



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

RESPUESTA TRANSCRIPTÓMICA AL ÁCIDO
TAUROCÓLICO EN LA ACTIVACIÓN Y
DESARROLLO TEMPRANO *IN VITRO* DEL
CISTICERCO DE *TAENIA SOLIUM*:

EXPRESIÓN DIFERENCIAL, ANÁLISIS
FUNCIONAL Y REDES DE COEXPRESIÓN E
INTERACCIÓN PROTEÍNA-PROTEÍNA

TESIS PARA OPTAR EL GRADO DE GRADO
DE MAESTRO EN BIOQUÍMICA Y
BIOLOGÍA MOLECULAR

DAVID CASTAÑEDA CARPIO

LIMA – PERÚ

2025

ASESORA

Dr. rer. nat. Cristina Guerra Giraldez

CO ASESOR

Renzo Marcelo Gutierrez Loli, M.Sc.

JURADO DE TESIS

Luis Julio Cesar De Stefano Beltran, Ph.D.
Presidente

Rosa Patricia Herrera Velit, Ph.D.
Vocal

Dr. rer. nat. Felipe Carlos Yon Torres
Secretario

DEDICATORIA

A mi familia, por todo el apoyo y las enseñanzas

AGRADECIMIENTOS

Agradezco a toda mi familia por todo su apoyo desde el inicio de mi formación como científico. A mis papás, Delia y David, que todo el tiempo han estado alentándome incondicionalmente y dándome las fuerzas para seguir adelante. A Sofia, Daniel y Juan Pablo, por toda la alegría que me dan en los momentos difíciles. A Leonela, por siempre acompañarme, escucharme y apoyarme en esta aventura.

A la Dra. Cristina Guerra, por permitirme aprender junto a su equipo desde mi tercer año en el pregrado y por contar conmigo para un proyecto tan ambicioso como lo ha sido este. Muchas gracias por confiar en mí como persona, y también en mi potencial y mis capacidades.

Renzo, Juan, José, Vale, Shego, Maca y Camila. Gracias por todo lo que me han ayudado a aprender, los consejos (profesionales y de vida) y por todo el tiempo compartido. En ningún lugar me he sentido tan a gusto como en el 218.

Y agradezco a mis amigos de la universidad y del colegio, sin ellos no habría llegado hasta aquí. Se que siempre estaremos compartiendo en donde sea que la vida nos lleve.

FUENTES DE FINANCIAMIENTO

Beca de Posgrado otorgada por la Fundación Cobián
Facultad de Ciencias e Ingeniería, UPCH

CONCYTEC - Programa Nacional de Investigación Científica y Estudios
Avanzados (Prociencia).

CONTRATO No PE501079376-2022-PROCIENCIA al proyecto “Identificación transcriptómica de determinantes celulares en la transición *in vitro* de larva metacestode (cisticerco) a gusano adulto en *Taenia solium*”; convocatoria del Esquema Financiero E041-2022-01 “PROYECTOS DE INVESTIGACIÓN BÁSICA”

DECLARACIÓN DE AUTOR			
FECHA	24	07	2025
APELLIDOS Y NOMBRES DEL EGRESADO	CASTAÑEDA CARPIO DAVID		
PROGRAMA DE POSGRADO	MAESTRÍA EN BIOQUÍMICA Y BIOLOGÍA MOLECULAR		
AÑO DE INICIO DE LOS ESTUDIOS	2025		
TITULO DEL TRABAJO DE INVESTIGACIÓN DE GRADO	“Respuesta transcriptómica al ácido taurocólico en la activación y desarrollo temprano <i>in vitro</i> del cisticerco de <i>Taenia solium</i> : expresión diferencial, análisis funcional y redes de coexpresión e interacción proteína-proteína”		
MODALIDAD DE TRABAJO DE GRADO	TESIS		
Declaración del Autor El presente Trabajo de Grado es original y no es el resultado de un trabajo en colaboración con otros, excepto cuando así está citado explícitamente en el texto. No ha sido ni enviado ni sometido a evaluación para la obtención de otro grado o diploma que no sea el presente.			
Teléfono de contacto (fijo / móvil)	943543505		
E-mail	david.castaneda.c@upch.pe		



Firma del Egresado
DNI: 71878478

ÍNDICE

RESUMEN

ABSTRACT

I.	INTRODUCCIÓN	1
II.	PREGUNTA DE INVESTIGACIÓN Y OBJETIVOS	10
III.	MARCO TEÓRICO	12
	1. Estructura, metabolismo, biología molecular y del desarrollo de céstodos	12
	2. Transcriptómica de céstodos	29
	3. Herramientas para el análisis transcriptómico en <i>T. solium</i>	32
IV.	METODOLOGÍA	50
	1. Diseño general	50
	2. Mapeo al genoma de referencia y generación de archivo de conteo, y análisis de expresión diferencial	50
	3. Anotación funcional del transcriptoma de <i>Taenia solium</i>	51
	4. Análisis de sobrerrepresentación de ontología génica (GO)	52
	5. Análisis de enriquecimiento en conjuntos de genes (GSEA)	52
	6. Análisis de redes de coexpresión genética ponderadas (WGCNA)	53
	7. Redes de interacción/asociación proteína-proteína (STRING)	53
	8. Aspectos éticos y de bioseguridad	54
V.	RESULTADOS	55
VI.	DISCUSIONES	106
	1. Consideraciones del cultivo y su efecto en el perfil transcripcional de <i>T. solium</i>	109
	2. Evidencias de la regulación transcripcional durante la evaginación del escólex y la estimulación por AT	110
	3. Los procesos biológicos asociados a la evaginación del escólex y el efecto del AT	113
VII.	LIMITACIONES Y RECOMENDACIONES	132

VIII. CONCLUSIONES	134
IX. REFERENCIAS BIBLIOGRÁFICAS	136
X. ANEXOS	155

RESUMEN

La fijación de la larva metacéstode de *Taenia solium* al intestino humano es fundamental para la transición del parásito a su estadio adulto. Un proceso clave es la evaginación del escólex, la cual es inducida por sales biliares y enzimas digestivas. Los mecanismos moleculares asociados a este evento no están completamente entendidos. El presente proyecto busca caracterizar la evaginación del escólex a nivel transcriptómico en un modelo *in vitro* de inducción por ácido taurocólico (AT). Se validaron las condiciones experimentales de acuerdo con los perfiles de expresión de parásitos antes y después de evaginar el escólex mediante PCA y agrupamiento jerárquico. Se anotó funcionalmente el transcriptoma de *T. solium* mediante Blast2GO y se identificaron transcritos diferencialmente expresados entre parásitos evaginados y no evaginados en presencia o ausencia de AT mediante DESeq2. Se realizaron análisis de enriquecimiento de ontología génica (GO) y de conjuntos de genes (GSEA). Se determinaron módulos de genes coexpresados y asociados a la fase de evaginación y presencia de AT. Estos fueron validados mediante redes de interacción proteína-proteína en la base de datos STRING. Los resultados evidencian una reprogramación transcripcional de procesos y genes asociados al metabolismo, proliferación, formación del tegumento, síntesis de proteínas, percepción del entorno, sistema nervioso, y regulación génica. Asimismo, la expresión de elementos de vías de señalización conservadas se asocia al efecto del AT. Estos hallazgos aportan nueva evidencia sobre los mecanismos moleculares de la transición de metacéstode a adulto y posicionan al AT como una señal moduladora que guía el desarrollo del parásito.

PALABRAS CLAVE

Taenia solium; Evaginación del escólex; Transcriptómica; Ácido taurocólico;
Desarrollo de céstodos.

ABSTRACT

The attachment of the *Taenia solium* metacestode larva to the human intestine is essential for the parasite's transition to its adult stage. A key step in this process is scolex evagination, which is induced by bile salts and digestive enzymes. However, the molecular mechanisms underlying this event remain poorly understood. This study aims to characterize scolex evagination at the transcriptomic level using an *in vitro* model induced with taurocholic acid (TA). Experimental conditions were validated based on gene expression profiles of parasites before and after evagination using PCA and hierarchical clustering. The *T. solium* transcriptome was functionally annotated using Blast2GO, and differentially expressed transcripts were identified between evaginated and non-evaginated parasites in the presence or absence of TA using DESeq2. Gene Ontology (GO) and Gene Set Enrichment Analyses (GSEA) were performed. Modules of co-expressed genes associated with evagination and TA exposure were identified and validated through protein-protein interaction networks using the STRING database. Results reveal transcriptional reprogramming of processes and genes involved in metabolism, proliferation, tegument formation, protein synthesis, environmental sensing, nervous system function, and gene regulation. Additionally, the expression of components from conserved signaling pathways was associated with TA exposure. These findings provide new insights into the molecular mechanisms underlying the metacestode-to-adult transition and position TA as a modulatory signal that guides parasite development.

KEYWORDS

Taenia solium; Evagination of the scolex; Transcriptomics; Taurocholic acid;
Cestode development

I. INTRODUCCIÓN

Taenia solium es un céstodo que en estadio larval (larva metacéstode o cisticerco) se adhiere al intestino humano mediante su órgano de fijación, el escólex, donde se desarrolla hacia su estadio adulto. Previo a ello, la larva consiste en una vesícula de cuyo interior el escólex evagina en un proceso de activación gatillado por componentes propios del hospedero como enzimas digestivas y sales biliares (1,2).

Se desconocen las implicancias de este proceso a nivel molecular y qué eventos son los que desencadena la activación en el intestino del hospedero. Asimismo, se desconoce si el proceso de evaginación del escólex es un evento mecánico o si involucra un programa molecular finamente regulado. Este trabajo es el primero en utilizar secuenciación de siguiente generación (NGS) para obtener una caracterización transcriptómica de este proceso *in vitro* en condiciones que buscan emular algunas de las variables del ambiente interno del hospedero. Entender estos mecanismos es clave para la comprensión de la biología celular y del desarrollo del parásito, así como también la búsqueda de nuevos fármacos antihelmínticos.

Importancia clínica y prevalencia de las enfermedades causadas por *Taenia solium*

Las enfermedades por *Taenia solium* son una de las mayores causas de mortalidad por enfermedades alimentarias de acuerdo a la Organización Mundial de la Salud (OMS) (3). La infección del cerebro por el estadio larval del parásito se denomina neurocisticercosis, y representa la mayor causa de epilepsia en regiones endémicas. Entre estas, áreas rurales en el Perú con condiciones de saneamiento deficientes donde la crianza de cerdos es artesanal y estos se encuentran expuestos a heces

humanas (4,5). A modo de referencia, la OMS reportó 51 países endémicos y 14 posibles endémicos de 196 territorios evaluados en su actualización del mapa de endemicidad de *T. solium*, donde también se destaca la necesidad de monitoreo subnacional (3). La prevalencia de estas enfermedades es difícil de estimar dado que tanto la infección por el estadio larval y el adulto pueden ser asintomáticas (6).

Ciclo de vida y desarrollo de *Taenia solium*

Taenia solium es un organismo parásito bilateral acelomado perteneciente a la clase Cestoda del phylum Platyhelminthes. Los céstodos son gusanos planos con forma de cinta cuyos estadios adultos parasitan mayoritariamente el tracto digestivo de sus hospederos (1).

La figura 1 ilustra el ciclo de vida de este parásito. Los huevecillos de *T. solium*, producidos por su estadio adulto, son liberados en las excretas del hospedero definitivo y son consumidos por un hospedero intermediario, sea el humano o el cerdo (2). En el tracto digestivo, por medio de enzimas digestivas y sales biliares, se libera el embrión hexacanto mediante la digestión de la membrana oncosférica (7). Este estadio corresponde a la larva del parásito (1,8), la cual atraviesa el epitelio intestinal hasta llegar al torrente sanguíneo, por el cual migrará al músculo esquelético, corazón (produciendo cisticercosis) y/o cerebro (produciendo neurocisticercosis) principalmente (2). La infección por el estadio larval (cisticercosis) o adulto (teniasis) de *T. solium* es un problema de salud pública en países de economías emergentes (2,9).

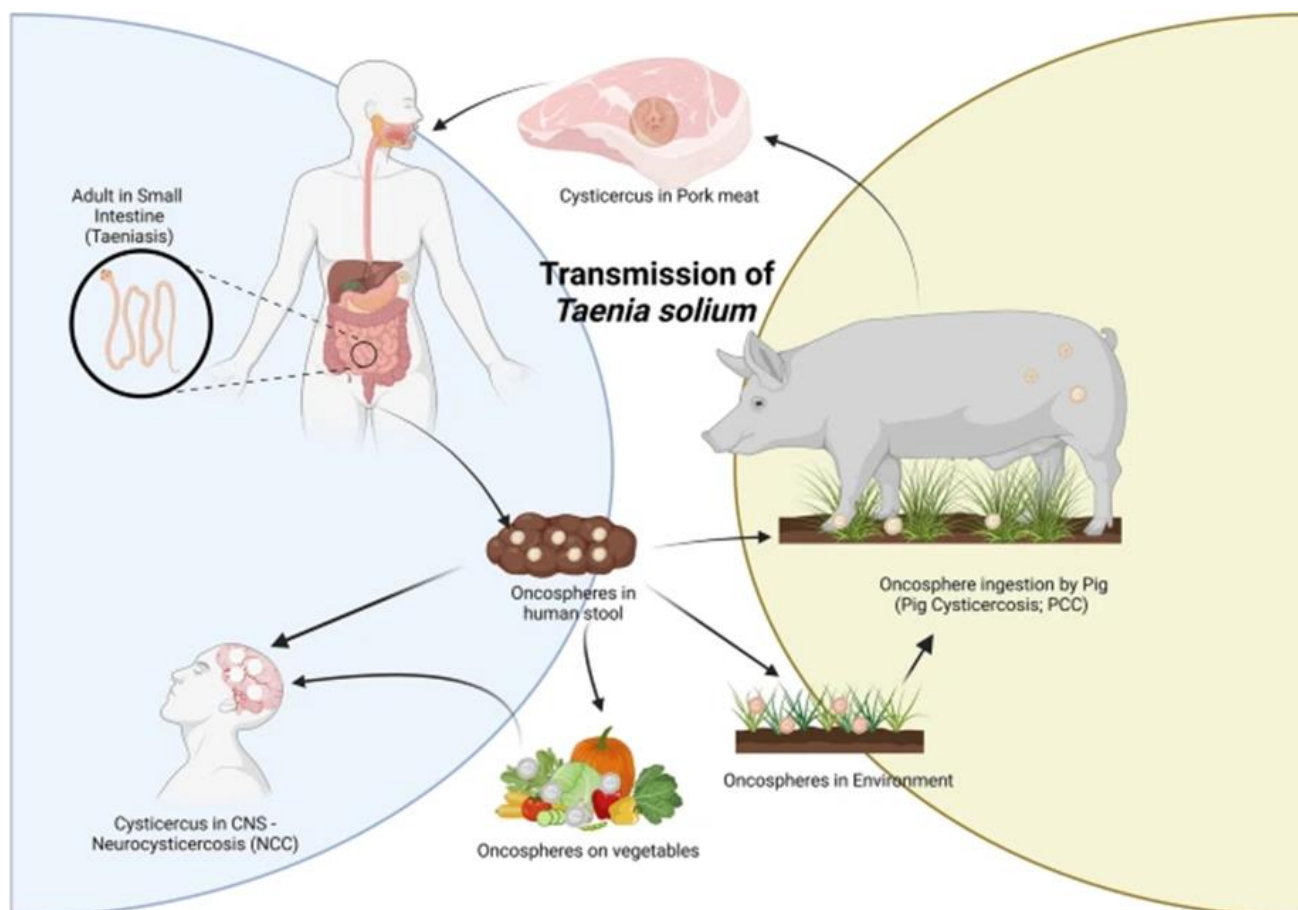


Figura 1. Ciclo de vida y transmisión de *Taenia solium*. Tomado de Hossain MS, *et al.* (4)

El estadio larval madura en estos tejidos hacia su estadio juvenil conocido como larva metacéstodo. Este se compone de una vesícula que contiene una forma del parásito que ya desarrolló un escólex, cuello y algunos segmentos. Estas estructuras se encuentran conservadas en la mayoría de céstodos (1,2,10). El escólex es una estructura de fijación compuesta por 4 ventosas y un rostelo. El cuello es una estructura no segmentada compuesta por células totipotentes. En el céstodo *Hymenolepis diminuta*, el cuello dirige la formación de proglótides (o segmentos) hacia su extremo posterior y presenta capacidad regenerativa (11).

