1 precursor, putative	1
<b>EHI_070450</b>	<b>1.5</b>	<b>1.36E-02</b>	<b>heat shock protein70, mitochondrial, putative</b>	2
<b>Genes reprimidos</b>				
EHI_174150	-6.3	6.54E-04	hypothetical protein	2
EHI_068380	-4.6	1.16E-02	WD domain containing protein	1
EHI_032670	-3.7	1.57E-03	hypothetical protein/NADPH-dependent FMN reductase domain containing protein	2
EHI_074380	-3.3	1.68E-03	hypothetical protein/transporter, major facilitator family	1
EHI_141050*	-3.3	1.04E-02	hypothetical protein	2
EHI_051690	-2.7	3.20E-03	hypothetical protein/eukaryotic translation initiation factor 3 subunit	1
EHI_017660	-2.4	9.78E-04	hypothetical protein/ataxia telangiectasia mutated, putative	1
EHI_194430	-2.4	8.44E-04	iron-sulfur cluster-binding protein, putative	1
EHI_083100	-2.4	1.09E-03	hypothetical protein	1
EHI_010020	-2.4	1.16E-04	calmodulin, putative	1
<b>EHI_118270</b>	<b>-2.3</b>	<b>1.24E-03</b>	<b>amoebapore C</b>	<b>1</b>
EHI_162230	-2.3	9.46E-03	alpha-1,3-mannosyltransferase ALG2, putative	2
EHI_118670	-2.3	2.05E-03	Skp1 family protein	1
EHI_142900	-2.3	1.55E-02	hypothetical protein	1
EHI_048140	-2.3	3.26E-04	hypothetical protein	1
EHI_032780	-2.3	8.02E-03	hypothetical protein	2
EHI_075530	-2.2	1.15E-03	coatomer subunit beta, putative	1
EHI_100350	-2.2	5.13E-03	hypothetical protein,	1

			conserved/cupin domain containing protein	
EHI_019860	-2.1	1.03E-02	hypothetical protein	1
EHI_048300	-2.1	9.15E-03	WD repeat protein 68, putative	1
EHI_050000	-2.1	3.53E-03	hypothetical protein/COMM domain containing protein	1
EHI_187780	-2.1	8.09E-04	hypothetical protein	1
EHI_000130	-2.1	1.20E-03	calmodulin, putative	2
EHI_062500	-2	1.68E-04	hypothetical protein/SNAP receptor complex member 1	3
EHI_156310	-2	6.48E-03	ribonuclease, putative	1
EHI_135060	-2	1.63E-03	ribosomal protein S17, putative	1
EHI_126130	-2	1.24E-02	hypothetical protein	2
<b>EHI_048860</b>	<b>-2</b>	<b>1.83E-02</b>	<b>U3 small nucleolar ribonucleo protein MPP10, putative</b>	<b>1</b>
EHI_134840	-2	6.07E-03	ADP-ribosylation factor, putative	1
EHI_058320	-1.9	4.69E-03	transmembrane protein tmp21, putative	1
EHI_010030	-1.9	2.46E-03	hypothetical protein/Ubiquitin-related modifier 1 family protein, putative	2
EHI_009940	-1.9	7.86E-03	clathrin adaptor complex small chain, putative	3
EHI_011880	-1.8	1.94E-02	hypothetical protein/COMM domain containing protein	1
EHI_035670	-1.8	2.99E-04	hypothetical protein	1
EHI_180000	-1.8	8.59E-03	eukaryotic translation initiation factor 4E, putative	1
EHI_028990	-1.8	1.92E-02	hypothetical protein/proteasome maturation factor ump1	2
EHI_157360	-1.8	5.07E-03	AIG1 family protein, putative	2
EHI_095830	-1.7	1.50E-03	hypothetical protein/GCN5 family protein 1 (GCN5L1) protein	1
EHI_150530	-1.7	1.77E-02	ATP-binding cassette sub-family E member 1, putative	1
EHI_148190	-1.7	1.04E-03	Ran family GTPase	1
EHI_006850	-1.7	1.43E-02	hypothetical protein/zinc finger protein, putative	1
EHI_017750	-1.7	3.04E-04	hypothetical protein/axoneme-associated protein mst101, putative	1
EHI_119540	-1.7	7.11E-04	eukaryotic translation elongation factor 1 gamma, putative	1
EHI_182550	-1.7	9.09E-03	Vacuolar ATP synthase subunit E, putative	1
EHI_046600	-1.7	9.70E-03	acyl-CoA hydrolase, putative	1
EHI_086080	-1.7	1.98E-02	proteasome subunit alpha type 1, putative	1
EHI_170060	-1.6	1.21E-02	ubiquitin like proteinsilico	1
EHI_151350	-1.6	1.29E-02	hypothetical protein	1
EHI_092930	-1.6	5.27E-03	hypothetical protein/methionyl-tRNA synthetase, putative	1
EHI_148470	-1.6	3.56E-03	asparagine synthetase A, putative	1
EHI_103510	-1.6	1.59E-02	ubiquitin domain protein, putative	1
<b>EHI_188960</b>	<b>-1.6</b>	<b>2.51E-03</b>	<b>histone H2A, putative</b>	<b>1</b>

EHI\_167310      -1.5      1.52E-02      grainin 2      1

\*Genes que coinciden con la predicción *in silico* realizada en el objetivo 1.

En negritas se muestran los genes que codifican para proteínas relevantes e involucradas en procesos relacionados con la virulencia de este parásito.

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a la proteína EhMyb10 (sobreexpresados).

Gene ID	FC	p Value	Anotación	# Exones
EHI_195110	12.82	1.32E-03	S-adenosylmethionine synthetase, putative	1
EHI_006770	7.38	2.21E-05	hypothetical protein, conserved / phosphopantetheine adenylyltransferase	2
EHI_123020	7.29	2.01E-05	hypothetical protein/ protein phosphatase domain-containing protein	1
EHI_018630	7.10	6.58E-05	hypothetical protein/ ER membrane protein, BIG1 family	1
EHI_178640	6.69	1.80E-04	Yos1-like domain-containing protein	2
EHI_083380	5.97	2.26E-03	hypothetical protein, conserved/ WD repeat protein*	1
EHI_109930	3.96	2.46E-03	hypothetical protein/ Proteasome assembly chaperone 4	1
EHI_159450	3.73	6.82E-03	hypothetical protein/ METALLOPROTEASE	1
EHI_103400	3.34	9.77E-05	hypothetical protein/ Tetraspanin	2
EHI_117270	3.13	3.69E-04	fumarate hydratase class I, anaerobic, putative	1
EHI_120650	2.92	1.07E-02	hypothetical protein/ proteasome regulatory subunit	1
EHI_050640	2.76	3.76E-03	P-glycoprotein 6, putative	1
EHI_073540	2.71	1.07E-03	heat shock protein 101, putative	1
EHI_166800	2.49	2.30E-03	ubiquitin, putative	1
EHI_179380	2.44	2.76E-02	hypothetical protein/ Signal recognition particle 9 kDa protein*	2
EHI_167620	2.32	6.09E-03	protein phosphatase domain-containing protein	2
EHI_146540	2.27	3.39E-03	hypothetical protein/ Synaptonemal complex protein 1	2
EHI_108640	2.14	1.93E-02	LSM domain containing protein	1
EHI_044220	2.01	4.20E-02	hypothetical protein	1
EHI_181610	2.00	4.41E-05	hypothetical protein/Swi5 protein	2
EHI_165810	1.91	3.83E-05	hypothetical protein, pseudogene	1
EHI_051580	1.87	1.96E-02	hypothetical protein, Eukaryotic translation initiation factor 3 subunit, putative	1
EHI_023400	1.87	3.25E-03	Ribosomal protein S30, putative	1
EHI_100810	1.84	9.85E-03	heat shock protein 70, mitochondrial, putative	1
EHI_079930	1.77	8.08E-03	hypothetical protein	1
EHI_040370	1.77	1.39E-02	hypothetical protein	1
EHI_107170	1.71	3.29E-02	hypothetical protein/ surface antigen ariel1, putative*	1
EHI_157010	1.69	8.87E-04	alcohol dehydrogenase, putative	1
EHI_057550	1.66	2.21E-04	methionine gamma-lyase, putative	1
EHI_094470	1.56	3.28E-02	heat shock protein 101, putative	1
EHI_048170	1.56	1.85E-02	Protein transport protein SEC61 gamma subunit, putative	2

EHI_068610	1.56	1.49E-02	hypothetical protein/ pre-mRNA-splicing factor CWC25, putative*	1
EHI_193300	1.50	2.28E-02	phosphoribulokinase/uridine kinase family protein	1

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a la proteína EhMyb10 (reprimidos).

Gene ID	FC	p Value	Anotación	# Exones
EHI_098470	-14.58	3.36E-05	ribosomal protein L18, putative	1
EHI_035060	-9.57	4.94E-10	fumarate hydratase class I, anaerobic, putative	1
EHI_155900	-9.26	9.89E-09	hypothetical protein, conserved/ Adapter-related protein complex 3 (AP-3)	1
EHI_049990	-8.43	1.00E-07	chaperone proteinDNAJ, putative	1
EHI_004430	-8.37	1.26E-07	hypothetical protein	1
EHI_089770	-7.64	2.73E-06	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit	1
EHI_183200	-7.29	1.90E-05	hypothetical protein, C-Myc-binding protein	2
EHI_179850	-6.32	7.23E-04	hypothetical protein, Pre-mRNA-splicing factor CWC25	2
EHI_083140	-6.31	6.39E-06	hypothetical protein, Calponin-homology (CH) domain-containing protein	2
EHI_033240	-6.25	8.19E-04	Riboflavin kinase , FAD synthetase, putative	1
EHI_085680	-5.68	4.09E-03	WD repeat protein	1
EHI_000690	-4.74	4.90E-02	seryl-tRNA synthetase, putative	1
EHI_169350	-3.80	2.32E-07	nonpathogenic pore-forming peptide precursor, putative	1
EHI_159910	-3.67	1.46E-02	ariadne-1, putative	1
EHI_005030	-3.61	4.98E-03	hypothetical protein, psq-type domain-containing protein	1
EHI_134600	-3.61	4.98E-03	hypothetical protein, psq-type domain-containing protein	1
EHI_172280	-3.57	1.56E-07	hypothetical protein, RanBD1 domain-containing protein	1
EHI_156380	-3.48	1.71E-03	U2 snRNP auxiliary factor small subunit, putative	1
EHI_180390	-3.41	1.40E-03	AIG1 family protein, putative	1
EHI_178340	-3.33	4.40E-07	ubiquitin, putative	1
EHI_032670	-3.29	1.86E-03	hypothetical protein, NADPH-dependent FMN reductase domain containing protein	2
EHI_170160	-3.26	3.35E-02	hypothetical protein, Protein phosphatase 2, regulatory subunit B	2
EHI_126980	-3.19	8.15E-05	hypothetical protein	1
EHI_044610	-3.00	4.90E-03	ribosome biogenesis protein NEP1, putative	2
EHI_194430	-3.00	8.10E-06	iron-sulfur cluster-binding protein, putative	1
EHI_116040	-2.92	3.80E-03	hypothetical protein, dab2-interacting protein, putative*	1
EHI_155680	-2.92	4.38E-09	hypothetical protein, Syntaxin F	3
EHI_099860	-2.89	5.34E-06	iron-sulfur cluster-binding protein, putative	1
EHI_120950	-2.86	6.04E-04	hypothetical protein	1
EHI_047920	-2.82	2.32E-02	hypothetical protein, Ras guanine nucleotide exchange factor, putative	2
EHI_152800	-2.82	7.48E-06	hypothetical protein	2

EHI_048140	-2.81	1.30E-06	hypothetical protein	1
EHI_179390	-2.74	4.25E-07	Translation initiation factor, putative	2
EHI_187780	-2.72	1.77E-05	hypothetical protein, EamA domain-containing protein	1
EHI_050000	-2.71	9.24E-04	hypothetical protein, COMM domain containing protein	1
EHI_118270	-2.71	1.10E-04	amoebapore C	1
EHI_072640	-2.59	3.64E-02	hypothetical protein, conserved, Calpastatin	1
EHI_196370	-2.56	1.25E-03	hypothetical protein	2
EHI_006960	-2.56	1.71E-02	hypothetical protein	1
EHI_140770	-2.55	8.53E-03	Rab family GTPase	4
EHI_001060	-2.44	2.12E-02	hypothetical protein, C2 domain-containing protein	1
EHI_068380	-2.44	4.45E-02	WD domain containing protein	1
EHI_035670	-2.42	1.63E-06	hypothetical protein	1
EHI_100350	-2.41	6.63E-03	hypothetical protein, Cupin 2 conserved barrel domain-containing protein	1
EHI_028990	-2.39	1.98E-04	hypothetical protein, Proteasome maturation factor UMP1	2
EHI_164840	-2.36	3.53E-02	hypothetical protein, Tetraspanin family protein	1
EHI_058320	-2.35	4.22E-04	transmembrane protein tmp21, putative	1
EHI_092720	-2.30	2.60E-02	hypothetical protein	2
EHI_095830	-2.29	2.70E-04	hypothetical protein	1
EHI_094810	-2.28	1.04E-02	Ras family protein	1
EHI_010030	-2.28	3.18E-05	hypothetical protein	2
EHI_092170	-2.28	1.70E-05	hypothetical protein	1
EHI_038300	-2.26	5.18E-04	hypothetical protein	1
EHI_160910	-2.25	1.23E-02	hypothetical protein	1
EHI_143440	-2.23	2.61E-05	calmodulin, putative	1
EHI_092930	-2.23	2.88E-07	hypothetical protein, methionyl-tRNA synthetase, putative	1
EHI_010020	-2.21	5.57E-04	calmodulin, putative	1
EHI_180000	-2.18	2.41E-03	eukaryotic translation initiation factor 4E, putative	1
EHI_077770	-2.13	5.18E-03	UDP-glucose:glyco protein glucosyltransferase 1 precursor, putative	1
EHI_069950	-2.12	1.31E-03	C2 domain containing protein	1
EHI_000620	-2.12	3.24E-02	PP-loop family protein	1
EHI_103510	-2.11	1.42E-03	ubiquitin domain protein, putative	1
EHI_138470	-2.11	3.67E-04	hypothetical protein	1
EHI_118230	-2.10	4.49E-04	hypothetical protein	1
EHI_032780	-2.10	3.02E-03	hypothetical protein	2
EHI_188960	-2.10	9.31E-07	histone H2A, putative	1
EHI_062500	-2.08	7.12E-05	hypothetical protein, golgi snap receptor complex member 1 isoform x1	3
EHI_117850	-2.08	5.84E-03	hypothetical protein	1
EHI_083270	-2.04	2.50E-03	ubiquitin, putative	1
EHI_083410	-2.04	2.50E-03	ubiquitin, putative	1
EHI_103370	-2.03	8.07E-04	hypothetical protein	3

EHI_118670	-2.02	5.13E-04	Skp1 family protein	1
EHI_103310	-2.02	5.22E-03	60S ribosomal protein L7a, putative	1
EHI_156310	-2.01	1.05E-02	ribonuclease, putative	1
EHI_045200	-2.00	4.03E-02	hypothetical protein	2
EHI_090020	-2.00	1.02E-02	ribosomal protein L18a, putative	1
EHI_012420	-2.00	6.94E-04	myb-like DNA-binding domain containing protein	1
EHI_144610	-1.98	4.69E-05	methionine gamma-lyase	1
EHI_022010	-1.97	5.50E-03	hypothetical protein	1
EHI_000130	-1.96	7.09E-04	calmodulin, putative	2
EHI_202460	-1.95	7.18E-04	calponin-homology domain containing protein	1
EHI_182950	-1.95	8.69E-03	CXXC-rich protein	1
EHI_017750	-1.93	1.03E-05	hypothetical protein	1
EHI_157360	-1.92	1.16E-03	AIG1 family protein, putative	2
EHI_056690	-1.90	5.09E-04	DNA-directed RNA polymerase II subunit, putative	1
EHI_046600	-1.90	2.94E-03	acyl-CoA hydrolase, putative	1
EHI_135060	-1.89	2.87E-03	ribosomal protein S17, putative	1
EHI_182550	-1.88	1.66E-03	Vacuolar ATP synthase subunit E, putative	1
EHI_009490	-1.87	3.69E-03	hypothetical protein	1
EHI_136740	-1.86	1.38E-02	hypothetical protein, conserved	1
EHI_151350	-1.85	5.72E-03	hypothetical protein	1
EHI_126210	-1.84	7.51E-04	histone H2A, putative	1
EHI_125600	-1.84	2.81E-02	hypothetical protein	1
EHI_093910	-1.84	2.78E-02	hypothetical protein, conserved	1
EHI_134810	-1.84	5.30E-05	rubrerythrin, putative	1
EHI_177520	-1.82	1.86E-03	Rab family GTPase	3
EHI_194190	-1.81	8.35E-03	hypothetical protein	1
EHI_134840	-1.81	5.30E-03	ADP-ribosylation factor, putative	1
EHI_148190	-1.81	6.03E-04	Ran family GTPase	1
EHI_098060	-1.80	1.31E-02	proteasome subunit alpha type 3, putative	1
EHI_112800	-1.79	1.78E-04	Mov34/MPN/PAD-1 family protein	2
EHI_008040	-1.78	1.28E-02	Thymidylate kinase, putative	2
EHI_105280	-1.76	1.01E-02	hypothetical protein, transcription initiation factor TFIID subunit 10B, putative	3
EHI_069280	-1.75	1.09E-03	hypothetical protein	1
EHI_091470	-1.75	4.91E-02	hypothetical protein	2
EHI_153940	-1.75	4.09E-02	hypothetical protein, pseudogene	1
EHI_093840	-1.74	2.87E-02	zinc finger protein, putative	2
EHI_119540	-1.74	1.81E-03	eukaryotic translation elongation factor 1 gamma, putative	1
EHI_086080	-1.73	1.40E-02	proteasome subunit alpha type 1, putative	1
EHI_045280	-1.73	4.44E-03	RNA recognition motif domain containing protein	2
EHI_170300	-1.73	4.74E-02	hypothetical protein	1
EHI_151890	-1.72	7.82E-03	EF-hand calcium-binding domain containing protein	1
EHI_141800	-1.71	6.38E-04	hypothetical protein	1
EHI_045930	-1.70	2.00E-02	hypothetical protein	1