Las larvas metacéstodos (también conocidas como quistes o cisticercos), que habitan la carne de cerdos infectados, son consumidas por el humano. Al llegar al

tracto digestivo, el juvenil emerge de la vesícula evaginando el escólex. Este proceso es gatillado por la presencia de sales biliares y enzimas digestivas del hospedero (1,12), y es indispensable para que el parásito se fije al epitelio intestinal y se desarrolle a su estadio adulto. La evaginación del escólex *in vitro* ocurre espontáneamente. Sin embargo, se puede inducir (y acelerar) incluyendo sales biliares y enzimas en el medio de cultivo (1,12,13). La multiplicación de las células proliferativas del cuello del parásito dirige y permite su crecimiento (13). Se desconoce, sin embargo, si durante la evaginación del escólex componentes de las sales biliares inducen la expresión de genes asociados a mecanismos de división celular y morfogénesis, así como también reguladores de la expresión de proteínas que contribuyen al aumento de la biomasa del parásito durante su crecimiento.

El estadio adulto de *T. solium* se alcanza únicamente tras la fijación del juvenil al epitelio intestinal humano (su hospedero definitivo), y puede llegar a los 2 a 4 metros de longitud, en promedio (2). Adicionalmente al escólex y cuello del juvenil, el estadio adulto presenta un estróbilo pseudosegmentado compuesto por cientos de proglótides. Cada una de estas contiene los aparatos reproductores masculinos y femeninos, responsables de la producción continua de miles de huevecillos, los cuales serán liberados al ambiente a través de las heces del hospedero (9). El estadio adulto es por ende el agente diseminador del parásito, por lo que la interrupción de su desarrollo hacia este estadio resulta un blanco atractivo para la búsqueda de drogas y vacunas.

Trabajo de laboratorio con *T. solium*

En el caso de *T. solium*, el mantenimiento del ciclo de vida completo en el laboratorio no se ha logrado dado que el hospedero definitivo es el humano. En este ocurre la activación del metacéstodo y su desarrollo hasta el estadio adulto. Adicionalmente, la parte más sencilla del trabajo experimental con este parásito, la obtención de cisticercos viables del músculo de cerdos infectados implica un considerable esfuerzo logístico. Por ende, el estudio del desarrollo de este parásito afronta dificultades técnicas, biológicas, éticas y legales. Adicionalmente, no existen modelos *in vitro* que permitan reconstruir transiciones completas entre estadios en *T. solium*. El estudio del desarrollo de este parásito se limita a conocer mecanismos de desarrollo de organismos emparentados de fácil mantenimiento, como el caso de los céstodos del género *Hymenolepis* (11). Por otra parte, estos estudios se enfocan en evaluar la abundancia de proteínas o transcritos en estadios bien descritos del desarrollo, como por ejemplo tetratiridio y adulto en *Mesocostoides corti* (14,15), o varias etapas de cisticercoide y adulto en *Hymenolepis microstoma* (16).

Como se ha evidenciado, a pesar del estudio de diferentes estadios de desarrollo de céstodos a nivel fisiológico y molecular, se desconocen los procesos activados por acción de las sales biliares que inducen la transición desde el estadio larval metacéstode (con el escólex no evaginado) a juvenil (con el escólex evaginado) en *T. solium*. Según reportes en otros céstodos (principalmente en su estadio adulto), la transición de larva a adulto regularía la expresión de transcritos asociados al transporte y excreción de metabolitos, metabolismo energético (asociado

principalmente a glucosa), proliferación celular, formación y remodelamiento de matriz extracelular, síntesis y degradación de proteínas (14–16).

El transcriptoma de *T. solium*

En el laboratorio de Proliferación Celular y Regeneración de la Universidad Peruana Cayetano Heredia (LID-218), se estandarizó el cultivo y evaginación *in vitro* de cisticercos en medio RPMI 1640 en presencia y ausencia de ácido taurocólico al 0.1% (AT) como parte del proyecto de investigación “Identificación transcriptómica de determinantes celulares en la transición *in vitro* de larva metacestode (cisticerco) a gusano adulto en *Taenia solium*” (PE501079376-2022; SIDISI: 209290). Se describen a continuación las condiciones para construir este transcriptoma, publicado en *Scientific data*, como antecedente fundamental para este trabajo de tesis (Figura 2A, 2B) (17).

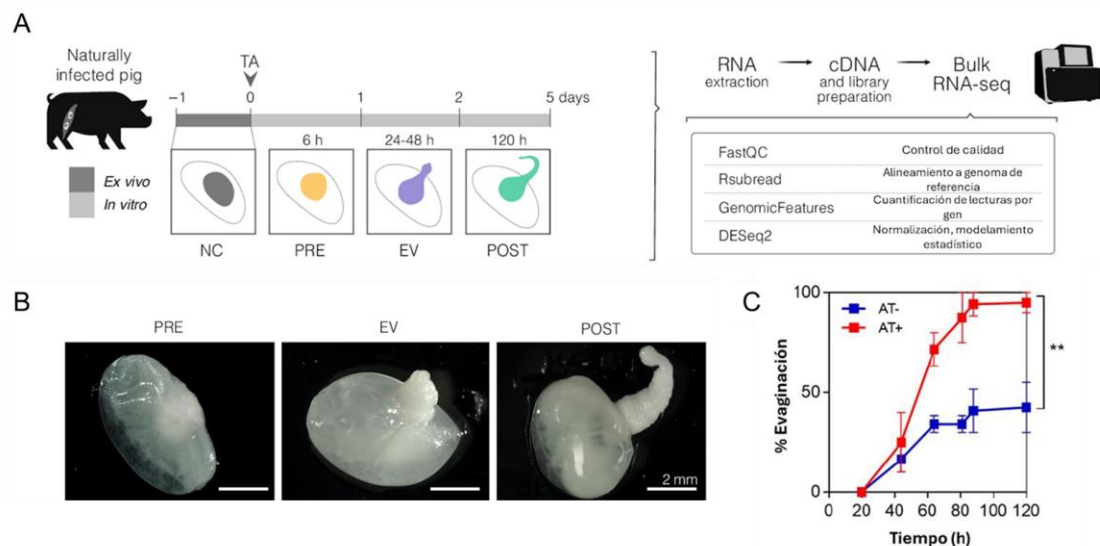


Figura 2. Descripción del Diseño experimental para la construcción de bibliotecas y secuenciación. (A) Grupos de parásitos según fases de desarrollo y flujo de análisis de datos. (B) Morfología de las fases de desarrollo definidas para la secuenciación. (C) Curva de evaginación de *T. solium* en presencia y ausencia de 0.1% AT (dos experimentos: n1 = 34, n2 = 60). Evaluado mediante prueba-t de dos colas al comparar porcentajes de evaginación escólex a 120h (**p<0.01). Adaptado de Castañeda *et al.* (17)

En dos experimentos independientes, se observó un incremento considerable en el porcentaje de individuos con el escólex evaginado (% Evaginación) en los grupos cultivados en presencia de AT con respecto a los que no tuvieron AT en el medio (Figura 2C).

Para determinar los genes diferencialmente expresados con un rol esencial durante la evaginación del escólex y el efecto del ácido taurocólico, se secuenció el RNA extraído de los siguientes grupos de parásitos, donde los individuos fueron escogidos por su morfología:

- **Parásitos no cultivados (NC):** Individuos que no se colocaron en cultivo; su escólex no había evaginado.
- **Parásitos “no evaginados” (PRE):** mantenidos por 6 horas en medio RPMI 1640 en presencia o ausencia de 0.1% AT (PRE_AT o PRE_SAT, respectivamente); escólex no evaginado.
- **Parásitos “recién evaginados” (EV):** cultivados en medio RPMI 1640 en presencia o ausencia de 0.1% AT (EV_AT o EV_SAT, respectivamente), colectados cuando mostraron el escólex fuera de la vesícula (cabeza y cuello visibles, o retraídos en el parénquima vestibular). Estas formas se observaron entre las 24 y 48 h en cultivo.
- **Parásitos “evaginados en cultivo 120h” (POST):** parásitos evaginados (cabeza y cuello visibles, o retraídos en el parénquima vestibular) cultivados en medio RPMI 1640 en presencia o ausencia de 0.1% AT (POST_AT o POST_SAT respectivamente) con 120h de cultivo.

La metodología de extracción de RNA, construcción de librerías y control de calidad inicial mediante FastQC (18) de este experimento fueron publicados (17). Adicionalmente, los datos de secuenciación son accesibles desde el *Gene Expression Omnibus* (<https://www.ncbi.nlm.nih.gov/geo/>, GSE288552). Este transcriptoma fue construido en base al alineamiento de lecturas al genoma de *T. solium* disponible en WormBase ParaSite (19).

Problema de investigación

Previamente, se ha reportado el uso de bilis para inducir la evaginación del escólex de *T. solium* (1), y la eclosión de oncósferas (7,10). Una tesis de nuestro laboratorio optimizó el uso de ácido taurocólico como inductor de este proceso, comparando los efectos de tres concentraciones de AT y también tripsina, además de lograr la inhibición de la evaginación usando como blanco farmacológico la vía de desarrollo Notch (20).

De acuerdo con estudios en modelos de laboratorio de otros céstodos, se espera un efecto inductor o represor de la expresión de genes por parte de la presencia del ácido taurocólico en el cultivo, lo cual se refleja en cambios en el perfil transcripcional de conjuntos de genes con funciones específicas asociadas al desarrollo o metabolismo.

Desde la publicación de los primeros cuatro genomas nucleares de céstodos en el año 2013 (19) se ha expandido el conocimiento de la biología de estos organismos con el objetivo de atacar las enfermedades que causan. Sin embargo, la mayoría de estos genomas se construyen en datos ómicos de organismos que han sido adoptados en los últimos años como modelos de laboratorio por su relativa facilidad de cultivo. Consecuentemente, los genomas de estos céstodos son de mayor calidad

y son origen de inferencias atribuidas a todo este grupo de organismos. La genómica de céstodos es un área del conocimiento en expansión (21).

El presente trabajo es el primero en utilizar técnicas de alto rendimiento para determinar el papel de las sales biliares en el desarrollo de *T. solium*, particularmente a nivel de la evaginación del escólex, como elemento necesario para el establecimiento del parásito y su crecimiento en el intestino humano, lo cual conlleva a su reproducción y por ende la diseminación de huevecillos infecciosos. Los resultados mostraron expresión diferencial en procesos estructurales, celulares y moleculares, los cuales se organizan y sintetizan en categorías para presentar un panorama general de este evento.

II. PREGUNTA DE INVESTIGACIÓN Y OBJETIVOS

¿Cuáles genes y procesos biológicos se asocian a la evaginación del escólex de *Taenia solium* y la presencia de ácido taurocólico, como preámbulo de eventos en el desarrollo de la forma adulta?

1. OBJETIVO GENERAL:

Describir el efecto a nivel transcripcional del AT y la evaginación del escólex en el desarrollo *in vitro* de *T. solium*.

2. OBJETIVOS ESPECÍFICOS:

- a. Determinar la expresión diferencial de genes entre estadios metacéstode antes y después de evaginar el escólex, cultivados en presencia y ausencia de ácido taurocólico.
- b. Realizar la anotación funcional, basada en homología, de los genes con expresión diferencial identificados.
- c. Determinar las categorías funcionales sobrerrepresentadas en el conjunto de genes diferencialmente expresados al comparar fases de desarrollo o presencia del inductor de evaginación.
- d. Determinar las categorías funcionales sobrerrepresentadas en conjuntos de genes con perfiles de expresión correlacionados y establecer su asociación a fases de desarrollo o presencia del inductor de evaginación.

- e. Validar *in silico* conjuntos de genes con perfiles de expresión correlacionados estableciendo redes de interacción/asociación proteína-proteína y enriquecimiento de vías de señalización.

III. MARCO TEÓRICO

1. Estructura, metabolismo, biología molecular y del desarrollo de los céstodos

1.1. Estructuras anatómicas generales de los céstodos

La anatomía de la larva metacéstode (o juvenil) varía de acuerdo con la especie del céstodo. En el caso de *T. solium*, está formada por una vesícula que proyecta hacia su interior una doble invaginación que forman dos espacios conocidos como vestíbulo y canal espiral. El interior de este último contiene las ventosas y ganchos del escólex del parásito. Durante el proceso de evaginación, las ventosas y el rostelo recorren el canal espiral hasta que este último desaparece. Las estructuras resultantes son el escólex y cuello evaginados (13).

El estadio juvenil y el adulto de los céstodos se caracterizan por no tener un sistema digestivo. La obtención de nutrientes se da a través de su tegumento. Este consiste en una serie de células (citones) con una proyección citoplasmática fusionada a los citones contiguos en lo que se conoce como citoplasma distal. Asimismo, proyecciones denominadas microtrichias emergen de la superficie del tegumento. La estructura del tegumento de los céstodos comparte similitudes estructurales con las células del epitelio intestinal humano, tejido con el cual compite por nutrientes (1). Adicionalmente, el tegumento posee un glicocáliz de mucopolisacáridos y glicoproteínas, cuya función, se especula, es la retención de nutrientes y resistencia a la digestión por enzimas del hospedero (1). A pesar de los estudios morfológicos y de ultraestructura del tegumento, las proteínas que

constituyen y contribuyen a la formación de este tejido en el desarrollo del parásito no se han estudiado y en su mayoría resultan desconocidas.

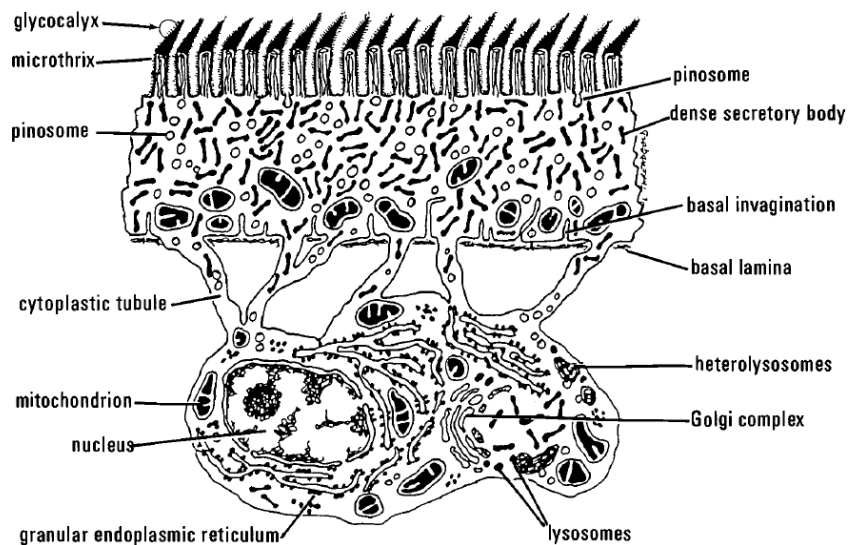


Figura 3. El tegumento de los cestodos. Se representa en la figura un único citón tegumentario. Tomado de Smyth y McManus (1).

A pesar de su diversidad anatómica en el estadio metacéstode, los cestodos adultos comparten algunas características. Estas se describen en el atlas de arquitectura tisular de *H. diminuta* por Rozario y Newmark (22): Poseen un sistema muscular a nivel de rostelo y ventosas. Adicionalmente, fibras musculares longitudinales que se extienden a lo largo del parásito y fibras transversales.

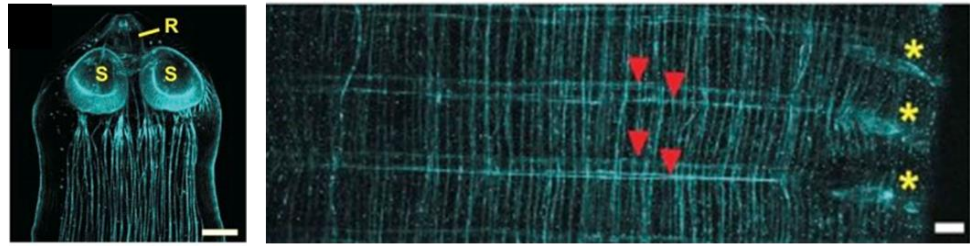


Figura 4. Fibras musculares longitudinales incluyendo ventosas (S) y rostelo (R) (izquierda) y fibras musculares transversales (flechas rojas, derecha). Tomado de Rozario y Newmark (21).

Los ganglios cefálicos (cerebro) se ubican a la altura de las ventosas que extienden dos fibras nerviosas laterales y dos fibras medias. En el caso de *H. diminuta* tres fibras nerviosas transversales se extienden en cada proglótido. Asimismo, prolongaciones dendríticas se proyectan hacia el tegumento formando un bulbo sensorial.

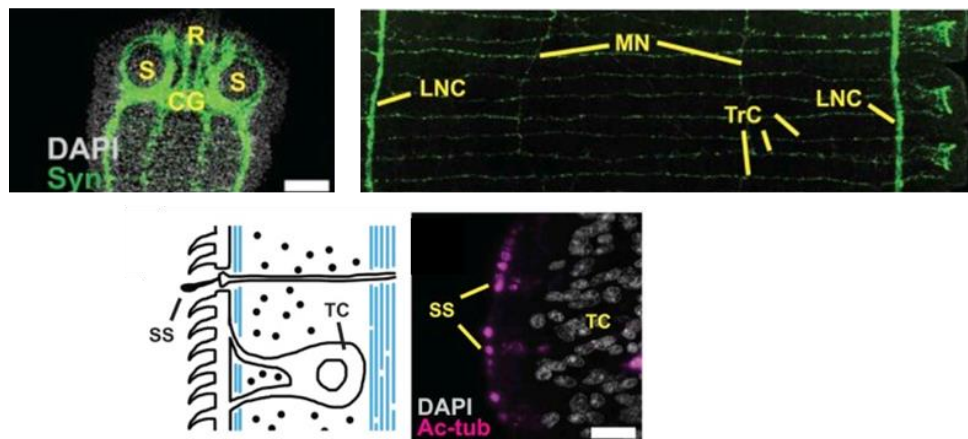


Figura 5. Estructura del sistema nervioso. Arriba izquierda: Ganglio cefálico (CG) en el escólex. Arriba derecha: Nervios medianos (MN), cordón nervioso lateral (LNC) y fibras nerviosas transversales (TrC). Representación del bulbo sensorial (SS). Tomado de Rozario y Newmark (22).

Poseen también un sistema osmorregulatorio y excretor formado de canales excretores, uno dorsal y otro ventral. Células flamígeras se conectan a estos canales excretores.

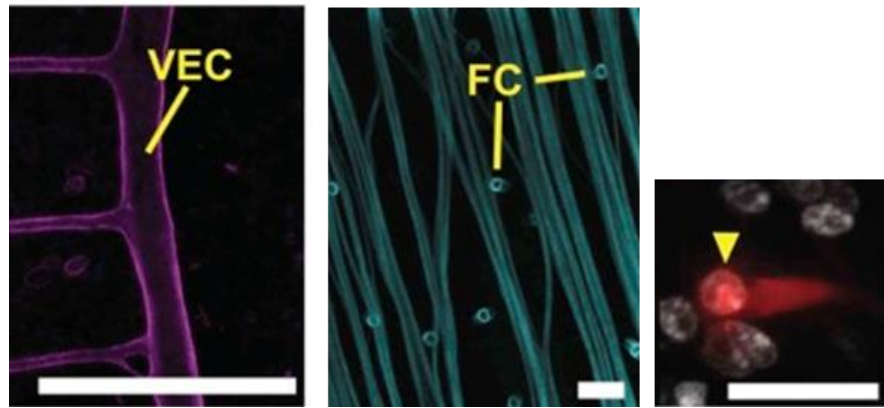


Figura 6. Sistema osmoregulatorio y excretor. A la izquierda: Canal excretor ventral. En el centro: Células flamígeras. Derecha: Célula flamígera a mayor aumento. Tomado de Rozario y Newmark (21).

Finalmente, al ser organismos hermafroditas, poseen aparatos sexuales masculinos (testis, cirro) y femeninos (atrio genital, vagina, receptáculo seminal, ovario) en cada proglótide. La fertilización se da en el oviducto. Posteriormente, el cigote se junta a una célula vitelina, la cual generará las membranas protectoras del embrión. Los embriones en desarrollo se depositan en el útero.

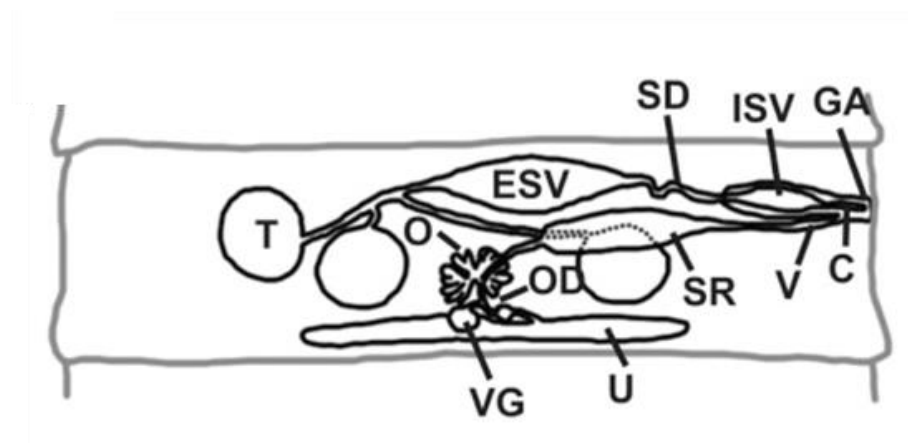


Figura 7. Aparato reproductor en una proglótide. T: Testis, C: Cirro, GA: Atrio genital, V: Vagina, SR: Receptáculo seminal, O: Ovario. OD: Oviducto, U: Útero. Tomado de Rozario y Newmark (21).

1.2. Biología del desarrollo de céstodos (*Cyclophillidea*)

Los céstodos tienen un ciclo de vida complejo. Los estadios definidos en estos organismos son huevo, oncosfera, juvenil metacéstode y adulto. El desarrollo de cada uno se da en condiciones diferentes e incluso distintos hospederos. Asimismo, las especies de céstodos pueden ser muy diversas entre sí respecto a las características de cada estadio de desarrollo. Por ello, la presente sección se centrará en los *Cyclophillidea*, que agrupa los céstodos con ciclos de vida que involucran animales terrestres (1).

1.2.1. La formación del escólex

En céstodos modelo, como los del género *Echinococcus* o *Hymenolepis* se ha estudiado ampliamente el desarrollo hacia el estadio juvenil metacéstode.

En el caso de los primeros, el quiste hidatídico se caracteriza por presentar un tejido con capacidad proliferativa donde se desarrollan constantemente protoescólices. Esto lo hace un modelo ideal para el estudio del desarrollo del escólex en céstodos. Se ha comprobado la existencia de regiones de alta proliferación celular donde se forman estas estructuras (23). En protoescólices formados, las células proliferativas se mantienen en la región posterior a las ventosas y se caracterizan por la expresión de *h2b* y *nanos* (24). Algunos otros marcadores para tejidos diferenciados que se han definido son: las fosfatasa alcalinas *alp-1/2* propias del tegumento de la capa

germinativa del quiste hidatídico y *alp-3* propio del sistema excretor del protoescólex; tubulina acetilada como marcador del sistema nervioso y tropomiosina de alto peso molecular como un marcador muscular (25).

El desarrollo de la larva metacéstode requiere el establecimiento de ejes de desarrollo. La vía de señalización Wnt define el eje anteroposterior (AP). Particularmente, los ligandos Wnt posteriores se expresan de manera ubicua y los inhibidores de Wnt especifican el polo anterior. En el caso de *E. multilocularis* y *H. microstoma*, Wnt11a se restringe progresivamente hacia el polo posterior durante la formación del protoescólex. *wnt1* se expresa hacia la región posterior al liberarse y activarse el protoescólex. Los inhibidores *sfrp* y *sfl* se restringen hacia el polo anterior en el escólex (25). Otra vía de señalización importante para la formación de protoescólices en *E. multilocularis* es la vía TGF- β /BMP (26).

1.2.2. La evaginación del escólex y estrobilación en céstodos (*Cyclophillidea*)

El juvenil metacéstode se puede encontrar libre, enquistado (la vesícula que cubre al juvenil es producto del propio parásito) o encapsulado (aislado por el organismo del hospedero) en el tejido del hospedero intermediario. El parásito por ende debe liberarse de la membrana que lo aísla (exquistar), evaginar su escólex y fijarse en su hospedero definitivo. El proceso de exquistación y evaginación dependen de

enzimas proteolíticas (para la degradación de la membrana que enquistaba al juvenil principalmente) y/o sales biliares (1,27). En *T. solium* el cultivo con bilis (12), progesterona (28) y ácido taurocólico (20) inducen la evaginación del escólex. Los mecanismos moleculares por los cuales se induce la evaginación no están del todo explorados.

La estrobilación es el paso subsecuente a la activación y fijación, e involucra el crecimiento y desarrollo de proglótides. En *H. microstoma*, Wnt1 y Wnt11 se localizan en la región posterior. Estos marcadores se posicionan en la región posterior de cada segmento. SFL se expresa en la región anterior de cada segmento. Wnt11a se expresa en la región de transición entre el cuello y las proglótides inmaduras (29).

Un análisis en 10 platelmintos (5 céstodos estrobilados y 5 tremátodos) identificó 34 secuencias asociadas a estrobilación al comparar datos entre parásitos adultos (estrobilados) y metacéstodes. Entre ellas, 6 pertenecen a vías de señalización conocidas (Wnt, TGF- β /BMP y un receptor acoplado a proteínas G). Por otro lado, 22 secuencias codifican proteínas no anotadas. Un WGCNA permitió identificar vías asociadas a algunas de estas proteínas: Vía Wnt, neurogénesis, ensamblaje del complejo de dineínas del axonema, regulación positiva de la vía JNK, procesos basados en microtúbulos, respuesta a estrés oxidativo, movimiento en base a microtúbulos, transporte intracelular, vías asociadas a receptores acoplados a proteína G, regulación positiva de

la vía Notch, organización lisosomal, adhesión celular, división celular y regulación positiva de TORC1 (30).

1.3. Metabolismo de los céstodos:

1.3.1. Metabolismo de glucosa de los céstodos

Los céstodos adultos dependen principalmente de la glucosa como sustrato metabólico (1,31). Consecuentemente, se valen de la vía glicolítica como eje central para obtener energía; también se debe considerar la pérdida a nivel genómico de múltiples rutas metabólicas (19). El primer punto de regulación metabólica en este caso se asocia al transporte de glucosa. En *Echinococcus granulosus*, se ha descrito un ortólogo de los transportadores GLUT, EgGLUT1, cuya inhibición resulta en una disminución en la viabilidad de protoescolices (juvenil) de este parásito (32). Por otra parte, en *T. solium* se caracterizaron transportadores de glucosa TGTP1 y TGTP2, análogos a los de la familia de transportadores GLUT, con actividad de transporte facilitado y sensibles a floretina (33). Adicionalmente, un reporte en el exquistamiento de *Hymenolepis diminuta*, denota mayor sensibilidad a florizina (un inhibidor del transporte de glucosa dependiente de sodio) en cisticercoides exquistados respecto a cisticercoides intactos (34). Sin embargo, la identidad de otros transportadores asociados a carbohidratos y cambios en su expresión se desconocen (35).