EHI_177460	-1.68	6.45E-03	translation initiation factor eIF-5A, putative	1
EHI_188170	-1.66	9.56E-04	hypothetical protein	1
EHI_044820	-1.66	1.63E-03	calmodulin, putative	1
EHI_105140	-1.66	4.90E-03	40S ribosomal protein S15a, putative	1
EHI_167270	-1.66	4.90E-03	40S ribosomal protein S15a, putative	1
EHI_036560	-1.63	1.80E-03	hypothetical protein	2
EHI_079890	-1.63	4.70E-03	Rab family GTPase	3
EHI_075530	-1.63	3.63E-02	coatamer subunit beta', putative	1
EHI_182610	-1.61	2.26E-02	myb-like DNA-binding domain containing protein	1
EHI_126860	-1.61	5.46E-03	hypothetical protein	2
EHI_020080	-1.60	8.47E-05	hypothetical protein	1
EHI_040310	-1.60	4.46E-05	Rab family GTPase	2
EHI_053420	-1.59	2.92E-02	Rab family GTPase	2
EHI_017890	-1.58	4.30E-02	60S ribosome subunit biogenesis protein NIP7, putative	1
EHI_191530	-1.58	4.30E-02	60S ribosome subunit biogenesis protein NIP7, putative	1
EHI_058090	-1.58	3.92E-04	Ras family GTPase	1
EHI_001940	-1.58	2.04E-03	hypothetical protein	2
EHI_170060	-1.57	1.22E-02	ubiquitin like protein	1
EHI_015310	-1.57	1.74E-02	60S ribosomal protein L35a, putative	1
EHI_073180	-1.57	4.76E-02	zinc finger domain containing protein	1
EHI_013220	-1.56	7.09E-03	eukaryotic translation initiation factor 4E, putative	2
EHI_187800	-1.55	2.26E-02	hypothetical protein, conserved	1
EHI_110570	-1.54	6.01E-03	protein tyrosine phosphatase, putative	1
EHI_076020	-1.54	4.93E-03	hypothetical protein	1
EHI_156290	-1.53	2.15E-02	BRAC1 domain containing protein	2
EHI_086540	-1.52	1.82E-02	replication factor C subunit 4, putative	1
EHI_127400	-1.52	1.48E-02	hypothetical protein	1
EHI_035650	-1.52	5.27E-03	ribose-phosphate pyrophosphokinase, putative	1
EHI_005070	-1.52	2.66E-03	acetyltransferase, GNAT family	3
EHI_037810	-1.51	3.24E-02	hypothetical protein	2
EHI_052110	-1.51	3.24E-02	hypothetical protein	2
EHI_176820	-1.51	4.48E-03	N-acetylmuraminidase, putative	1

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a la proteína EhCDC5 (sobreexpresados).

Gene ID	FC	p Value	Anotación	# Exones
EHI_195110	12.55	9.81E-17	S-adenosylmethionine synthetase, putative	1
EHI_118220	8.55	3.01E-07	vacuolar ATP synthase subunit B, putative	2
EHI_025840	6.89	1.67E-04	LSM domain containing protein	2
EHI_109930	4.42	3.76E-04	hypothetical protein, Proteasome assembly chaperone	3
EHI_117270	3.09	3.70E-02	fumarate hydratase class I, anaerobic, putative	1
EHI_087650	2.83	1.37E-02	hypothetical protein	1

EHI_050640	2.80	1.41E-03	P-glycoprotein 6, putative	1
EHI_192420	2.72	2.31E-02	hypothetical protein, conserved	1
EHI_101730	2.68	1.72E-02	surface antigen ariel1, putative	1
EHI_197140	2.67	3.52E-02	hypothetical protein/ Nuclear receptor coactivator 7B	1
EHI_103900	2.56	2.06E-11	nucleosome-binding protein 1, putative	1
EHI_080100	2.54	7.37E-03	cylicin-2, putative	1
EHI_073540	2.52	5.21E-04	heat shock protein 101, putative	1
EHI_011400	2.51	9.95E-03	hypothetical protein/ TATA-binding protein-associated phosphoprotein, putative *	1
EHI_166800	2.37	2.11E-04	ubiquitin, putative	1
EHI_166860	2.31	2.24E-02	hypothetical protein	1
EHI_148260	2.24	4.45E-03	hypothetical protein,U1 zinc finger protein*	1
EHI_023400	2.24	9.40E-08	Ribosomal protein S30, putative	1
EHI_140480	2.23	3.36E-02	hypothetical protein, extracellular protein with coiled coils(Interpro)**	1
EHI_103400	2.15	3.49E-02	hypothetical protein/ Tetraspanin**	2
EHI_044220	2.15	3.80E-02	hypothetical protein	1
EHI_086030	2.11	3.76E-02	hypothetical protein, cytoskeletal protein Sojo, putative*	1
EHI_030150	2.09	3.30E-02	hypothetical protein, conserved, synaptonemal complex protein, putative*	1
EHI_167620	2.09	1.16E-03	protein phosphatase domain-containing protein	2
EHI_057550	2.08	8.91E-09	methionine gamma-lyase, putative	1
EHI_069510	2.06	4.39E-02	hypothetical protein	1
EHI_170470	1.92	4.26E-02	ER lumen protein retaining receptor	2
EHI_107170	1.88	1.68E-03	hypothetical protein, surface antigen ariel1, putative*	1
EHI_181610	1.85	1.03E-06	hypothetical protein, Swi5 protein, putative*	2
EHI_130480	1.85	3.97E-02	hypothetical protein	1
EHI_037820	1.80	3.47E-02	hypothetical protein, palmitoyltransferase PFA3, putative*	1
EHI_041580	1.75	1.53E-03	hypothetical protein, conserved	1
EHI_008130	1.74	1.62E-02	Myb family DNA-binding protein, SHAQKYF family	2
EHI_158130	1.73	9.05E-03	hypothetical protein	1
EHI_003240	1.72	3.38E-02	N-acetylmuraminidase, pseudogene	1
EHI_105350	1.72	4.73E-03	hypothetical protein, TATA-binding protein-associated phosphoprotein, putative*	1
EHI_058110	1.71	1.54E-02	hypothetical protein	1
EHI_100810	1.69	2.90E-03	heat shock protein 70, mitochondrial, putative	1
EHI_079930	1.67	1.54E-02	hypothetical protein	1
EHI_068140	1.66	2.30E-02	hypothetical protein, polyamine-modulated factor 1-binding protein, putative*	1
EHI_087610	1.64	1.50E-02	hypothetical protein, prp4*	3
EHI_159330	1.64	1.50E-02	hypothetical protein, prp4, putative*	3
EHI_161900	1.61	4.81E-02	hypothetical protein/,splicing factor, arginine/serine-rich, putative*	1
EHI_159490	1.60	1.03E-02	hypothetical protein, TATA-binding protein-associated phosphoprotein, putative*	1

EHI_126270	1.59	9.66E-03	Ras GTPase-activating protein, putative	2
EHI_014250	1.58	5.14E-04	hypothetical protein	1
EHI_068610	1.55	1.05E-04	hypothetical protein, pre-mRNA-splicing factor CWC25, putative*	1
EHI_193300	1.54	2.47E-02	phosphoribulokinase/uridine kinase family protein	1
EHI_176850	1.52	3.86E-02	hypothetical protein, conserved/ GRIP domain containing protein RUD3*	1
EHI_122900	1.50	1.23E-03	hypothetical protein, conserved/ DNA double-strand break repair Rad50 ATPase, putative*	1

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhCDC5 (reprimidos)

Gene ID	FC	p Value	Anotación	# Exones
EHI_098470	-14.28	4.76E-05	ribosomal protein L18, putative	1
EHI_094020	-10.42	6.42E-12	hypothetical protein, filamin,ABP280 repeat containing protein	1
EHI_000690	-9.98	1.11E-10	seryl-tRNA synthetase, putative	1
EHI_035060	-9.28	1.23E-09	fumarate hydratase class I, anaerobic, putative	1
EHI_155900	-8.96	1.73E-08	hypothetical protein, conserved, Adapter-related protein complex 3 (AP-3)	1
EHI_083140	-8.46	2.21E-03	hypothetical protein, cortexillin II, putative	2
EHI_140640	-7.43	1.10E-05	F-actin capping protein subunit alpha, putative	1
EHI_089770	-7.35	9.64E-06	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit, putative	1
EHI_048300	-6.98	5.88E-06	WD repeat protein 68, putative	1
EHI_020990	-6.66	2.44E-04	hypothetical protein, Nucleoporin NSP1	1
EHI_041880	-6.27	7.09E-04	hypothetical protein	2
EHI_193940	-6.10	4.37E-09	hypothetical protein, calponin domain containing protein*	2
EHI_116040	-5.20	2.36E-03	hypothetical protein, dab2-interacting protein*	1
EHI_156380	-5.20	2.59E-03	U2 snRNP auxiliary factor small subunit, putative	1
EHI_039020	-4.82	4.39E-03	actobindin, putative	1
EHI_083100	-4.57	2.06E-07	hypothetical protein	1
EHI_188880	-4.43	1.65E-02	Chaperonin 10 Kd subunit, putative	1
EHI_000620	-4.29	2.06E-03	PP-loop family protein	1
EHI_099860	-4.15	2.97E-09	iron-sulfur cluster-binding protein, putative	1
EHI_169350	-4.13	1.37E-04	nonpathogenic pore-forming peptide precursor, putative	1
EHI_114920	-4.08	5.70E-23	60S ribosomal protein L35a, putative	1
EHI_172280	-4.05	9.61E-10	hypothetical protein, leukocyte-endothelial cell adhesion molecule 3, putative*	1
EHI_152800	-3.90	2.05E-12	hypothetical protein	2
EHI_048860	-3.82	2.24E-03	U3 small nucleolar ribonucleo protein MPP10, putative	1
EHI_158570	-3.58	7.47E-11	actobindin, putative	1
EHI_178340	-3.57	7.09E-11	ubiquitin, putative	1
EHI_126980	-3.52	7.37E-06	hypothetical protein	1
EHI_148400	-3.48	1.78E-03	hypothetical protein	1