A pesar de que los céstodos poseen las enzimas del ciclo de ácidos tricarboxílicos y los componentes de la cadena de transporte de

electrones, rutas hacia las cuales desemboca la vía glicolítica, su metabolismo se caracteriza por ser mayoritariamente anaeróbico tanto en presencia o ausencia de oxígeno (1,36). Por ejemplo, *Echinococcus granulosus*, *Hymenolepis diminuta* e *Hymenolepis microstoma* excretan lactato tanto en presencia como ausencia de oxígeno (1,13). Adicionalmente, la capa proliferativa de *E. granulosus*, la cual contiene células germinativas (el linaje celular totipotente en céstodos) presenta una sobreexpresión de mensajeros de la enzima lactato deshidrogenasa (37). En el caso de *H. microstoma*, se observa mayor expresión de una isoforma de lactato deshidrogenasa en el estadio adulto respecto al estadio de larva cisticercoide (análogo al estadio de larva metacéstode de *T. solium*) y huevo (16). La relación entre fermentación láctica y proliferación celular en condiciones aeróbicas se conoce en algunos modelos de biología del desarrollo. Esta vía permite mantener disponibles intermediarios metabólicos, cuyos esqueletos hidrocarbonados son utilizados para la construcción de biomoléculas requeridas para la proliferación celular (38). En los últimos años se ha evidenciado el rol de la fermentación láctica en tejidos con alta proliferación celular y durante el desarrollo embrionario de organismos modelo como *Drosophila melanogaster* (39) o *Xenopus laevis* (40). Los cambios en la actividad metabólica se ven influenciados por la abundancia de enzimas en un estadio dado. En *T. solium*, no se sabe de cambios a nivel transcripcional en las enzimas clave del metabolismo glicolítico, pero estos cambios estarían potencialmente asociados a proliferación celular durante el proceso de evaginación.

Los céstodos adultos tienen fuerte preferencia por la vía de dismutación del malato (1,16,19,31,36). A esta se acopla una cadena de transporte de electrones alternativa que tiene como aceptor final de electrones al fumarato en lugar del oxígeno (1,16,19). Esto se refleja en que, además del lactato, los productos de excreción mayoritario de muchos céstodos (entre ellos los ya mencionados *H. diminuta*, *H. microstoma* y *E. granulosus*) son el acetato y succinato (1,36). La enzima limitante de esta ruta es la fosfoenolpiruvato carboxiquinasa (PEPCK). En muchos organismos esta enzima permite producir fosfoenolpiruvato a partir de oxalacetato durante la gluconeogénesis. En el caso de los céstodos, esta enzima opera hacia la producción de oxalacetato, el cual alimenta una malato deshidrogenasa citosólica que produce malato (1). Este es importado por la mitocondria, donde puede oxidarse y reducirse. La ruta oxidativa implica la reducción de NADP⁺ en NADPH y la oxidación del malato a piruvato por la enzima málica mitocondrial. Este último, mediante el complejo piruvato deshidrogenasa y la acetato:succinato coenzima-A transferasa, se convierte en acetato con la consecuente producción de succinil-CoA. La ruta reductora, por otro lado, se vale de la fumarasa para producir fumarato, el cual mediante el complejo fumarato reductasa, acepta electrones del NADH producido en la ruta oxidativa y se reduce a succinato (1,16,19). En el desarrollo adulto de *H. diminuta*, estos productos metabólicos se producen en tejidos diferenciados: lactato predominantemente en la región anterior (escólex y cuello), y succinato y acetato en la región posterior más

diferenciada (41). En *T. solium* se desconocen los transportadores asociados a la excreción de estos productos finales del metabolismo.

1.3.2. Metabolismo de ácidos grasos y colesterol en céstodos

Los céstodos han perdido la capacidad de sintetizar ácidos grasos (1,19) y colesterol *de novo* (42), además de algunos aminoácidos, como por ejemplo serina y prolina (19). Por este motivo, los céstodos se valen de dos tipos de proteínas para absorber lípidos provenientes de la dieta de su hospedero: Proteínas de unión a ligando hidrofóbico (HLBPs) y proteínas de unión a ácidos grasos (FABP) (43). En cuanto a las HLBPs, en *T. solium* se ha reportado la capacidad de unión a lípidos por parte de un complejo proteico de 150 kDa con actividad de unión a ligandos hidrofóbicos localizado en la región sincicial del tegumento (44).

Por otra parte, respecto a las FABPs, se ha reportado una proteína de unión a ácidos grasos, TsFABP1, con preferencia por los ácidos palmítico y esteárico, y ubicada en el tegumento del canal espiral y en los citones subtegumentarios (45). Adicionalmente, se ha caracterizado un ortólogo de la proteína SEC14 con capacidad de unión a lípidos, específicamente fosfatidilinositoles (46). En un reporte reciente, se evidenció expresión de un posible translocador lipídico (CD36) en una forma hiperproliferativa del cisticerco de *T. solium* conocido como cisticerco racemoso (47). Se desconoce el rol de este complejo en el desarrollo hacia el estadio adulto del parásito.

De la misma forma que en el caso de ácidos grasos, los céstodos requieren proteínas de transporte para la captación de esteroides. En céstodos, particularmente en *E. multilocularis*, se propone a la proteína 1 de tipo Niemann-Pick C1 (NPC1B), homólogo a la proteína transportadora de colesterol del intestino delgado NPC1L1, como un posible transportador de colesterol (48). Sin embargo, al ser un estudio *in silico*, se requiere una validación funcional. Se desconocen otras proteínas con potencial de ser captadoras de colesterol.

1.3.3. El metabolismo de las sales biliares en céstodos

Las sales biliares y enzimas digestivas son capaces de inducir la evaginación o exquistación del escólex en múltiples especies de céstodos, incluyendo *T. solium* (1). En adultos de *H. diminuta* y *H. microstoma* se ha evidenciado que sales biliares como el ácido taurocólico y glicocólico son adsorbidos por el tegumento de estos organismos (49). Estas observaciones sugieren que estas moléculas no cumplen un rol a nivel intracelular en céstodos. Sin embargo, esta información contradice evidencia encontrada en *E. granulosus* a nivel genético, en donde se identificaron candidatos de receptores nucleares para sales biliares con homología al receptor X farnesoide (FXR) y cotransportadores de sodio-taurocolato (50).

Adicionalmente, se conoce que el ácido taurocólico inhibe el transporte de glucosa en estas mismas especies (51). Se desconoce de rutas metabólicas involucradas en la oxidación y transporte de sales biliares.

1.4. Aspectos clave de la biología molecular de los céstodos

Los céstodos conservan la maquinaria molecular esencial de los organismos metazoos para los eventos de replicación, transcripción y traducción. Sin embargo, poseen características únicas respecto a la conservación de funciones y regulación de la expresión de genes que son un reflejo de la reducción de sistemas de órganos, necesidades nutricionales y su capacidad de adaptación a una forma de vida parasitaria.

Los céstodos poseen un ciclo de vida que involucra dos transiciones: la transición de oncosfera a larva metacéstode, y la transición de larva metacéstode a adulto (1). Los programas transcripcionales son altamente variables entre cada uno de estos estadios. Por ejemplo, se ha reportado cambios en los programas de transcripción de genes del metabolismo en *H. microstoma*, caracterizado por expresión diferencial de genes que codifican HLBP, SOAPS (proteínas ácidas secretadas por oncósferas), ASAPs (Proteínas ancladas de exón único de exón único) y expresión de isoformas únicas de lactato deshidrogenasa, así como también del conjunto de genes de la ruta de dismutación del malato expresado mayormente en el adulto respecto a la oncosfera y larva cisticercoide (16). En el caso de *M. corti*, se reportó 1811 genes diferencialmente expresados al comparar el estadio metacéstode y adulto estrobilado (14). Adicionalmente, estos cambios pueden estar confinados a periodos más discretos del desarrollo de los céstodos. Por ejemplo, protoscolices de *E. granulosus* activados durante 15

minutos con pepsina exhiben patrones diferenciales de corte y empalme alternativo (52).

Por ende, la regulación transcripcional resulta importante en el ciclo de vida de los céstodos. A continuación, se detallan algunos mecanismos de regulación transcripcional que son particulares para estos organismos.

1.4.1. Metilación del ADN

La epigenética involucra el estudio de los cambios de expresión de genes en base a mecanismos que no alteran la secuencia de estos. Estas implican modificaciones enzimáticas en bases del DNA o en los componentes estructurales de la cromatina.

En platelmintos, estudios del año 1998 (53) y 2001 (54) exponen la ausencia de modificaciones a nivel del DNA, concretamente, 5-metilcitosina (5-mC), una marca epigenética asociada a represión de la expresión de genes. Sin embargo, el desarrollo de nuevas tecnologías permitió determinar que el phylum Platyhelminthes posee esta modificación epigenética característica y las enzimas necesarias para esta. Se encontró a lo largo del phylum secuencias candidatas que poseen los motivos característicos de los dominios catalíticos de las enzimas DNA metil transferasa (DNMTs), así como también dominios de unión a metilcitosina (MBDs). Adicionalmente, se determinó la presencia de 5-mC en platelmintos representativos (*E. multilocularis* para los céstodos, *Protopolystoma xenopodis* en el caso de los

monogéneos, *Polycelis nigra* en el caso de los turbelarios y *Fasciola hepatica* en el caso de los trematodos) (55). Además de las secuencias identificadas en este estudio, no existen estudios funcionales que caractericen estas proteínas en *T. solium*.

1.4.2. Regulación por RNAs pequeños no codificantes

Los RNAs pequeños no codificantes son moléculas que regulan la expresión de genes a nivel postranscripcional.

Los mecanismos que regulan su biogénesis, maduración y actividad se encuentran conservados en los metazoos. Brevemente, los precursores de estos RNAs son procesados inicialmente por la ribonucleasa III DROSHA o el spliceosoma. Posteriormente, son modificados en su estructura de horquilla por la ribonucleasa III DICER a secuencias de entre 22 y 23 pares de bases, y secuestrados por las proteínas Argonauta. Estas últimas permiten la unión de estos RNAs pequeños con sus RNA blanco en conjunto a diversas proteínas, como por ejemplo deasilasas o el complejo RISC. El resultado es un silenciamiento postranscripcional (56).

En el caso de los platelmintos, las enzimas ribonucleasas III DICER (*dcr-1* y *dcr-2*; en trematodos adicionalmente, *dcr-3*) y DROSHA (*drsh-1*) se encuentran conservadas (57). Adicionalmente, las proteínas Argonauta también presentan al menos un ortólogo de la secuencia Ago canónica en especies parasitarias. Sin embargo, una de las proteínas

pertenecientes a esta familia, PIWI, no se encuentra presente en platelmintos parásitos, al igual que una proteína con interacción directa a PIWI, la metilasa HEN-1 (57). La vía PIWI es responsable de la regulación de la expresión y movilización de elementos transponibles. Los céstodos carecen de los elementos de la vía PIWI/piRNA, lo que implica que no regulan transposones mediante piRNAs ni utilizan PIWI para controlar la inestabilidad genómica en células madre germinales. Se puede especular que hayan desarrollado mecanismos alternativos para mantener la estabilidad genómica y controlar la expresión génica (56).

Se desconoce por tanto cómo los platelmintos parásitos, en particular los céstodos, mantienen la estabilidad genética por represión de la actividad de elementos transponibles, siendo estos últimos una causa de diversidad genética o regulación de la expresión de genes.

1.4.3. *Trans-splicing*

El corte y empalme (o *splicing* o *cis-splicing*) de mRNA es un mecanismo conservado a lo largo de los linajes eucariotas. Sin embargo, una forma particular de modificación postranscripcional es el *trans-splicing*. Este fue descrito inicialmente en Kinetoplastidos y posteriormente en el nemátodo modelo *Caenorhabditis elegans* como un mecanismo para introducir sitios de inicio de la traducción y que eventualmente permiten la transcripción policistrónica. Brevemente, un transcrito (RNA pequeño nuclear no poliadenilado) con una secuencia

de corte y empalme líder (SL), el cual puede contener un codón de inicio de la traducción, cataliza la reacción de transesterificación de un segundo transcrito. Este puede darse al inicio del primer exón en el extremo 5' o en alguno de los exones internos del transcrito (58). Por ende, la mayoría de las secuencias que sufren *trans-splicing* pueden identificarse evaluando una única secuencia compartida en el extremo 5' de diferentes transcritos.

Este mecanismo ha sido confirmado en platelmintos. Inicialmente se evidenció en el trematodo *Schistosoma mansoni* una SL de 36pb derivado de una secuencia de 90pb con un sitio de unión Sm (característico de snRNAs involucrados en mecanismos de *cis-splicing*) que realiza *trans-splicing* en el exón 3 de la 3-hidroxi-3-metilglutaril-CoA reductasa. Se confirmó adicionalmente que esta secuencia puede sufrir *cis-splicing* con su exón 2 (59). Resultados similares se obtuvieron al analizar la enolasa de *Fasciola hepatica* donde se identificó un SL de 37pb proveniente de una secuencia de 108pb ubicada en una región con repeticiones en tándem y es estructuralmente distinta a la encontrada previamente en *S. mansoni* (60). En céstodos particularmente, se identificó en *E. multilocularis* un SL-RNA de 104pb que incluye un SL de 36pb (61).

Las secuencias SL en platelmintos rondan aproximadamente entre los 35 y 51pb. Sin embargo, su tamaño ni secuencia están conservadas a lo largo de los platelmintos. Incluso, los elementos comunes a estas

secuencias como las horquillas en los extremos 5'y 3' de las secuencias SL y los dominios Sm resultan altamente variables en platelmintos a diferencia de Nemátodos y Kinetoplástidos (62).

2. Transcriptómica de céstodos

2.1. El genoma de *Taenia solium*:

Hasta el momento existen solamente dos genomas disponibles de *T. solium*. El primero, obtenido por el *Taenia solium genome consortium* y disponible en la base de datos WormBase ParaSite, consta de 12352 contigs y 11237 scaffolds, un N50 de 68 kilobases compuesto por 439 scaffolds, y 12467 genes codantes (19). El segundo, obtenido en Perú, de 7979 contigs y un N50 de 43.9 kilobases, sin una anotación de genes (63). Ambos proyectos fueron realizados en los años 2013 y 2015, respectivamente, mediante tecnologías de secuenciación de segunda generación.

La ausencia de genomas representativos es una limitante para el estudio de la biología del parásito. Asimismo, la identidad y función de las secuencias reportadas en bases de datos ómicas de libre acceso se basa en la búsqueda de secuencias ortólogas con funciones ya asignadas en organismos más estudiados. La escasez de estudios de secuenciación de genomas y transcriptomas de *T. solium* resulta en una limitada anotación funcional de genes para este organismo. Por otro lado, para elucidar la secuencia de isoformas a lo largo de los genes anotados, el genoma de referencia (PRJNA170813) carece de anotaciones de isoformas por corte y empalme alternativo. En comparación, *E. granulosus* posee al menos 33.31% y 38.1%

de genes totales con eventos de corte y empalme alternativo para sus dos genomas de referencia (PRJNA182977 y PRJEB121 respectivamente). *E. multilocularis* posee al menos 35.75% de genes totales con eventos de corte y empalme alternativo (PRJEB122) (52).

Las tecnologías de secuenciación de tercera generación han permitido obtener genomas con resolución a nivel de cromosoma, como es el caso de *H. microstoma*, con 7 scaffolds representando cada uno un cromosoma del parásito y el genoma mitocondrial. Este adicionalmente integra los genomas secuenciados en 2013 (1132 scaffolds) (19) y 2015 (3643 scaffolds) para dicho parásito (64).

2.2. Transcriptomas de *T. solium* y céstodos modelo en el contexto de la evaginación/exquistación:

En la actualidad, se han reportado 5 transcriptomas de *T. solium*. El primer transcriptoma se obtuvo mediante Open Reading Frame Expressed Sequence Tags (ORESTES) al secuenciar librerías de cDNA y resultó en la anotación de 191 secuencias (65). El segundo corresponde a secuenciación por capilares de una biblioteca de cDNA reportada junto al genoma de referencia con 12,490 genes anotados (19). El tercer transcriptoma fue obtenido mediante secuenciación de segunda generación Illumina del estadio cisticerco con el objetivo de observar efectos en el metiloma del parásito (66). El cuarto, corresponde a la comparación entre vesículas de un cisticerco racemoso y vesículas o escólex de un cisticerco típico (47). Finalmente, (y como parte del presente trabajo) se reportó el transcriptoma

de cisticercos antes y después de evaginar el escólex por efecto del ácido taurocólico, una sal biliar (17). Este último es el único enfocado en describir un evento del ciclo de vida y desarrollo de *T. solium*.

Existen varios transcriptomas de céstodos fáciles de manipular en el laboratorio. En el contexto del desarrollo, por ejemplo, el transcriptoma de *Mesocostoides corti* presenta transcritos con expresión diferencial en el estadio adulto respecto al tetratiridio que se asocian a procesos biológicos como glicólisis, transcripción, señalización, recambio de proteínas y proliferación celular (14), los cuales estarían relacionados con las funciones fisiológicas del estadio adulto de los céstodos.

En *E. granulosus*, se identificaron módulos de genes mediante WGCNA al comparar oncosfera, protoescólex y adulto. Los módulos asociados a procesos de citoesqueleto y motilidad celular poseen alta expresión en protoescólices, con respecto a adultos. Lo opuesto se observa en el módulo asociado a metabolismo de carbohidratos y vitaminas (67).

En la misma especie, otro estudio determinó splicing diferencial en protoescólices inducidos o no inducidos con pepsina. Se determinó que los transcritos con múltiples isoformas corresponden mayormente a categorías funcionales de biogénesis ribosomal. Sin embargo, en presencia de pepsina, los transcritos con mayor splicing diferencial son componentes del spliceosoma (proteínas SR, y factores asociados a U1 y U2) (52).

Otro estudio, en presencia o ausencia de bilis en *E. granulosus*, determinó que este inductor permite la estrobilación del parásito. Los procesos inducidos se asocian al metabolismo de carbohidratos y desarrollo del

sistema nervioso. En ausencia de bilis se favorecen procesos relacionados con transporte de iones y actividad dependiente de nucleósido-trifosfato. Procesos asociados al citoesqueleto (procesos asociados a microtúbulos), homeostasis iónica y celular se ven enriquecidos en individuos inducidos y no inducidos. Sin embargo, otros genes son diferencialmente expresados de manera independiente de la bilis (68).

En *Taenia pisiformis*, en presencia de bilis se encontró sobreexpresión de términos GO “Proceso metabólico de péptidos”, “Proceso biosintético de amidas”, “Traducción”; en cuanto a KEGG: “Vía de señalización de la quinasa activada por AMP”, “Vía TGF- β ”, “Elongación de ácidos grasos”, “Biosíntesis de pantotenato y CoA”, “Fosforilación oxidativa” (estas 3 últimas considerando que los céstodos no sintetizan ácidos grasos *de novo* (19)), “Ribosoma”, “Proteosoma”, “Spliceosoma” y “RNA polimerasa”. Por otro lado, se encontraron reprimidos los términos GO: “Procesos de oxidación-reducción” y “Metabolismo de carbohidratos”; respecto a KEGG: “Amino azúcar y metabolismo de nucleótido-azúcar”, “Degradación de otros glicanos” y “metabolismo de inositol-fosfato” (69).

3. Herramientas para el análisis transcriptómico en *T. solium*

3.1. La base de datos WormBase ParaSite

La base de datos WormBase ParaSite es el repositorio principal de información ómica de nemátodos, platelmintos parásitos y de vida libre. En total, la versión WBPS19 incluye 274 genomas de 208 especies distintas (70). Esta base de datos también integra información de bases de datos

externas como por ejemplo InterPro (71). En el caso de *T. solium*, existe un genoma de referencia en esta plataforma con la anotación de 12467 genes (19).

Sin embargo, como se expuso anteriormente, la calidad de este genoma impide una anotación de genes fiable. Herramientas como BioMart, disponibles en esta base de datos, facilitan la búsqueda de secuencias ortólogas y la asignación de categorías funcionales en base a homología. Adicionalmente, esta plataforma cuenta con la posibilidad de realizar BLAST sobre los genomas, transcriptomas y proteomas disponibles en la totalidad de base de datos.

3.1.1. Anotación funcional de genes: Blast2GO

Tomando en consideración la calidad de los genomas disponibles, es posible realizar una nueva anotación funcional de genes. Blast2GO es una herramienta que permite la anotación de secuencias en gran escala, lo cual permite obtener términos GO, KEGG y códigos de enzimas para todas las secuencias de un transcriptoma. Para ello, realiza alineamientos de todas las secuencias a bases de datos (por ejemplo: NCBI nr, RefSeq) mediante BLAST (72). Adicionalmente, integra InterProScan, lo cual permite búsquedas de dominios estructurales y funcionales en bases curadas como Pfam, SMART, PROSITE, etc (73). Finalmente, incluye también la herramienta eggNOG-mapper, la cual realiza alineamientos de secuencias contra una base de datos de grupos de ortólogos construida a partir de predicciones funcionales basadas en

secuencias ortólogas de más de 5000 organismos. Esta herramienta es especialmente útil para organismos no-modelo como el caso de *T. solium* (74).

3.2. Herramientas de exploración no-supervisada de datos de RNAseq

3.2.1. Análisis de componentes principales (PCA)

Un experimento de RNAseq de “n” muestras y “p” genes puede representarse como un punto en un espacio dimensional donde cada punto es una muestra y cada gen es una dimensión. El análisis de componentes principales establece dos variables arbitrarias que capturan la máxima varianza, no están correlacionadas (ortogonales) y permiten proyectar la totalidad de datos en 2 o 3 dimensiones. Es útil para detectar patrones, outliers e identificar otras fuentes de variación (75,76).

Se parte de una matriz X de conteos de “p” genes en “n” condiciones normalizados y centrados a la media de cada gen. El cálculo de una matriz de covarianza permite capturar la varianza de la expresión de genes a lo largo de las muestras.

$$COV = \frac{1}{n-1} X^T X$$

La descomposición propia de la matriz de covarianzas permite encontrar los vectores propios v_k (componentes principales). Estos valores capturan una proporción de la varianza total. Para ello, se calcula los valores propios λ resolviendo:

$$\det(COV - \lambda I) = 0$$

Los componentes principales (v_k) se definen como la combinación lineal de genes que capturan un porcentaje de la varianza de los datos.

Para ello se resuelve:

$$COV v_k = \lambda_k v_k$$

Finalmente, se proyectan los datos sobre 2 o 3 de los componentes principales con mayor varianza explicada (mayor valor propio para dicho vector) calculando el producto escalar de los conteos normalizados y centrados para cada gen con el componente principal v_1, v_2 y v_3 .

$$Z = X \cdot v_k$$

3.2.2. Agrupamiento jerarquizado

Este es un algoritmo basado en distancias. Requiere el cálculo de una matriz de distancias en el caso de una matriz X de conteos normalizados de genes de “n” muestras y “p” genes:

$$D_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Donde “i” y “j” representan dos muestras distintas y “k” un gen. D es una matriz de distancias euclidianas. Posteriormente se utiliza agrupamiento por enlazamiento completo. Es decir, se calcula la máxima distancia entre grupos (partiendo por grupos compuestos por una única muestra) y fusiona el par con la menor distancia máxima.

Posteriormente se calcula nuevamente la matriz de distancias con este nuevo grupo y se repite el procedimiento hasta finalizar con un único grupo con múltiples subgrupos (76,77).

3.3. DESeq2

DESeq2 es un algoritmo para determinar diferencias de conteos entre grupos (78). Particularmente, el diseño de experimentos de secuenciación de RNA involucran un número bajo de réplicas y los datos de conteo obtenidos representan variables cuantitativas discretas con un amplio rango dinámico, que no siguen los supuestos de normalidad que requieren las pruebas estadísticas convencionales (76,78).

En un contexto de variables cuantitativas continuas con distribución normal, se puede definir un modelo lineal ordinario de la siguiente forma:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Donde “Y” es la variable respuesta, “X” la variable predictora, “ β_0 ” es el valor de “Y” cuando “X” es 0, “ β_1 ” es el cambio en el valor de “Y” cuando “X” incrementa en una unidad, y “ ϵ ” el error estándar o la diferencia entre los valores muestreados y el valor obtenido por el modelo. Adicionalmente, el modelo involucra asumir varianza constante para todos los valores tomados por la variable independiente (homocedasticidad) y una distribución normal del error estándar (en el caso del modelo, los residuales). El modelo se puede generalizar de la siguiente manera:

$$Y|X \sim N(\mu, \sigma^2)$$

Donde la variable respuesta sigue una distribución normal (N) condicionada al valor de X y posee una media “ μ ”, correspondiente al valor de Y estimado por el modelo para cada X, con varianza σ^2 dada por el error estándar. La media se puede definir de la siguiente forma:

$$\mu = \eta$$

En este caso, “ η ” es el predictor lineal, que en el caso de un modelo lineal ordinario se define como:

$$\eta = \beta_0 + \beta_1 X + \epsilon$$

A esto se le conoce como un modelo lineal generalizado (GLM) y permite utilizar distribuciones de probabilidad que se ajusten a la naturaleza de los datos al aplicar una función de enlace a la media o parámetro a estimar (en el caso de un modelo lineal ordinario, multiplicación del predictor lineal por 1) (79).

En el caso de los datos de conteo de RNAseq, por las características previamente establecidas, estos requieren un tratamiento estadístico específico donde la variable respuesta es el valor de conteo para cada gen y la variable predictora es el tratamiento/condición. DESeq2 asume que los valores de conteo poseen una distribución binomial negativa (NB), la cual es útil para datos de conteo donde la varianza excede el valor de la media (76,78):

$$Y|X \sim NB(\mu, \alpha)$$

Donde la varianza, con dispersión “ α ”, se define como:

$$Var(y) = \mu + \alpha\mu^2$$

En este caso, a la media se le aplica la función de enlace “log”. Por ello, se define como:

$$\log(\mu) = \eta$$

Con un modelo predictor lineal:

$$\eta = X\beta$$

Este modelo es aplicado por DESeq2 para los conteos (K) del gen “i” en la muestra “j”:

$$K_{ij} \sim NB(\mu_{ij}, \alpha_i)$$

Donde μ es la media esperada de conteos del gen “i” en la muestra “j” y α es la dispersión específica al gen “i”. Como se demostró con anterioridad al definir el GLM, la varianza de los conteos “K” del gen “i” en la muestra “j” dependen de la media esperada y de la dispersión.

Debido al bajo número de réplicas en los experimentos de secuenciación, la estimación de la dispersión puede resultar altamente variable. DESeq2 asume que los genes con una expresión promedio similar tienen una dispersión similar. La dispersión “ α ” de cada gen se estima por separado

mediante el método de máxima verosimilitud. Se ajusta una curva de estos valores estimados. De acuerdo con esta, los valores estimados son comprimidos de modo que se ajusten mejor a esta curva mediante método empírico de Bayes, lo cual implica que la compresión de los valores estimados dependerá del ajuste estimado *a priori* y el número de muestras. Mientras menos muestras se incluyen, el ajuste es mayor y viceversa. Esto permite que, ante la ausencia de información, se estime la dispersión de cada gen en base a la tendencia global de la dispersión manteniendo la variabilidad específica de cada gen.

La media esperada de conteos del gen “i” en la muestra “j” se puede establecer de la siguiente forma:

$$\mu_{ij} = q_{ij} \cdot s_j$$

En este caso “s” representa el factor de tamaño que aplica una corrección a las diferencias en la profundidad de secuenciación y el tamaño de las bibliotecas entre las muestras “j” a los valores de “q”, que son proporcionales a la concentración real de lecturas para la muestra “j”.

El factor de tamaño se establece a partir de la matriz de conteos “K” mediante el método de la mediana de razones: Se computa la media geométrica “G” para cada gen “i” a lo largo de todas las muestras (n).

$$G_i = \left(\prod_{j=1}^n K_{ij} \right)^{1/n}$$

Conteos de valor 0 son omitidos de este cálculo. Para cada muestra “j” se calcula la razón de cada gen respecto a “G”.

$$R_{ij} = \frac{K_{ij}}{G_i}$$

Finalmente, el factor de tamaño se define como:

$$s_j = \text{mediana}_i(R_{ij})$$

Dado que “s” es considerada una constante, “q” es el parámetro por estimar en el modelo:

$$\log_2(q_{ij}) = \eta$$

$$\eta = X_j\beta_i$$

En este caso, “X” es un vector de diseño experimental que indica la condición de la muestra “j”, y “β” un vector de coeficientes que representa la respuesta del gen “i” en cada una de las condiciones del diseño. El producto resulta en un valor escalar que representa el logaritmo en base 2 de la expresión relativa para el gen “i” para cada muestra “j”.

Una vez ajustado el GLM, DESeq2 utiliza la prueba estadística de Wald para determinar expresión diferencial al evaluar si los coeficientes “β” en el modelo son estadísticamente diferentes de 0. Para ello, para cada coeficiente “β” del gen “i” bajo la condición “k” se calcula el estadístico:

$$W = \frac{\hat{\beta}_{ik}}{SE(\hat{\beta}_{ik})}$$

Bajo la hipótesis nula:

$$H_0: \beta_{ik} = 0$$

Con distribución nula del estadístico de Wald:

$$W \sim N(0,1)$$

Se realiza el cálculo del valor de p (dos colas) y se genera un ajuste de pruebas múltiples mediante el método de Benjamini-Hochberg.