EHI_035220	-3.40	7.01E-08	hypothetical protein	1
EHI_047920	-3.35	1.65E-02	hypothetical protein, Ras guanine nucleotide exchange factor, putative*	2
EHI_160910	-3.25	2.88E-04	hypothetical protein, Helix-turn-helix domain of resolvase domain containing protein*	1
EHI_050000	-3.24	8.79E-05	hypothetical protein, COMM domain containing protein*	1
EHI_001060	-3.23	2.36E-02	hypothetical protein	1
EHI_151350	-3.09	1.80E-05	hypothetical protein	1
EHI_141050	-3.09	1.58E-02	hypothetical protein	2
EHI_118270	-3.06	1.26E-09	amoebapore C	1
EHI_118670	-2.99	5.38E-07	Skp1 family protein	1
EHI_092320	-2.99	6.76E-03	hypothetical protein, Cofilin,tropomyosin-type actin-binding protein	1
EHI_155680	-2.98	2.41E-12	hypothetical protein, SNARE domain containing protein*	3
EHI_017660	-2.88	1.97E-08	hypothetical protein	1
EHI_032780	-2.85	9.85E-05	hypothetical protein	2
EHI_075530	-2.82	1.20E-03	coatomer subunit beta', putative	1
EHI_028990	-2.80	2.58E-09	hypothetical protein	2
EHI_083270	-2.78	1.54E-19	ubiquitin, putative	1
EHI_083410	-2.78	1.54E-19	ubiquitin, putative	1
EHI_010020	-2.74	9.84E-08	calmodulin, putative	1
EHI_194430	-2.74	2.65E-05	iron-sulfur cluster-binding protein, putative	1
EHI_109880	-2.71	2.53E-03	hypothetical protein	1
EHI_090020	-2.68	5.93E-12	ribosomal protein L18a, putative	1
EHI_038300	-2.66	8.62E-06	hypothetical protein	1
EHI_096200	-2.64	1.79E-02	thioredoxin, putative	1
EHI_103370	-2.60	6.29E-04	hypothetical protein	3
EHI_035670	-2.59	9.00E-08	hypothetical protein	1
EHI_048140	-2.58	2.27E-08	hypothetical protein	1
EHI_135060	-2.58	1.03E-09	ribosomal protein S17, putative	1
EHI_052720	-2.57	2.13E-03	hypothetical protein	1
EHI_187780	-2.54	3.61E-05	hypothetical protein	1
EHI_117850	-2.51	1.72E-11	hypothetical protein	1
EHI_179390	-2.49	1.89E-07	Translation initiation factor, putative	2
EHI_103510	-2.49	6.32E-09	ubiquitin domain protein, putative	1
EHI_049000	-2.48	1.11E-14	60S ribosomal protein L35a, putative	1
EHI_103310	-2.47	7.43E-09	60S ribosomal protein L7a, putative	1
EHI_058320	-2.47	2.15E-04	transmembrane protein tmp21, putative	1
EHI_007260	-2.46	3.21E-02	hypothetical protein	1
EHI_140770	-2.45	1.80E-02	Rab family GTPase	4
EHI_188960	-2.45	9.71E-14	histone H2A, putative	1
EHI_095830	-2.44	9.63E-09	hypothetical protein	1
EHI_044610	-2.44	2.72E-02	ribosome biogenesis protein NEP1, putative	2
EHI_179990	-2.42	5.37E-03	hypothetical protein	1

EHI_182550	-2.40	1.07E-06	Vacuolar ATP synthase subunit E, putative	1
EHI_000130	-2.40	1.02E-05	calmodulin, putative	2
EHI_162230	-2.38	1.96E-03	alpha-1,3-mannosyltransferase ALG2, putative	2
EHI_160900	-2.37	4.18E-02	Sec1 family protein	1
EHI_032670	-2.35	1.21E-02	hypothetical protein	2
EHI_023000	-2.34	6.47E-06	hypothetical protein	2
EHI_151890	-2.32	1.70E-03	EF-hand calcium-binding domain containing protein	1
EHI_197480	-2.31	6.84E-11	actophorin, putative	1
EHI_010030	-2.31	6.27E-06	hypothetical protein	2
EHI_143440	-2.29	6.39E-07	calmodulin, putative	1
EHI_125600	-2.29	1.09E-02	hypothetical protein	1
EHI_006960	-2.28	3.25E-02	hypothetical protein	1
EHI_105280	-2.24	3.26E-05	hypothetical protein	3
EHI_017030	-2.24	2.87E-05	hypothetical protein, conserved	1
EHI_126210	-2.22	5.28E-06	histone H2A, putative	1
EHI_097930	-2.22	1.71E-04	muT/nudix family protein	1
EHI_196370	-2.19	5.12E-03	hypothetical protein	2
EHI_092930	-2.18	4.40E-08	hypothetical protein	1
EHI_177520	-2.18	7.79E-09	Rab family GTPase	3
EHI_134810	-2.16	4.83E-11	rubrerythrin, putative	1
EHI_118750	-2.15	2.59E-08	hypothetical protein	1
EHI_182950	-2.15	6.24E-03	CXXC-rich protein	1
EHI_077770	-2.14	8.81E-05	UDP-glucose:glyco protein glucosyltransferase 1 precursor, putative	1
EHI_126870	-2.14	7.77E-11	40S ribosomal protein S21, putative	2
EHI_170300	-2.10	2.82E-03	hypothetical protein	1
EHI_044820	-2.10	2.51E-05	calmodulin, putative	1
EHI_138470	-2.09	1.46E-05	hypothetical protein, myosin-light-chain kinase, putative	1
EHI_148190	-2.09	7.71E-07	Ran family GTPase	1
EHI_170060	-2.09	1.41E-09	ubiquitin like protein	1
EHI_081130	-2.08	8.67E-04	P-glycoprotein 6, putative	2
EHI_092170	-2.07	7.39E-06	hypothetical protein	1
EHI_160210	-2.07	8.65E-04	hypothetical protein, conserved	2
EHI_148470	-2.07	1.27E-05	asparagine synthetase A, putative	1
EHI_119540	-2.07	2.88E-08	eukaryotic translation elongation factor 1 gamma, putative	1
EHI_186720	-2.06	1.79E-10	40S ribosomal protein S12, putative	1
EHI_065340	-2.05	5.74E-07	hypothetical protein	1
EHI_180000	-2.04	5.18E-03	eukaryotic translation initiation factor 4E, putative	1
EHI_017750	-2.04	1.60E-07	hypothetical protein	1
EHI_073180	-2.04	2.24E-08	zinc finger domain containing protein	1
EHI_027610	-2.03	2.38E-10	hypothetical protein	2
EHI_115360	-2.03	3.73E-04	hypothetical protein	2

EHI_037810	-2.00	8.35E-04	hypothetical protein	2
EHI_052110	-2.00	8.35E-04	hypothetical protein	2
EHI_086080	-1.98	3.36E-05	proteasome subunit alpha type 1, putative	1
EHI_069950	-1.98	4.08E-03	C2 domain containing protein	1
EHI_134840	-1.98	2.60E-03	ADP-ribosylation factor, putative	1
EHI_105140	-1.97	7.61E-09	40S ribosomal protein S15a, putative	1
EHI_167270	-1.97	7.61E-09	40S ribosomal protein S15a, putative	1
EHI_030390	-1.96	5.75E-04	hypothetical protein	1
EHI_026360	-1.95	5.66E-03	phosphoserine aminotransferase, putative	1
EHI_202460	-1.93	1.01E-05	calponin-homology domain containing protein	1
EHI_100350	-1.93	1.33E-02	hypothetical protein, conserved	1
EHI_150530	-1.91	1.75E-02	ATP-binding cassette sub-family E member 1, putative	1
EHI_139050	-1.89	1.57E-02	hypothetical protein	1
EHI_022010	-1.88	3.24E-11	hypothetical protein	1
EHI_045280	-1.88	3.34E-03	RNA recognition motif domain containing protein	2
EHI_046600	-1.88	4.63E-04	acyl-CoA hydrolase, putative	1
EHI_024240	-1.87	1.64E-04	aldehyde-alcohol dehydrogenase 2, putative	3
EHI_201940	-1.87	5.13E-03	clathrin heavy chain, putative	1
EHI_187800	-1.86	1.33E-02	hypothetical protein, conserved	1
EHI_087570	-1.85	1.04E-06	endoribonuclease L-PSP, putative	1
EHI_066770	-1.83	2.72E-08	Skp1 family protein	1
EHI_174180	-1.83	2.72E-08	Skp1 family protein	1
EHI_141890	-1.82	2.50E-04	zinc finger protein containing CCHC type domain	1
EHI_009940	-1.82	5.30E-03	clathrin adaptor complex small chain, putative	3
EHI_146780	-1.82	3.37E-03	hypothetical protein	1
EHI_194050	-1.81	3.99E-03	hypothetical protein	1
EHI_050290	-1.79	2.23E-02	SGS domain protein	1
EHI_153260	-1.78	3.40E-03	hypothetical protein, conserved	2
EHI_118290	-1.77	4.03E-02	hypothetical protein	1
EHI_086200	-1.76	1.27E-03	hypothetical protein	2
EHI_087410	-1.75	1.41E-04	high mobility group (HMG) box domain containing protein	2
EHI_001940	-1.75	3.67E-06	hypothetical protein	2
EHI_036560	-1.75	1.30E-04	hypothetical protein	2
EHI_136740	-1.72	4.15E-03	hypothetical protein, conserved	1
EHI_159820	-1.72	6.91E-04	hypothetical protein	2
EHI_144610	-1.72	6.03E-06	methionine gamma-lyase	1
EHI_167310	-1.72	2.83E-05	grainin 2	1
EHI_193350	-1.70	5.78E-04	3'(2'),5'-bisphosphate nucleotidase, putative	1
EHI_079890	-1.69	9.43E-03	Rab family GTPase	3
EHI_110570	-1.69	1.06E-04	protein tyrosine phosphatase, putative	1
EHI_031240	-1.69	4.71E-04	serine/threonine-protein phosphatase, putative	1
EHI_112800	-1.67	1.57E-04	Mov34/MPN/PAD-1 family protein	2
EHI_099720	-1.66	3.33E-03	hypothetical protein	2

EHI_169300	-1.66	3.74E-03	ribonuclease, putative	1
EHI_029210	-1.66	6.17E-03	proteasome subunit alpha type 3, putative	1
EHI_035650	-1.65	3.23E-04	ribose-phosphate pyrophosphokinase, putative	1
EHI_006170	-1.65	4.94E-04	eukaryotic translation initiation factor 6, putative	1
EHI_141010	-1.64	3.05E-02	hypothetical protein	1
EHI_015310	-1.64	4.23E-07	60S ribosomal protein L35a, putative	1
EHI_196580	-1.63	2.18E-03	phosphoribulokinase /uridine kinase family protein	1
EHI_050520	-1.63	8.58E-06	hypothetical protein	1
EHI_140550	-1.62	1.60E-06	60S ribosomal protein L6, putative	1
EHI_147570	-1.61	2.46E-06	Rho GDP exchange inhibitor, putative	1
EHI_069280	-1.60	1.13E-03	hypothetical protein	1
EHI_126130	-1.60	2.43E-02	hypothetical protein	2
EHI_157360	-1.60	6.28E-03	AIG1 family protein, putative	2
EHI_030070	-1.59	8.67E-07	hypothetical protein	1
EHI_015440	-1.58	8.39E-05	Rho family GTPase	2
EHI_176820	-1.57	3.80E-04	N-acetylmuraminidase, putative	1
EHI_105120	-1.57	2.04E-02	hypothetical protein, conserved	1
EHI_013340	-1.57	1.81E-03	hypothetical protein	1
EHI_174910	-1.57	1.22E-04	ARP2/3 complex 21 kDa subunit, putative	2
EHI_062500	-1.57	1.83E-03	hypothetical protein	3
EHI_197410	-1.56	4.28E-02	cation transporter, putative	1
EHI_093910	-1.56	4.88E-02	hypothetical protein, conserved	1
EHI_186950	-1.56	4.15E-05	protein phosphatase domain-containing protein	2
EHI_005070	-1.56	9.13E-04	acetyltransferase, GNAT family	3
EHI_056690	-1.55	1.19E-04	DNA-directed RNA polymerase II subunit, putative	1
EHI_104360	-1.55	1.08E-04	nucleoside diphosphate kinase, putative	1
EHI_163670	-1.55	5.84E-05	ribosomal protein L18a, putative	1
EHI_013220	-1.54	1.55E-05	eukaryotic translation initiation factor 4E, putative	2
EHI_134490	-1.54	1.19E-05	F-actin capping protein subunit beta, putative	1
EHI_056460	-1.54	1.02E-04	hypothetical protein, conserved	1
EHI_198550	-1.54	1.68E-02	glucosamine 6-phosphate N-acetyltransferase. putative	2
EHI_063340	-1.53	2.44E-02	prefoldin, alpha subunit	2
EHI_087380	-1.53	2.44E-02	prefoldin, alpha subunit	2
EHI_161870	-1.52	6.83E-03	hypothetical protein	1
EHI_035510	-1.52	1.26E-03	hypothetical protein	3
EHI_198930	-1.52	4.71E-05	Actin, putative	1
EHI_160980	-1.51	3.63E-04	hypothetical protein	1
EHI_170120	-1.51	1.08E-04	hypothetical protein	1
EHI_194190	-1.50	5.77E-03	hypothetical protein	1
EHI_058090	-1.50	3.00E-05	Ras family GTPase	1

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhMyb10+EhCDC5 (sobreexpresados)

Gene ID	FC	p Value	Anotación	# Exones
EHI_006770	8.58	5.682E-08	hypothetical protein, conserved,phosphopantetheine adenylyltransferase, putative	2
EHI_075040	7.74	1.737E-11	GTP-binding protein	1
EHI_123020	7.42	1.158E-05	hypothetical protein,protein phosphatase domain-containing protein	1
EHI_147990	3.90	3.360E-03	hypothetical protein, Ribosomal RNA large subunit methyltransferase H	1
EHI_109010	2.86	2.486E-02	hypothetical protein, surface antigen ariel1, putative*	2
EHI_103400	2.76	9.009E-03	hypothetical protein,Tetraspanin family protein	2
EHI_066970	2.52	2.590E-02	mucin-like protein 1 precursor, putative	1
EHI_114000	2.28	4.331E-02	hypothetical protein, Pecanex C-terminal domain-containing protein*	2
EHI_166860	2.22	3.307E-02	hypothetical protein, RNaseIII, non canonical Dicer, RNA procesing	1
EHI_023400	2.19	1.413E-04	Ribosomal protein S30, putative	1
EHI_103900	2.10	1.047E-03	nucleosome-binding protein 1, putative	1
EHI_073540	2.04	3.795E-03	heat shock protein 101, putative	1
EHI_107170	1.97	3.727E-03	hypothetical protein, surface antigen ariel1, putative*	1
EHI_166800	1.90	1.192E-02	ubiquitin, putative	1
EHI_112590	1.89	1.497E-02	heat shock protein 70 family	1
EHI_057550	1.85	3.824E-05	methionine gamma-lyase, putative	1
EHI_182530	1.69	3.874E-02	hypothetical protein, 15 kDa selenoprotein	2
EHI_079930	1.63	1.851E-02	hypothetical protein, Immunoglobulin-binding protein 1	1
EHI_166370	1.57	2.935E-03	hypothetical protein, surface antigen ariel1, putative*	1
EHI_194730	1.56	7.649E-03	mucin-like protein 1 precursor, putative	1
EHI_070450	1.53	1.358E-02	heat shock protein70, mitochondrial, putative	2

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhMyb10+EhCDC5 (reprimidos).

Gene ID	FC	p Value	Anotación	# Exones
EHI_155900	-9.14	1.49E-08	hypothetical protein, conserved, split ends, isoform B	1
EHI_036630	-8.41	1.20E-07	hypothetical protein, eukaryotic translation initiation factor 3 subunit H	1
EHI_006680	-8.24	1.95E-07	phosphopantetheine adenylyltransferase, putative	2
EHI_140640	-7.61	6.09E-06	F-actin capping protein subunit alpha, putative	1
EHI_089770	-7.52	4.22E-06	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit, putative*	1
EHI_116040	-6.83	7.34E-05	hypothetical protein, dab2-interacting protein, putative*	1
EHI_174150	-6.28	6.54E-04	hypothetical protein, Photosynthetic reaction center PufX protein	2
EHI_068380	-4.63	1.16E-02	WD domain containing protein	1
EHI_032670	-3.72	1.57E-03	hypothetical protein, NADPH-dependent FMN reductase domain containing protein	2
EHI_180390	-3.54	4.23E-02	AIG1 family protein, putative	1
EHI_169350	-3.40	7.59E-07	nonpathogenic pore-forming peptide precursor, putative	1
EHI_039020	-3.31	3.29E-02	actobindin, putative	1
EHI_074380	-3.30	1.68E-03	hypothetical protein, major facilitator superfamily protein	1

EHI_141050	-3.30	1.04E-02	hypothetical protein, Cox19-like CHCH family protein	2
EHI_188880	-3.12	3.68E-05	Chaperonin 10 Kd subunit, putative	1
EHI_172280	-3.11	6.39E-05	hypothetical protein, leukocyte-endothelial cell adhesion molecule 3, putative	1
EHI_099860	-3.04	3.19E-05	iron-sulfur cluster-binding protein, putative	1
EHI_104690	-3.03	3.86E-02	zinc knuckle domain-containing protein	1
EHI_178340	-2.92	1.04E-05	ubiquitin, putative	1
EHI_051690	-2.74	3.20E-03	hypothetical protein, eukaryotic translation initiation factor 3 subunit, putative	1
EHI_166410	-2.65	4.67E-02	myb-like DNA-binding domain containing protein	2
EHI_164840	-2.44	2.56E-02	hypothetical protein, Tetrapanin	1
EHI_017660	-2.43	9.78E-04	hypothetical protein, ataxia telangiectasia mutated, putative	1
EHI_016110	-2.43	4.80E-02	hypothetical protein, RIBBON-HELIX-HELIX, TRANSCRIPTION REPRESSOR	1
EHI_194430	-2.38	8.44E-04	iron-sulfur cluster-binding protein, putative	1
EHI_083100	-2.36	1.09E-03	hypothetical protein, splicing regulatory glutamine/lysine-rich protein 1 isoform b	1
EHI_010020	-2.35	1.16E-04	calmodulin, putative	1
EHI_118270	-2.34	1.24E-03	amoebapore C	1
EHI_162230	-2.32	9.46E-03	alpha-1,3-mannosyltransferase ALG2, putative	2
EHI_118670	-2.32	2.05E-03	Skp1 family protein	1
EHI_142900	-2.30	1.55E-02	hypothetical protein, Ragulator complex protein LAMTOR1	1
EHI_044610	-2.30	2.06E-02	ribosome biogenesis protein NEP1, putative	2
EHI_148400	-2.27	4.76E-02	hypothetical protein, Cohesin subunit SA-1	1
EHI_048140	-2.27	3.26E-04	hypothetical protein, Cysteine proteinases superfamily protein	1
EHI_032780	-2.27	8.02E-03	hypothetical protein, Galactose oxidase/kelch repeat superfamily protein	2
EHI_075530	-2.23	1.15E-03	coatomer subunit beta', putative	1
EHI_179390	-2.23	3.50E-05	Translation initiation factor, putative	2
EHI_156380	-2.22	2.53E-02	U2 snRNP auxiliary factor small subunit, putative	1
EHI_100350	-2.22	5.13E-03	hypothetical protein, conserved, cupin 2 barrel domain containing protein	1
EHI_140770	-2.17	2.62E-02	Rab family GTPase	4
EHI_000620	-2.15	2.05E-02	PP-loop family protein	1
EHI_019860	-2.14	1.03E-02	hypothetical protein	1
EHI_048300	-2.10	9.15E-03	WD repeat protein 68, putative	1
EHI_050000	-2.10	3.53E-03	hypothetical protein	1
EHI_187780	-2.07	8.09E-04	hypothetical protein	1
EHI_000130	-2.06	1.20E-03	calmodulin, putative	2
EHI_159740	-2.06	4.11E-02	hypothetical protein	2
EHI_062500	-2.03	1.68E-04	hypothetical protein, SNAP receptor complex member 1*	3
EHI_156310	-2.00	6.48E-03	ribonuclease, putative	1
EHI_135060	-1.97	1.63E-03	ribosomal protein S17, putative	1
EHI_126130	-1.96	1.24E-02	hypothetical protein, PIG-P protein, putative*	2
EHI_048860	-1.96	1.83E-02	U3 small nucleolar ribonucleo protein MPP10, putative	1