3.4. Análisis de enriquecimiento funcional

Los análisis de enriquecimiento funcional permiten determinar las categorías funcionales que se encuentran mayormente representadas en un conjunto de genes. Las categorías funcionales son términos que agrupan genes con características funcionales similares. La base de datos de categorías funcionales más utilizada corresponde al Consorcio de Ontología Génica (*Gene Ontology Consortium*). Esta agrupa términos de ontología génica (GO) que se asocian a tres grandes grupos/ontologías: Procesos biológicos (BP), función molecular (MF) y componente celular (CC) (80,81).

El paquete de R/Bioconductor clusterProfiler incluye múltiples algoritmos de análisis de enriquecimiento funcional y reducción de redundancia de categorías funcionales (82). A continuación, se explicarán los algoritmos de mayor relevancia para el presente trabajo.

3.4.1. Análisis de sobrerrepresentación (ORA)

Los análisis de sobrerrepresentación requieren de establecer previamente un conjunto de genes de interés y especificar el conjunto universo de genes o conjunto de genes de fondo. Esto permite estimar estadísticamente el enriquecimiento de categorías funcionales de un conjunto de genes con características particulares (p. ej. Genes diferencialmente expresados bajo valores umbrales de expresión relativa y significancia estadística) (82,83).

La prueba estadística utilizada es una prueba hipergeométrica (prueba exacta de Fisher con una cola) con hipótesis nula: Los genes de interés están distribuidos aleatoriamente a lo largo de los genes de fondo respecto a una categoría funcional.

El valor de p se calcula de la siguiente forma:

$$p = \frac{\sum_{k=x}^{\min(M,K)} \binom{M}{k} \binom{N-M}{K-k}}{\binom{N}{K}}$$

Donde:

- N: número de elementos en el conjunto de genes de fondo.
- M: número de elementos en el conjunto de genes pertenecientes a una categoría funcional a evaluar.
- K: número de elementos en el conjunto de genes de interés.

- x : número de elementos en la intersección del conjunto de genes de interés y el conjunto de genes de la categoría funcional a evaluar.

El valor de p es ajustado posteriormente para pruebas múltiples mediante el método de Benjamini-Hochberg.

3.4.2. Análisis de enriquecimiento de conjuntos de genes (GSEA)

A diferencia de los ORA, GSEA permite determinar el enriquecimiento de categorías funcionales utilizando la información de expresión de genes obtenida mediante análisis de expresión diferencial. Esto permite detectar el enriquecimiento de categorías funcionales con cambios de expresión sutiles a lo largo de las condiciones incluidas en el diseño experimental (82,84).

Sea “G” el conjunto de genes “g” secuenciados, se aplica un ordenamiento decreciente en base a un parámetro “r” (en el caso datos analizados mediante DESeq2, se utiliza el estadístico de Wald, el logaritmo en base 2 de la expresión relativa o el valor de p ajustado) obteniendo el vector de genes L:

$$L = [g_{(1)}, g_{(2)}, \dots, g_{(N)}]; r(g_{(1)}) \geq r(g_{(2)}) \geq \dots \geq r(g_{(N)})$$

Se calcula una puntuación partiendo del inicio de la lista y de forma consecutiva. El valor es proporcional a la magnitud del estadístico. La curva de puntuación de enriquecimiento, “RES”, para cada gen de la categoría funcional “S” se construye de la siguiente forma:

$$RES_i = \sum_{k=1}^i \begin{cases} \frac{|r_k|^p}{\sum_{g_j \in S} |r_j|^p} & \text{si } g_k \in S \\ -\frac{1}{N - N_H} & \text{si } g_k \notin S \end{cases}$$

Si el gen de la lista está incluido en la categoría funcional “S”, la puntuación de enriquecimiento es positiva. En este caso el numerador de la expresión es el valor absoluto del estadístico “r” para dicho gen elevado a un exponente de ponderación, “p” (que toma el valor de 1 usualmente), y el denominador es un parámetro normalizador que toma el valor de la suma total de valores absolutos del estadístico “r” para la categoría funcional “S” elevado al exponente de ponderación “p”.

Si el gen de la lista no está incluido en la categoría funcional “S”, la puntuación de enriquecimiento es negativa. La expresión para su cálculo es el valor negativo de la inversa de la resta del número total de genes de la lista “N” y el número total de genes pertenecientes a la categoría funcional “S”, “N_H”.

El valor de máxima desviación de 0 en la curva “RES” corresponde a un estadístico de tipo Kolmogorov-Smirnov que es utilizado para el cálculo del valor del estadístico “p” y se conoce como puntuación de enriquecimiento, “ES”.

El algoritmo calcula una distribución nula mediante la permutación de los valores del parámetro “r” a partir del cual se establece “L”. El valor

de p es calculado de acuerdo con el valor de “ES” relativo a la distribución nula (el número de permutaciones que excede el valor ES respecto al total de permutaciones).

Posteriormente se evalúa si los genes de una categoría funcional “S” se encuentran en los extremos de “L”, o si por el contrario se encuentran distribuidos de forma aleatoria.

Para corregir el caso de pruebas múltiples, se toma en cuenta que el valor de “ES” depende del tamaño del conjunto de genes de cada categoría funcional, la cual es variable para cada una de estas. Por ello, se aplica una normalización respecto a la media de los valores de “ES” al permutar los valores del parámetro “r” cuando se construye la distribución nula:

$$NES = \frac{ES}{media(ES_{H_0})}$$

De la misma forma se construye un conjunto de valores NES nulos, “ NES_{H_0} ”. Finalmente se estima la tasa de descubrimientos falsos (FDR) para los valores NES de cada categoría funcional “S”, al dividir el número de NES nulos por encima del valor NES de “S” entre los valores NES observados por encima del valor NES de “S”:

$$FDR(NES) = \frac{\#(NES_{H_0} \geq NES)}{\#(NES_{obs} \geq NES)}$$

En el caso de NES negativos:

$$FDR(NES) = \frac{\#(NES_{H_0} \leq NES)}{\#(NES_{obs} \leq NES)}$$

3.5. Redes de Coexpresión Genética Ponderadas (WGCNA)

El análisis WGCNA permite describir la correlación de patrones de expresión de genes a lo largo de las muestras de RNAseq. Es una herramienta que permite la identificación de conjuntos de genes con expresión altamente correlacionada, denominados módulos (85,86).

De igual forma que el cálculo de una matriz de distancias para agrupamiento jerárquico, se establece una matriz “ s_{ij} ” mediante el cálculo del coeficiente de correlación de Pearson de los conteos normalizados de todos o un grupo de genes que abarque una proporción de la varianza en la expresión global:

$$s_{ij} = cor(x_i, x_j)$$

Esta matriz es transformada a una matriz de adyacencia al aplicar una función de potencia a la matriz de correlaciones:

$$a_{ij} = |s_{ij}|^\beta$$

El uso de una función de potencia permite penalizar valores bajos de correlación sin perder información, a diferencia de usar un valor umbral fijo de correlación. El valor del exponente de la función de potencia “ β ” se establece de manera empírica de modo que la red resultante se asemeje a una topología de red sin escala. Este tipo de redes se caracteriza por tener una topología establecida por pocos nodos altamente conectados con

muchos nodos de baja conectividad. Este tipo de topología es común en sistemas biológicos (87).

Se establece una matriz de superposición topológica (TOM), la cual mide la superposición de vecinos compartidos entre pares de genes. Es una medida de la interconectividad de una red de genes, donde “ l_{ij} ” es el número de vecinos compartidos entre los genes “ i ” y “ j ”, y “ k_i ” es la conectividad del gen “ i ”:

$$TOM_{ij} = \frac{l_{ij} + a_{ij}}{\min(k_i, k_j) + 1 - a_{ij}}$$

$$l_{ij} = \sum_u a_{iu} a_{uj}$$

$$k_i = \sum_u a_{iu}$$

Se construye una matriz de disimilitud de TOM, la cual es una matriz de distancias que toma valores de 0 si los genes están altamente correlacionados y de 1 si los genes se encuentran desconectados. Se aplica agrupamiento jerárquico y se definen módulos mediante el algoritmo de corte dinámico de ramas (*Dynamic tree cut*) tomando como parámetro un tamaño mínimo de genes por módulo especificado por el usuario (30 genes por defecto):

$$dissTOM_{ij} = 1 - TOM$$

Los valores propios de los genes (*eigengenes*) de cada módulo se calculan como su primer componente principal, el cual representa el patrón de expresión promedio de dicho módulo.

$$ME_k = PC_1(modulo_k)$$

La relación de variables experimentales (*trait*, T) con cada módulo se define como la correlación de Pearson entre el valor propio de los genes de cada módulo con una matriz de diseño experimental:

$$cor(ME_k, T)$$

La asociación de cada gen con cada variable experimental (significancia de un gen, GS) y la correlación del perfil de expresión de un gen “i” con el módulo (pertenencia a módulo, MM) se cuantifican de la siguiente forma:

$$GS_i = |cor(x_i, T)|$$

$$MM_i = cor(x_i, ME_k)$$

Los genes con valores altos de “GS” y “MM” se denominan genes modulares, genes de alta conectividad o genes nodo (*hub*), los cuales explican la mayor varianza en los niveles de expresión de dicho módulo asociado a una variable experimental.

3.6. La herramienta de interacción proteína-proteína STRING

STRING es el acrónimo de Herramienta de Búsqueda para la recuperación de genes/proteínas que interactúan. Es una base de datos que posee información de interacciones proteína-proteína conocidas o predichas. Para

ello toda la información disponible en esta base de datos se categoriza en siete “canales”: tres de contexto genómico (vecindad, unión o co-ocurrencia), canal de coexpresión (incluye experimentos de RNAseq en el Gene Expression Omnibus del NCBI y proteómica), canal de minado de texto, canal de datos bioquímicos/genéticos experimentales y canal de bases de datos curadas de vías y complejos externas (Reactome, Gene Ontology, KEGG y BioCyc). Esta herramienta permite establecer redes de interacción proteína-proteína y realizar análisis de enriquecimiento funcional (87).

IV. METODOLOGÍA

1. Diseño general:

Se realizó una nueva anotación funcional a partir de las secuencias identificadas en el genoma/transcriptoma de *T. solium* previamente reportado por Tsai y colaboradores (19), dado que las funciones y términos GO no se encuentran completamente anotados. A partir de los genes diferencialmente expresados, se realizó un análisis de sobrerrepresentación de categorías de ontología génica (GO). Adicionalmente, se utilizó la información de expresión de todos los transcritos para el análisis de enriquecimiento de conjuntos de genes (GSEA). Para determinar conjuntos de genes con tendencias transcripcionales similares, se construyeron redes de coexpresión genética ponderadas (WGCNA) y se aplicó un análisis de sobrerrepresentación (ORA) para determinar las categorías funcionales de los módulos de genes identificados. Los módulos de genes identificados se asociaron a condiciones experimentales y se validaron funcionalmente mediante redes de interacción/asociación proteína-proteína. Los procesos de interés son aquellos asociados a metabolismo energético, transporte de nutrientes, proliferación celular, excreción, elementos de matriz extracelular y síntesis de proteínas.

2. Mapeo al genoma de referencia y generación de archivo de conteo, y análisis de expresión diferencial

Para obtener las abundancias de los transcritos, las lecturas fueron mapeadas al genoma de referencia, reportado previamente por Tsai y colaboradores (19), mediante el paquete RSubread (88). Posteriormente, se generó un archivo de conteos de transcritos mapeados mediante el paquete bioinformático

GenomicFeatures (89). A partir de este se realizó la normalización de los conteos para cada gen y el análisis de expresión diferencial mediante el paquete DESeq2 (78). Este algoritmo permitió obtener conjuntos de genes diferencialmente expresados al comparar dos condiciones. Se consideraron genes diferencialmente expresados a aquellos que presentaron una tasa de descubrimientos falsos (FDR) menor a 0.05 y un valor de $\log_2$ *Fold Change* superior a 1 o inferior a -1.

Los grupos experimentales se validaron mediante análisis de reducción dimensional: PCA y agrupamiento jerarquizado. Esto permitió determinar las condiciones a comparar en el análisis de expresión diferencial.

3. Anotación funcional del transcriptoma de *Taenia solium*

El transcriptoma de *Taenia solium* se anotó funcionalmente mediante transferencia de función utilizando la herramienta bioinformática BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), utilizando la base de datos NR (Non Redundant) del NCBI e incluyendo un filtro taxonómico para platelmintos (Platyhelminthes, taxid:6157) y nemátodos (Nematoda, taxid:6231) con un valor de E menor a 1×10^{-5} . Estos últimos al ser protostomados con algunas especies parasitarias ampliamente estudiadas a nivel genético. Los parámetros adicionales se mantuvieron en los valores por defecto. También, se utilizó la herramienta bioinformática INTERPRO (<https://www.ebi.ac.uk/interpro/>) para la anotación de ontologías génicas de acuerdo con dominios conservados asociados a la secuencia de aminoácidos codificada por cada transcrito bajo los parámetros por defecto (valor de E menor a 1×10^{-6}). Esta utiliza múltiples

bases de datos (CDD, Pfam, Prosite, PANTHER, CATH, entre otras) para asignar en base a secuencia la identidad de dominios conservados con funciones conocidas y asignar esta secuencia a una familia o superfamilia de proteínas. Adicionalmente, se realizó una anotación mediante la herramienta EggNOG mapper (57), la cual utiliza una base de datos de grupos de ortólogos y filogenias previamente establecidas. Se utilizó los parámetros por defecto (valor de E menor a 1×10^{-6}). Se elaboró un consolidado de estas anotaciones. Los términos de ontología génica fueron asignados mediante la integración de las anotaciones de BLAST, INTERPRO y EggNOG utilizando la herramienta de mapeo BLAST2GO (58).

4. Análisis de Sobrerepresentación de ontología génica (GO)

Se utilizaron los genes obtenidos por análisis de expresión diferencial. Se utilizará la función ‘compareCluster’ del paquete de R clusterProfiler (82) y la ontología de procesos biológicos (BP). El algoritmo calcula un valor de significancia estadística mediante una distribución hipergeométrica que permite determinar si un término GO, en base a una lista de genes de interés, se encuentra representado a una frecuencia mayor a la esperada por azar. El valor de significancia elegido es de $p < 0.1$. Este análisis se realizó para todas las comparaciones validadas.

5. Análisis de Enriquecimiento en Conjuntos de Genes (GSEA)

Se ordenaron todos los genes secuenciados en base al valor del estadístico de Wald obtenido en el análisis de expresión diferencial. Se filtraron los genes con

un valor de estadístico 0. Se utilizó la función gseGO con un valor de significancia de $p < 0.05$. Este análisis se realizó para todas las comparaciones validadas.

6. Análisis de Redes de Coexpresión Genética Ponderadas (WGCNA)

Se transformó la matriz de conteos normalizados obtenida en el análisis de expresión diferencial mediante la función estabilizadora de la varianza (VST). Se estimó el valor de potencia para correlaciones genéticas ponderadas de $\beta = 9$. Se realizó el análisis con la función WGCNA del paquete con el mismo nombre. El valor de correlación mínimo considerado fue 0.8 y se establecieron módulos de al menos 30 elementos. Se correlacionaron estos módulos a las condiciones experimentales (fase de desarrollo, presencia de inductor) y se realizó un ORA como se describió previamente.

7. Redes de interacción/asociación proteína-proteína (STRING)

Se anotó funcionalmente el proteoma de *T. solium* obtenido desde la base de datos WormBase ParaSite en la base de datos STRING (87). Se evaluaron los genes identificados en los módulos establecidos previamente mediante WGCNA y adicionalmente se realizó un análisis de sobrerrepresentación de elementos identificados en la base de datos Reactome. Los parámetros usados fueron los establecidos por defecto (puntaje de interacción mínima requerida = 0.4, agrupamiento de elementos por similitud > 0.8)

8. Aspectos éticos y de bioseguridad

El proyecto sombilla, “Identificación transcriptómica de determinantes celulares en la transición *in vitro* de larva metacestode (cisticerco) a gusano adulto en *Taenia solium*” (SIDISI 209290), se ejecutó bajo aprobación del Comité Institucional de Ética para el uso de Animales (CIEA) de la UPCH, vigente entre octubre de 2022 y septiembre de 2024 (CONSTANCIA 040-09-22 y CONSTANCIA-CIEA-R-040-09-23). Este periodo abarca el sacrificio de cerdos para coleccionar los cisticercos de *T. solium* analizados en el presente trabajo de tesis. Se solicitó y obtuvo exoneración de revisión al CIEA.

El parásito *Taenia solium* es un agente patógeno perteneciente al grupo de riesgo dirigido a humanos 2 (RG2) de acuerdo con la base de datos ePATHOGEN de la agencia de salud pública de Canadá (<https://health.canada.ca/en/epathogen>).

El manejo de datos bioinformáticos no requiere la manipulación directa de cisticercos. Los análisis se realizaron en un área del laboratorio libre de contaminación biológica/química destinada exclusivamente al trabajo con computadoras con acceso a internet. Los resultados de los análisis se colocaron encriptados en un disco duro del laboratorio.

V. RESULTADOS

Los resultados detallan los hallazgos principales usando diferentes análisis. Primero, para verificar posible contaminación se alinearon las lecturas al genoma de cerdo (*Sus scrofa*). Posteriormente, se evaluaron de forma exploratoria los perfiles transcripcionales y la calidad de las réplicas de los grupos estudiados mediante agrupamiento jerarquizado y PCA. A continuación, se determinaron los genes diferencialmente expresados entre individuos con escólex no evaginado y evaginado. Se determinaron los procesos biológicos modulados mediante análisis de enriquecimiento funcional (ORA, GSEA). Luego, se determinaron conjuntos de genes con expresión correlativa y se asociaron tanto al AT como a la evaginación del escólex. Las funciones de estos conjuntos se verificaron mediante análisis de enriquecimiento funcional.

En la sección de Discusión se reúnen los resultados en una visión panorámica de los alcances del transcriptoma para describir el proceso de evaginación del escólex inducido por AT como momento crucial en el desarrollo de *T. solium*.

1. Las lecturas obtenidas por secuenciación Illumina de RNA no tuvieron contaminación con material genético del hospedero, lo cual permite realizar análisis de expresión diferencial, previa normalización

Siete condiciones experimentales fueron secuenciadas, cada una con tres réplicas biológicas compuestas por cuatro cisticercos cada una. El mapeo de lecturas mediante Rsubread dio como resultado la distribución mostrada en la figura 8A. La profundidad de la secuenciación se encuentra en el rango de 9.27 (EV en cultivo 120h en presencia de AT, POST_AT_3) a 31.19 (EV temprano en presencia de AT, EV_AT_2) millones de lecturas mapeadas, con una mediana

de 14.45 millones de lecturas mapeadas. El amplio rango en el número de lecturas exige normalización como parte del análisis de expresión diferencial mediante DESeq2.

Adicionalmente, se evaluó el mapeo de lecturas sobre el genoma de cerdo (*Sus scrofa*) (Figura 8B) y el ya realizado sobre *T. solium* (Figura 8C) para descartar contaminación por parte de material biológico del hospedero. En promedio, se obtuvo que solo 1.23% de las lecturas obtenidas se pueden mapear al genoma de *S. scrofa*, mientras que sobre el genoma de *T. solium* se mapearon en promedio 98.47% de lecturas únicas. Estas observaciones descartan la posibilidad de contaminación cruzada con transcritos de *S. scrofa*.

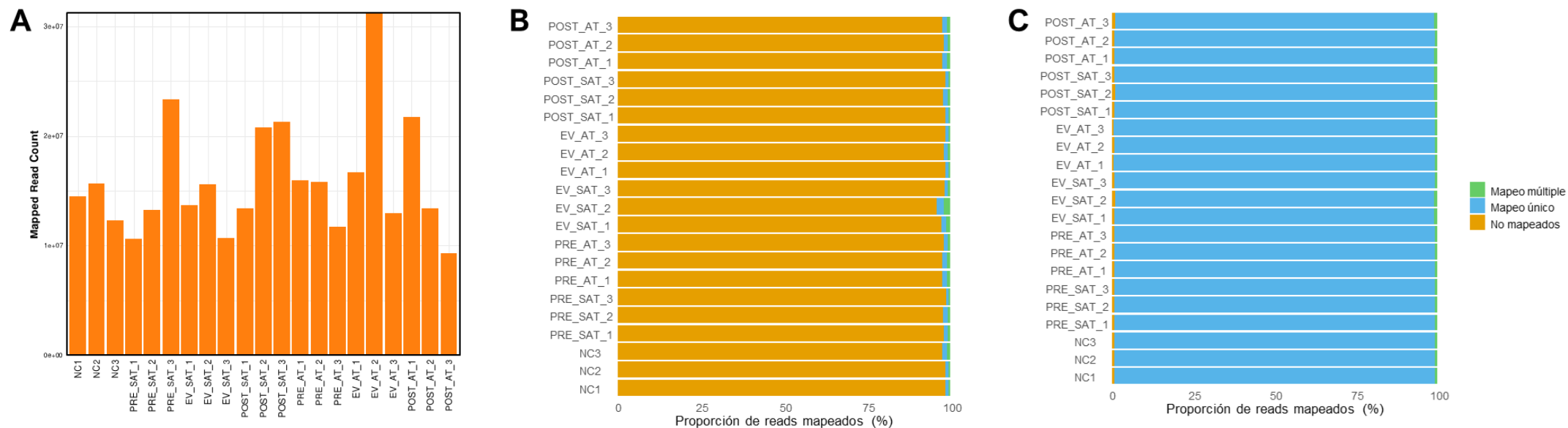


Figura 8. Control de calidad de los datos de secuenciación por RNAseq. (A) Distribución de la abundancia de lecturas mapeadas al genoma de *T. solium*. (B y C) Proporción de lecturas mapeadas de forma única, multiple o no mapeadas al genoma de *Sus scrofa* (B) o al genoma de *Taenia solium* (C) con respecto al total de lecturas secuenciadas.

2. Los análisis de agrupamiento jerárquico y reducción dimensional evidencian cambios en el perfil transcripcional de acuerdo con la fase de desarrollo y la presencia del inductor de evaginación

Para determinar a nivel global cambios a nivel transcripcional en cada una de las muestras y por efecto de las condiciones experimentales (presencia y ausencia de AT) se aplicó agrupamiento jerárquico y PCA.

Para el agrupamiento jerárquico, a partir de la matriz normalizada de conteo de lecturas, se aplicó transformación estabilizadora de la varianza para construir la matriz de distancias Euclidianas. Esta matriz está representada en la figura 9A. El dendrograma distingue dos grandes grupos: NC-PRE y EV-POST.

Dentro de NC-PRE, se distinguen los subgrupos NC y PRE, donde este último se subdivide de acuerdo con AT o SAT. Dentro de EV-POST, el subgrupo POST no se resuelve de forma clara, a diferencia del subgrupo EV, que se subdivide según las condiciones experimentales.

El PCA se construyó tomando en cuenta los 500 genes de mayor varianza. Se obtuvieron los componentes principales (PC) 1 y 2 que explican el 49% y 24% de la varianza total (Figura 9B). Las tendencias son similares a lo observado por agrupamiento jerárquico: sobre el PC1, NC se distribuido hacia valores negativos seguido de PRE (tanto en presencia como ausencia de AT). En el caso de PRE, existe una estratificación entre AT y SAT sobre el PC2.

Hacia los valores positivos EV_AT y EV_SAT se encuentran distantes (tanto en PC1 como PC2), mientras que sólo dos muestras de POST (tanto AT como SAT) tienen se agrupan entre sí.

Estos análisis permiten determinar que el perfil transcripcional de cada muestra depende directamente de la condición experimental, la cual involucra la fase de desarrollo y la presencia de AT. Particularmente, el efecto de AT es claro en las muestras PRE y EV.

Adicionalmente, NC no tiene semejanza con las demás condiciones experimentales al ser individuos no cultivados y es imposible determinar si los cambios transcripcionales respecto a las demás condiciones se deben al efecto del cultivo o la fase de desarrollo. Por otro lado, POST es un grupo altamente heterogéneo a pesar de estar sujeto a las mismas condiciones de cultivo que PRE y EV, además de estar condicionado a una variable temporal. Por estos motivos, los análisis posteriores se centrarán únicamente en las respuestas transcripcionales inducidas por AT en las fases de desarrollo PRE y EV. Estas fases de desarrollo reflejan un momento de activación temprano caracterizado por el evento de evaginación del escólex en condiciones de cultivo que pretenden simular la inducción fisiológica por bilis.

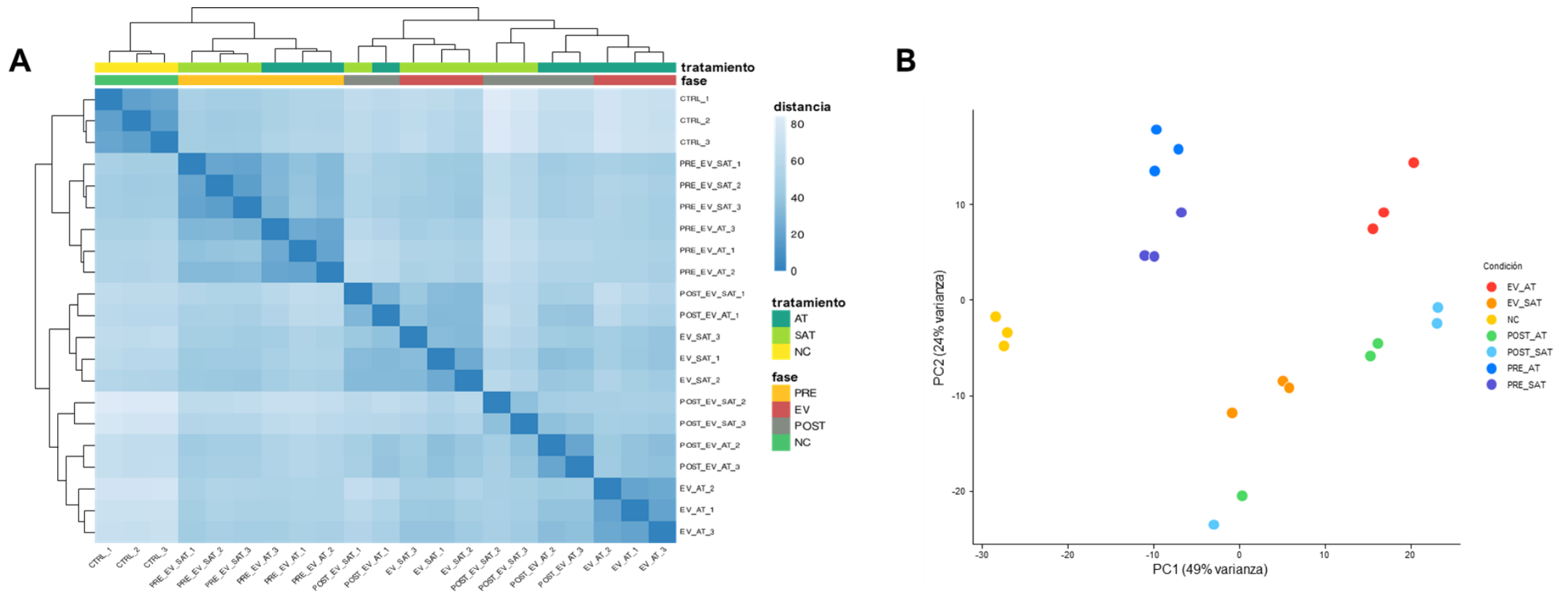


Figura 9. Análisis de agrupamiento jerárquico y reducción dimensional sobre una matriz de conteos sometida a transformación estabilizadora de la varianza (VST) (13). (A) Mapa de calor con agrupamiento jerárquico de los grupos de la matriz de distancias obtenida a partir de la matriz de conteo de genes. (B) Análisis de componentes principales (PCA) de la matriz de conteos con varianza estabilizada con componentes PC1 y PC2 que explican el 49% y 24% de la varianza de los conteos respectivamente.

3. La presencia de AT determina cambios drásticos en la expresión de genes antes y después de la evaginación del escólex

Se determinó la expresión diferencial de genes (differentially expressed genes, DEG) al comparar las condiciones EV y PRE, ambas en presencia de AT (Figura 10A). El resultado fueron 895 DEG (403 sobreexpresados y 492 reprimidos en EV respecto a PRE). Al inspeccionar los 10 genes con mayor nivel de sobreexpresión y represión (Anexo 1) no es posible obtener funciones biológicas relevantes: se encontró sobreexpresión en EV de dos genes que codifican fosfatasa alcalinas, tres ectonucleotido-pirofosfatasas fosfodiesterasas (particularmente *enpp-4*), el gen que codifica la proteína GLIPR-1; y represión de una glicosiltransferasa y dos antígenos de diagnóstico GP50.

La comparación de EV y PRE en ausencia de AT (Figura 10B) obtuvo 420 DEG, de los cuales 182 se encontraron sobreexpresados y 238 reprimidos en EV con respecto a PRE. Dentro de los 10 genes con mayor nivel de sobreexpresión o represión, se encontró sobreexpresión del gen ortólogo codificante de la proteína ECG_00274 en *E. granulosus* y una proteína de unión a ligando hidrofóbico, además de represión de una glicosiltransferasa y el antígeno de diagnóstico GP50 (Anexo 2). No fue posible obtener funciones biológicas relevantes.