EHI_134840	-1.90	6.07E-03	ADP-ribosylation factor, putative	1
EHI_058320	-1.90	4.69E-03	transmembrane protein tmp21, putative	1
EHI_189950	-1.89	3.14E-02	co-chaperone protein, putative	1
EHI_010030	-1.89	2.46E-03	hypothetical protein	2
EHI_009940	-1.86	7.86E-03	clathrin adaptor complex small chain, putative	3
EHI_134740	-1.84	2.72E-02	NADPH-dependent FMN reductase domain containing protein	1
EHI_011880	-1.82	1.94E-02	hypothetical protein	1
EHI_035670	-1.82	2.99E-04	hypothetical protein	1
EHI_180000	-1.80	8.59E-03	eukaryotic translation initiation factor 4E, putative	1
EHI_028990	-1.80	1.92E-02	hypothetical protein	2
EHI_157360	-1.80	5.07E-03	ALG1 family protein, putative	2
EHI_006960	-1.79	4.98E-02	hypothetical protein	1
EHI_155680	-1.78	3.01E-02	hypothetical protein, SNARE domain containing protein*	3
EHI_095830	-1.77	1.50E-03	hypothetical protein	1
EHI_150530	-1.73	1.77E-02	ATP-binding cassette sub-family E member 1, putative	1
EHI_148190	-1.71	1.04E-03	Ran family GTPase	1
EHI_138790	-1.71	3.02E-02	LSM domain containing protein	1
EHI_006850	-1.71	1.43E-02	hypothetical protein	1
EHI_017750	-1.69	3.04E-04	hypothetical protein	1
EHI_092100	-1.68	4.89E-02	chitinase, putative	1
EHI_090020	-1.68	2.88E-02	ribosomal protein L18a, putative	1
EHI_119540	-1.67	7.11E-04	eukaryotic translation elongation factor 1 gamma, putative	1
EHI_182550	-1.66	9.09E-03	Vacuolar ATP synthase subunit E, putative	1
EHI_046600	-1.66	9.70E-03	acyl-CoA hydrolase, putative	1
EHI_086080	-1.65	1.98E-02	proteasome subunit alpha type 1, putative	1
EHI_038540	-1.64	3.17E-02	hypothetical protein, conserved	2
EHI_017890	-1.63	4.19E-02	60S ribosome subunit biogenesis protein NIP7, putative	1
EHI_191530	-1.63	4.19E-02	60S ribosome subunit biogenesis protein NIP7, putative	1
EHI_126210	-1.63	3.10E-02	histone H2A, putative	1
EHI_170060	-1.62	1.21E-02	ubiquitin like protein	1
EHI_151350	-1.60	1.29E-02	hypothetical protein	1
EHI_092930	-1.60	5.27E-03	hypothetical protein	1
EHI_151890	-1.60	2.89E-02	EF-hand calcium-binding domain containing protein	1
EHI_069950	-1.59	2.48E-02	C2 domain containing protein	1
EHI_148470	-1.59	3.56E-03	asparagine synthetase A, putative	1
EHI_103510	-1.58	1.59E-02	ubiquitin domain protein, putative	1
EHI_188960	-1.57	2.51E-03	histone H2A, putative	1
EHI_103310	-1.54	3.62E-02	60S ribosomal protein L7a, putative	1
EHI_153250	-1.53	3.40E-02	hypothetical protein	1
EHI_125600	-1.52	3.48E-02	hypothetical protein	1
EHI_196370	-1.52	3.67E-02	hypothetical protein, chromosome partition protein, putative*	2
EHI_167310	-1.51	1.52E-02	grainin 2	1

Lista de genes expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhMyb10, EhCDC5 y EhMyb10+EhCDC5

Gene ID	FC	p Value	Anotación	# Exones
<b>Sobreexpresados</b>				
EHI_103400	3.34	9.77E-05	hypothetical protein, Tetraspanin family protein	2
EHI_073540	2.71	1.07E-03	heat shock protein 101, putative	1
EHI_166800	2.49	2.30E-03	ubiquitin, putative	1
EHI_023400	1.87	3.25E-03	Ribosomal protein S30, putative	1
EHI_079930	1.77	8.08E-03	hypothetical protein, Immunoglobulin-binding protein	1
EHI_107170	1.71	3.29E-02	hypothetical protein, surface antigen ariel1, putative*	1
EHI_057550	1.66	2.21E-04	methionine gamma-lyase, putative	1
<b>Reprimidos</b>				
EHI_155900	-9.14	1.49E-08	hypothetical protein, conserved, Adapter-related protein complex 3 (AP-3) subunit	1
EHI_089770	-7.52	4.22E-06	hypothetical protein, Adapter-related protein complex 3 (AP-3) subunit	1
EHI_116040	-6.83	7.34E-05	hypothetical protein, dab2-interacting protein, putative*	1
EHI_032670	-3.72	1.57E-03	hypothetical protein, NADPH-dependent FMN reductase domain containing protein	2
EHI_169350	-3.40	7.59E-07	nonpathogenic pore-forming peptide precursor, putative	1
EHI_172280	-3.11	6.39E-05	hypothetical protein, RanBD1 domain-containing protein	1
EHI_099860	-3.04	3.19E-05	iron-sulfur cluster-binding protein, putative	1
EHI_178340	-2.92	1.04E-05	ubiquitin, putative	1
EHI_194430	-2.38	8.44E-04	iron-sulfur cluster-binding protein, putative	1
EHI_010020	-2.35	1.16E-04	calmodulin, putative	1
EHI_118270	-2.34	1.24E-03	amoebapore C	1
EHI_118670	-2.32	2.05E-03	Skp1 family protein	1
EHI_044610	-2.30	2.06E-02	ribosome biogenesis protein NEP1, putative	2
EHI_048140	-2.27	3.26E-04	hypothetical protein, cysteine proteinases superfamily protein	1
EHI_032780	-2.27	8.02E-03	hypothetical protein, cysteine proteinases superfamily protein	2
EHI_075530	-2.23	1.15E-03	coatamer subunit beta', putative	1
EHI_179390	-2.23	3.50E-05	Translation initiation factor, putative	2
EHI_156380	-2.22	2.53E-02	U2 snRNP auxiliary factor small subunit, putative	1
EHI_100350	-2.22	5.13E-03	hypothetical protein, conserved cupin 2 barrel domain containing protein*	1
EHI_140770	-2.17	2.62E-02	Rab family GTPase	4
EHI_000620	-2.15	2.05E-02	PP-loop family protein	1
EHI_050000	-2.10	3.53E-03	hypothetical protein, COMM domain containing protein, putative*	1
EHI_187780	-2.07	8.09E-04	hypothetical protein, transmembrane protein 42-like*	1
EHI_000130	-2.06	1.20E-03	calmodulin, putative	2
EHI_062500	-2.03	1.68E-04	hypothetical protein, SNAP receptor complex member 1*	3
EHI_135060	-1.97	1.63E-03	ribosomal protein S17, putative	1

EHI_134840	-1.90	6.07E-03	ADP-ribosylation factor, putative	1
EHI_058320	-1.90	4.69E-03	transmembrane protein tmp21, putative	1
EHI_010030	-1.89	2.46E-03	hypothetical protein, Ubiquitin related modifier 1	2
EHI_035670	-1.82	2.99E-04	hypothetical protein, Myosin-4	1
EHI_180000	-1.80	8.59E-03	eukaryotic translation initiation factor 4E, putative	1
EHI_028990	-1.80	1.92E-02	hypothetical protein, Proteasome maturation factor UMP1	2
EHI_157360	-1.80	5.07E-03	AIG1 family protein, putative	2
EHI_006960	-1.79	4.98E-02	hypothetical protein, spindle and kinetochore-associated protein 2 isoform	1
EHI_155680	-1.78	3.01E-02	hypothetical protein, SNARE domain containing protein	3
EHI_095830	-1.77	1.50E-03	hypothetical protein, GCN5-like 1 domain-containing protein	1
EHI_148190	-1.71	1.04E-03	Ran family GTPase	1
EHI_017750	-1.69	3.04E-04	hypothetical protein, axoneme associated protein, putative	1
EHI_090020	-1.68	2.88E-02	ribosomal protein L18a, putative	1
EHI_119540	-1.67	7.11E-04	eukaryotic translation elongation factor 1 gamma, putative	1
EHI_182550	-1.66	9.09E-03	Vacuolar ATP synthase subunit E, putative	1
EHI_046600	-1.66	9.70E-03	acyl-CoA hydrolase, putative	1
EHI_086080	-1.65	1.98E-02	proteasome subunit alpha type 1, putative	1
EHI_126210	-1.63	3.10E-02	histone H2A, putative	1
EHI_170060	-1.62	1.21E-02	ubiquitin like protein	1
EHI_151350	-1.60	1.29E-02	hypothetical protein, vesicle transport protein SEC20 isoform BNIP1-a	1
EHI_092930	-1.60	5.27E-03	hypothetical protein, methionine-tRNA ligase	1
EHI_151890	-1.60	2.89E-02	EF-hand calcium-binding domain containing protein	1
EHI_069950	-1.59	2.48E-02	C2 domain containing protein	1
EHI_103510	-1.58	1.59E-02	ubiquitin domain protein, putative	1
EHI_188960	-1.57	2.51E-03	histone H2A, putative	1
EHI_103310	-1.54	3.62E-02	60S ribosomal protein L7a, putative	1
EHI_125600	-1.52	3.48E-02	hypothetical protein	1
EHI_196370	-1.52	3.67E-02	hypothetical protein, Rho-associated protein kinase, putative	2

Lista de genes exclusivos expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhMyb10.

Gene ID	FC	p Value	Anotación	# Exones
Sobreexpresados				
EHI_018630	7.10	6.58E-05	hypothetical protein, ER membrane protein, BIG1 family	1
EHI_178640	6.69	1.80E-04	Yos1-like domain-containing protein	2
EHI_083380	5.97	2.26E-03	hypothetical protein, protein SOF1, putative	1
EHI_159450	3.73	6.82E-03	hypothetical protein, METALLOPROTEASE	1
EHI_120650	2.92	1.07E-02	hypothetical protein, proteasome regulatory subunit	1

EHI_179380	2.44	2.76E-02	hypothetical protein, Signal recognition particle 9 kDa protein	2
EHI_146540	2.27	3.39E-03	hypothetical protein, Synaptonemal complex protein 1	2
EHI_108640	2.14	1.93E-02	LSM domain containing protein	1
EHI_165810	1.91	3.83E-05	hypothetical protein, pseudogene	1
EHI_051580	1.87	1.96E-02	hypothetical protein, Eukaryotic translation initiation factor 3	1
EHI_040370	1.77	1.39E-02	hypothetical protein,	1
EHI_157010	1.69	8.87E-04	alcohol dehydrogenase, putative	1
EHI_094470	1.56	3.28E-02	heat shock protein 101, putative	1
EHI_048170	1.56	1.85E-02	Protein transport protein SEC61 gamma subunit, putative	2
Reprimidos				
EHI_049990	-8.43	1.00E-07	chaperone proteinDNAJ, putative	1
EHI_004430	-8.37	1.26E-07	hypothetical protein,MYB PROTO-ONCOGENE PROTEIN	1
EHI_183200	-7.29	1.90E-05	hypothetical protein	2
EHI_179850	-6.32	7.23E-04	hypothetical protein	2
EHI_033240	-6.25	8.19E-04	Riboflavin kinase / FAD synthetase, putative	1
EHI_085680	-5.68	4.09E-03	WD repeat protein	1
EHI_159910	-3.67	1.46E-02	ariadne-1, putative	1
EHI_005030	-3.61	4.98E-03	hypothetical protein, ataxia telangiectasia mutated, putative	1
EHI_134600	-3.61	4.98E-03	hypothetical protein, ataxia telangiectasia mutated, putative	1
EHI_170160	-3.26	3.35E-02	hypothetical protein, putative serine threonine-protein phosphatase 2a regulatory subunit b subunit gamma	2
EHI_120950	-2.86	6.04E-04	hypothetical protein	1
EHI_072640	-2.59	3.64E-02	hypothetical protein, conserved, Calpastatin,	1
EHI_092720	-2.30	2.60E-02	hypothetical protein	2
EHI_094810	-2.28	1.04E-02	Ras family protein	1
EHI_118230	-2.10	4.49E-04	hypothetical protein, transcription factor TFIID	1
EHI_045200	-2.00	4.03E-02	hypothetical protein	2
EHI_012420	-2.00	6.94E-04	myb-like DNA-binding domain containing protein	1
EHI_009490	-1.87	3.69E-03	hypothetical protein, transcription factor TFIID	1
EHI_098060	-1.80	1.31E-02	proteasome subunit alpha type 3, putative	1
EHI_008040	-1.78	1.28E-02	Thymidylate kinase, putative	2
EHI_091470	-1.75	4.91E-02	hypothetical protein	2
EHI_153940	-1.75	4.09E-02	hypothetical protein, pseudogene	1
EHI_093840	-1.74	2.87E-02	zinc finger protein, putative	2
EHI_141800	-1.71	6.38E-04	hypothetical protein	1
EHI_045930	-1.70	2.00E-02	hypothetical protein, DNA double-strand break repair Rad50 ATPase, putative	1

EHI_177460	-1.68	6.45E-03	translation initiation factor eIF-5A, putative	1
EHI_188170	-1.66	9.56E-04	hypothetical protein, M protein, serotype 5 precursor, putative	1
EHI_182610	-1.61	2.26E-02	myb-like DNA-binding domain containing protein	1
EHI_126860	-1.61	5.46E-03	hypothetical protein	2
EHI_020080	-1.60	8.47E-05	hypothetical protein, M protein, serotype 5 precursor, putative	1
EHI_040310	-1.60	4.46E-05	Rab family GTPase	2
EHI_053420	-1.59	2.92E-02	Rab family GTPase	2
EHI_076020	-1.54	4.93E-03	hypothetical protein, TATA-binding protein-associated phosphoprotein, putative	1
EHI_156290	-1.53	2.15E-02	BRAC1 domain containing protein	2
EHI_086540	-1.52	1.82E-02	replication factor C subunit 4, putative	1
EHI_127400	-1.52	1.48E-02	hypothetical protein, bZIP transcription factor domain containing protein	1

Lista de genes exclusivos expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhCDC5.

Gene ID	FC	p Value	Anotación	# Exones
<b>Sobreexpresados</b>				
EHI_118220	8.55	3.01E-07	vacuolar ATP synthase subunit B, putative	2
EHI_025840	6.89	1.67E-04	LSM domain containing protein	2
EHI_087650	2.83	1.37E-02	hypothetical protein, extracellular protein (interpro)	1
EHI_192420	2.72	2.31E-02	hypothetical protein, conserved/ kinesin	1
EHI_101730	2.68	1.72E-02	surface antigen ariel1, putative	1
EHI_197140	2.67	3.52E-02	hypothetical protein	1
EHI_080100	2.54	7.37E-03	cylicin-2, putative	1
EHI_011400	2.51	9.95E-03	hypothetical protein, TATA-binding protein-associated phosphoprotein, putative	1
EHI_148260	2.24	4.45E-03	hypothetical protein, U1 zinc finger protein*	1
EHI_140480	2.23	3.36E-02	hypothetical protein	1
EHI_086030	2.11	3.76E-02	hypothetical protein, cytoskeletal protein Sojo, putative	1
EHI_030150	2.09	3.30E-02	hypothetical protein, conserved, synaptonemal complex protein, putative	1
EHI_069510	2.06	4.39E-02	hypothetical protein	1
EHI_170470	1.92	4.26E-02	ER lumen protein retaining receptor	2
EHI_130480	1.85	3.97E-02	hypothetical protein	1
EHI_037820	1.80	3.47E-02	hypothetical protein, conserved domain containing	1
EHI_041580	1.75	1.53E-03	hypothetical protein, conserved	1
EHI_008130	1.74	1.62E-02	Myb family DNA-binding protein, SHAQKYF family	2
EHI_158130	1.73	9.05E-03	hypothetical protein	1
EHI_003240	1.72	3.38E-02	N-acetylmuraminidase, pseudogene	1

EHI_105350	1.72	4.73E-03	hypothetical protein, TATA-binding protein-associated phosphoprotein, putative	1
EHI_058110	1.71	1.54E-02	hypothetical protein, golgin IMH1, putative	1
EHI_068140	1.66	2.30E-02	hypothetical protein, polyamine-modulated factor 1-binding protein, putative	1
EHI_087610	1.64	1.50E-02	hypothetical protein, prp4, putative	3
EHI_159330	1.64	1.50E-02	hypothetical protein, prp4, putative	3
EHI_161900	1.61	4.81E-02	hypothetical protein, E3 ubiquitin protein ligase BRE1, putative	1
EHI_159490	1.60	1.03E-02	hypothetical protein, TATA-binding protein-associated phosphoprotein, putative	1
EHI_126270	1.59	9.66E-03	Ras GTPase-activating protein, putative	2
EHI_014250	1.58	5.14E-04	hypothetical protein	1
EHI_176850	1.52	3.86E-02	hypothetical protein, conserved, GRIP domain containing protein RUD3	1
EHI_122900	1.50	1.23E-03	hypothetical protein, conserved, DNA double-strand break repair Rad50 ATPase, putative	1

---

#### Reprimidos

---

EHI_094020	-10.42	6.42E-12	hypothetical protein, filamin/ABP280 repeat containing protein	1
EHI_020990	-6.66	2.44E-04	hypothetical protein	1
EHI_041880	-6.27	7.09E-04	hypothetical protein	2
EHI_193940	-6.10	4.37E-09	hypothetical protein, calponin domain containing protein	2
EHI_114920	-4.08	5.70E-23	60S ribosomal protein L35a, putative	1
EHI_158570	-3.58	7.47E-11	actobindin, putative	1
EHI_035220	-3.40	7.01E-08	hypothetical protein	1
EHI_092320	-2.99	6.76E-03	hypothetical Cofilin/tropomyosin-type actin-binding protein	1
EHI_109880	-2.71	2.53E-03	hypothetical protein	1
EHI_096200	-2.64	1.79E-02	thioredoxin, putative	1
EHI_052720	-2.57	2.13E-03	hypothetical protein	1
EHI_049000	-2.48	1.11E-14	60S ribosomal protein L35a, putative	1
EHI_007260	-2.46	3.21E-02	hypothetical protein	1
EHI_179990	-2.42	5.37E-03	hypothetical protein, cwf21 protein	1
EHI_160900	-2.37	4.18E-02	Sec1 family protein	1
EHI_023000	-2.34	6.47E-06	hypothetical protein	2
EHI_197480	-2.31	6.84E-11	actophorin, putative	1
EHI_017030	-2.24	2.87E-05	hypothetical protein, conserved	1
EHI_097930	-2.22	1.71E-04	muT/nudix family protein	1
EHI_118750	-2.15	2.59E-08	hypothetical Cofilin,tropomyosin-type actin-binding protein	1
EHI_126870	-2.14	7.77E-11	40S ribosomal protein S21, putative	2
EHI_081130	-2.08	8.67E-04	P-glycoprotein 6, putative	2
EHI_160210	-2.07	8.65E-04	hypothetical protein, conserved, peptidyl-tRNA hydrolase, putative	2
EHI_186720	-2.06	1.79E-10	40S ribosomal protein S12, putative	1

EHI_065340	-2.05	5.74E-07	hypothetical protein	1
EHI_027610	-2.03	2.38E-10	hypothetical protein	2
EHI_115360	-2.03	3.73E-04	hypothetical protein	2
EHI_030390	-1.96	5.75E-04	hypothetical protein	1
EHI_026360	-1.95	5.66E-03	phosphoserine aminotransferase, putative	1
EHI_139050	-1.89	1.57E-02	hypothetical protein	1
EHI_024240	-1.87	1.64E-04	aldehyde-alcohol dehydrogenase 2, putative	3
EHI_201940	-1.87	5.13E-03	clathrin heavy chain, putative	1
EHI_087570	-1.85	1.04E-06	endoribonuclease L-PSP, putative	1
EHI_066770	-1.83	2.72E-08	Skp1 family protein	1
EHI_174180	-1.83	2.72E-08	Skp1 family protein	1
EHI_141890	-1.82	2.50E-04	zinc finger protein containing CCHC type domain	1
EHI_146780	-1.82	3.37E-03	hypothetical protein, serine/threonine-protein kinase, putative	1
EHI_194050	-1.81	3.99E-03	hypothetical protein	1
EHI_050290	-1.79	2.23E-02	SGS domain protein	1
EHI_153260	-1.78	3.40E-03	hypothetical protein, conserved, protein YIF1A, putative	2
EHI_118290	-1.77	4.03E-02	hypothetical protein, Hus1 family protein, putative	1
EHI_086200	-1.76	1.27E-03	hypothetical protein, EhSyntaxin K, putative	2
EHI_087410	-1.75	1.41E-04	high mobility group (HMG) box domain containing protein	2
EHI_159820	-1.72	6.91E-04	hypothetical protein	2
EHI_193350	-1.70	5.78E-04	3'(2'),5'-bisphosphate nucleotidase, putative	1
EHI_031240	-1.69	4.71E-04	serine/threonine-protein phosphatase, putative	1
EHI_099720	-1.66	3.33E-03	hypothetical protein	2
EHI_169300	-1.66	3.74E-03	ribonuclease, putative	1
EHI_029210	-1.66	6.17E-03	proteasome subunit alpha type 3, putative	1
EHI_006170	-1.65	4.94E-04	eukaryotic translation initiation factor 6, putative	1
EHI_141010	-1.64	3.05E-02	hypothetical protein	1
EHI_196580	-1.63	2.18E-03	phosphoribulokinase /uridine kinase family protein	1
EHI_050520	-1.63	8.58E-06	hypothetical protein	1
EHI_140550	-1.62	1.60E-06	60S ribosomal protein L6, putative	1
EHI_147570	-1.61	2.46E-06	Rho GDP exchange inhibitor, putative	1
EHI_030070	-1.59	8.67E-07	hypothetical protein, ribosomal protein S29, putative	1
EHI_015440	-1.58	8.39E-05	Rho family GTPase	2
EHI_105120	-1.57	2.04E-02	hypothetical protein, conserved/ protein TRANSPARENT TESTA GLABRA, putative	1
EHI_013340	-1.57	1.81E-03	hypothetical protein, Myb family DNA-binding protein, SHAQKYF family protein	1
EHI_174910	-1.57	1.22E-04	ARP2/3 complex 21 kDa subunit, putative	2

EHI_197410	-1.56	4.28E-02	cation transporter, putative	1
EHI_186950	-1.56	4.15E-05	protein phosphatase domain-containing protein	2
EHI_104360	-1.55	1.08E-04	nucleoside diphosphate kinase, putative	1
EHI_163670	-1.55	5.84E-05	ribosomal protein L18a, putative	1
EHI_134490	-1.54	1.19E-05	F-actin capping protein subunit beta, putative	1
EHI_056460	-1.54	1.02E-04	hypothetical protein, conserved, rRNA-processing protein FCF1 family protein, putative	1
EHI_198550	-1.54	1.68E-02	glucosamine 6-phosphate N-acetyltransferase. putative	2
EHI_063340	-1.53	2.44E-02	prefoldin, alpha subunit	2
EHI_087380	-1.53	2.44E-02	prefoldin, alpha subunit	2
EHI_161870	-1.52	6.83E-03	hypothetical protein	1
EHI_035510	-1.52	1.26E-03	hypothetical protein	3
EHI_198930	-1.52	4.71E-05	Actin, putative	1
EHI_160980	-1.51	3.63E-04	hypothetical protein, M protein, serotype 5 precursor, putative	1
EHI_170120	-1.51	1.08E-04	hypothetical protein	1

Lista de genes exclusivos expresados diferencialmente ( $\geq 1,5$  veces) en trofozoítos de *E. histolytica* que sobreexpresan a las proteínas EhMyb10+EhCDC5.

Gene ID	FC	p Value	Anotación	# Exones
<b>Sobreexpresados</b>				
EHI_075040	7.74	1.74E-11	GTP-binding protein	1
EHI_147990	3.90	3.36E-03	hypothetical protein	1
EHI_109010	2.86	2.49E-02	hypothetical protein, surface antigen ariel1, putative	2
EHI_066970	2.52	2.59E-02	mucin-like protein 1 precursor, putative	1
EHI_114000	2.28	4.33E-02	hypothetical protein, pecanex-like protein 1 isoform x5	2
EHI_112590	1.89	1.50E-02	heat shock protein 70 family	1
EHI_182530	1.69	3.87E-02	hypothetical protein	2
EHI_166370	1.57	2.94E-03	hypothetical protein, surface antigen ariel1, putative	1
EHI_194730	1.56	7.65E-03	mucin-like protein 1 precursor, putative	1
EHI_070450	1.53	1.36E-02	heat shock protein70, mitochondrial, putative	2
<b>Reprimidos</b>				
EHI_036630	-8.41	1.20E-07	hypothetical protein	1
EHI_006680	-8.24	1.95E-07	phosphopantetheine adenylyltransferase, putative	2
EHI_174150	-6.28	6.54E-04	hypothetical protein	2
EHI_074380	-3.30	1.68E-03	hypothetical protein, transporter, major facilitator family	1
EHI_104690	-3.03	3.86E-02	zinc knuckle domain-containing protein	1
EHI_051690	-2.74	3.20E-03	hypothetical protein, eukaryotic translation initiation factor 3 subunit, putative	1
EHI_166410	-2.65	4.67E-02	myb-like DNA-binding domain containing protein	2

EHI_016110	-2.43	4.80E-02	hypothetical protein	1
EHI_142900	-2.30	1.55E-02	hypothetical protein	1
EHI_019860	-2.14	1.03E-02	hypothetical protein	1
EHI_159740	-2.06	4.11E-02	hypothetical protein	2
EHI_189950	-1.89	3.14E-02	co-chaperone protein, putative	1
EHI_134740	-1.84	2.72E-02	NADPH-dependent FMN reductase domain containing protein	1
EHI_011880	-1.82	1.94E-02	hypothetical protein, COMM domain-containing protein, putative	1
EHI_138790	-1.71	3.02E-02	LSM domain containing protein	1
EHI_006850	-1.71	1.43E-02	hypothetical protein, zinc finger protein, putative	1
EHI_092100	-1.68	4.89E-02	chitinase, putative	1
EHI_038540	-1.64	3.17E-02	hypothetical protein, conserved, RIKEN cDNA family protein, putative	2
EHI_153250	-1.53	3.40E-02	hypothetical protein, U1 zinc finger protein, putative	1

---

## Congresos asistidos



Acknowledgements

Participation:

**ORAL**



**XXX**  
INTERNATIONAL  
SEMINAR ON  
AMEBIASIS  
**2023**

to: **Elisa Azuara Liceaga**

*"Unraveling the role of CDC5 protein in Entamoeba histolytica: bridging transcriptional factors and splicing machinery"*

Claudia Castillo Mercado, Patricia Cuellar Silva, Elizabeth J. Castañeda-Ortiz, Abigail Betanzos, Mario Alberto Rodríguez Rodríguez, Carlos Martínez Rodríguez, Jesús Valdes, Elisa Azuara Liceaga\*

For the presentation of his work, during the "XX International Seminar on Amebiasis 2023", held from October 24 to 27 in the city of Merida, Yucatan, Mexico.

Cecilia Jiménez PhD

Nancy Guillén PhD

Victor Tsutsumi PhD

Tomoyoshi Nozaki PhD