Al contrastar las comparaciones entre EV y PRE en presencia y ausencia de AT (Figura 12A) se encontraron 706 DEG propios de la presencia de AT en el medio y 231 genes exclusivos del cultivo en ausencia de AT. Además, 189 DEG son insensibles a la presencia del inductor.

Al comparar individuos EV en presencia y ausencia de AT se obtuvieron 452 DEGs: 349 genes sobreexpresados y 103 genes reprimidos (Figura 11A) en AT

con respecto a SAT. Al inspeccionar los 10 genes con mayor nivel de sobreexpresión (Anexo 3) se encuentran *glipr1*, dos genes que codifican fosfatasa alcalina, tres genes que codifican ectonucleótido-pirofosfatasas fosfodiesterasas y una arginasa-2. Al inspeccionar los 10 genes con mayor nivel de represión, se encuentra una glicosiltransferasa, dos proteínas de diagnóstico GP50 y una proteína que contiene repeticiones de dominios de ankirina. No se pueden establecer funciones biológicas relevantes en este subconjunto.

Por otra parte, en individuos PRE en presencia y ausencia de AT se encontraron 187 DEG, de los cuales 104 genes están sobreexpresados y 83 genes reprimidos en AT con respecto a SAT (Figura 11B). Entre los 10 genes con mayor nivel de sobreexpresión (Anexo 4), la proteína asociada a DAZ-1, una proteína de tegumento y citidina deaminasa. Entre los 10 genes con mayor represión, enolasa C, una proteína con dominios fibronectina y el potenciador transcripcional TEF-3.

Al contrastar las comparaciones de la presencia de AT en EV y PRE (Figura 12B) se encontraron 389 DEG propios de la condición EV, 121 DEG propios de la condición PRE y 63 DEG indiferentes a la fase de desarrollo.

Por ende, fue posible identificar conjuntos de DEG al comparar fases de desarrollo y presencia del inductor. Los conjuntos de mayor tamaño se observan al comparar EV y PRE en presencia de AT, y al comparar AT y SAT en EV. Por otro lado, se han identificado algunos genes que podrían tener importancia biológica, sin embargo, para tener una visión más integradora se evaluaron los conjuntos en su totalidad mediante análisis de sobrerrepresentación (ORA) o

conjuntos de genes con patrones de expresión correlacionados mediante WGCNA.

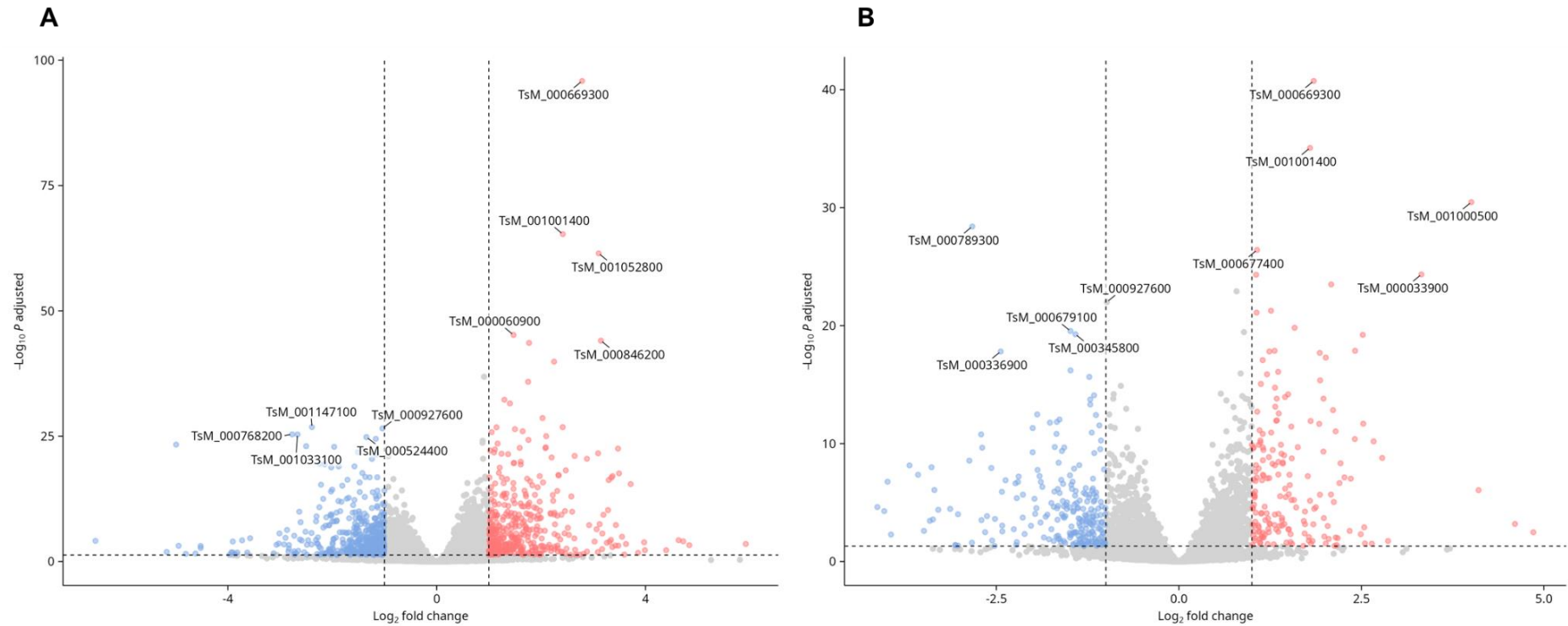


Figura 10. Análisis de expresión diferencial de genes al comparar las condiciones EV y PRE, y la presencia o ausencia de 0.1% AT. Se consideró como genes diferencialmente expresados (DEGs) aquellos que cumplen con los criterios de $|\text{Fold Change}| > 2$, y $p\text{-ajustado} < 0.05$. Los genes sobreexpresados (en EV con respecto a PRE, 3A Y 3B; o AT con respecto a SAT, 3C y 3D) se muestran en rojo, y los reprimidos, en azul (A) Gráfico de volcán en la comparación EV y PRE, ambos en presencia de AT. (B) Gráfico de volcán en la comparación EV y PRE, ambos en ausencia de AT.

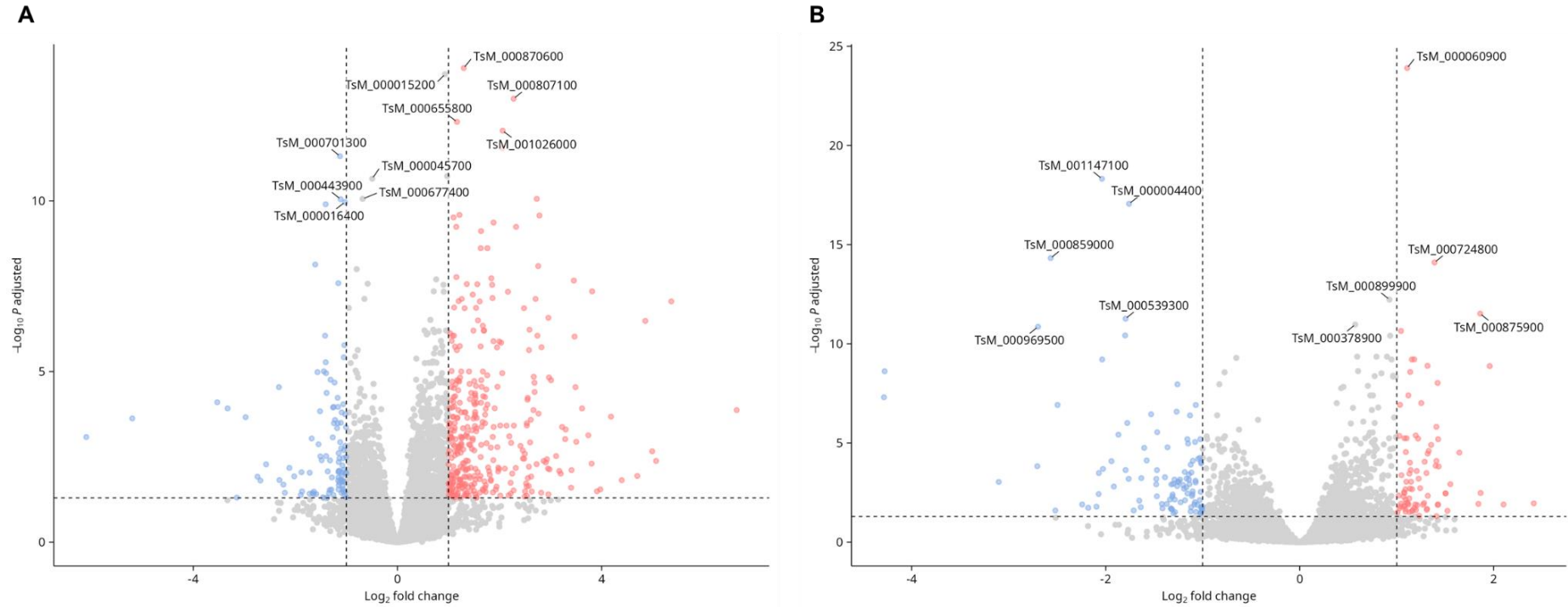


Figura 11. Análisis de expresión diferencial de genes al comparar las condiciones EV y PRE, y la presencia o ausencia de 0.1% AT. Se consideró como genes diferencialmente expresados (DEGs) aquellos que cumplen con los criterios de $|\text{Fold Change}| > 2$, y $p\text{-ajustado} < 0.05$. Los genes sobreexpresados (en EV con respecto a PRE, 3A Y 3B; o AT con respecto a SAT, 3C y 3D) se muestran en rojo, y los reprimidos, en azul. (A) Gráfico de volcán en la comparación AT+ y AT- ambos, en la fase de desarrollo EV. (B) Gráfico de volcán en la comparación AT y SAT ambos en la fase de desarrollo PRE.

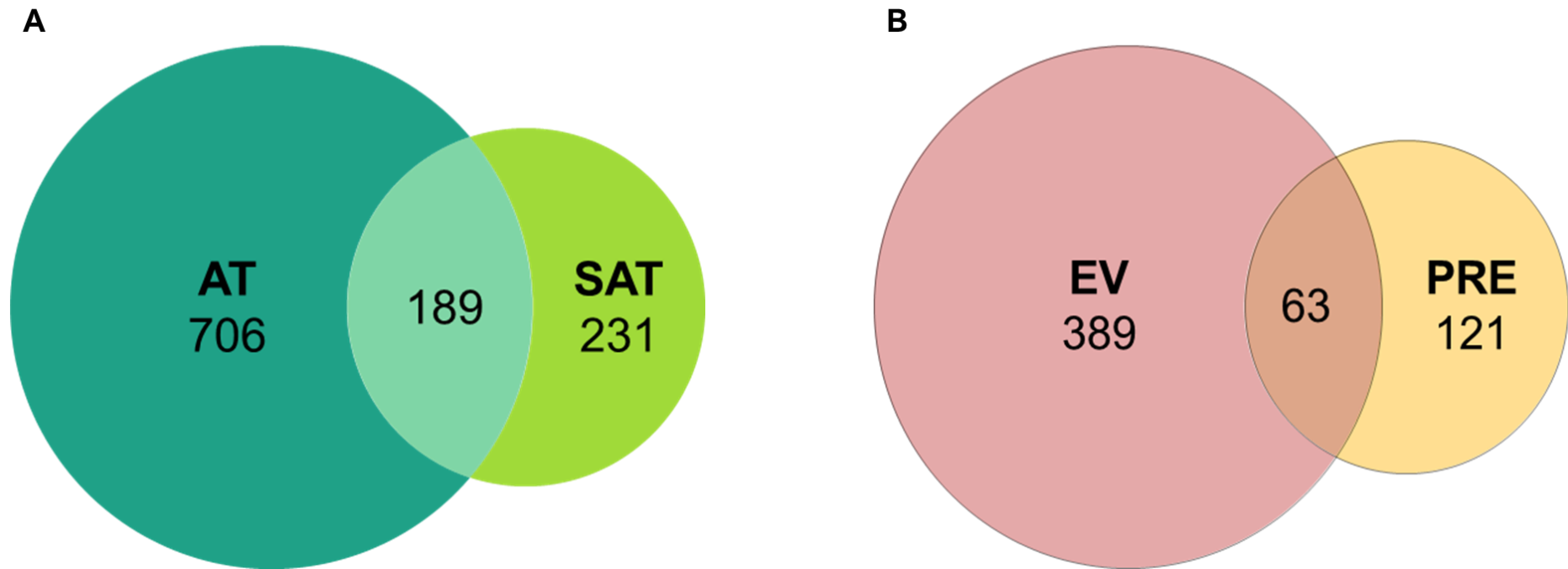


Figura 12. Análisis de expresión diferencial de genes al comparar las condiciones EV y PRE, y la presencia o ausencia de 0.1% AT. Se consideraron como genes diferencialmente expresados (DEGs) aquellos que cumplen con los criterios de $|\text{Fold Change}| > 2$, y $p\text{-ajustado} < 0.05$. (A) Comparación de los DEGs de EV con respecto a PRE en presencia y ausencia del inductor de evaginación. (B) Comparación de los DEGs al comparar AT y SAT en EV y PRE.

4. El análisis de sobrerrepresentación (ORA) de DEGs permitió identificar categorías funcionales GO al comparar fases de desarrollo y la presencia del inductor de evaginación

El análisis de sobrerrepresentación permite evaluar todas las categorías funcionales estadísticamente representativas en el conjunto de DEG. Se evaluaron las categorías funcionales GO pertenecientes a la ontología de procesos biológicos (BP). La figura 13 contiene los términos GO con redundancia reducida (se mantiene la categoría funcional con mayor significancia estadística entre categorías redundantes), mientras que la figura 14 representa todos los términos obtenidos en el análisis como nodos y su grado de redundancia dado por las interconexiones entre estos.

Al evaluar EV vs. PRE, tanto en presencia como ausencia de AT, los términos GO enriquecidos para la condición EV fueron “Proceso basado en microtúbulos” y “Estabilización del potencial de membrana” (Figuras 13 y 14). En el caso del primero, los genes enriquecidos corresponden a tres secuencias “no nombradas” (es decir, no caracterizadas o sin un nombre funcional no asignado) que poseen dominios de la superfamilia de cadenas ligeras de dineína, dos proteínas de tegumento y cuatro cadenas ligeras de dineínas (Figura 15C). En el caso del segundo término GO, solo se encontraron tres secuencias no-nombradas que no poseen ortólogos identificados (la anotación de dos de ellas fueron obtenidas por homología con *Hymenolepis diminuta* y *Taenia asiatica* de la secuencia restante; esta última TsM_000953600, aparentemente tiene elementos transmembrana y secuencia señal anotadas) y un canal de potasio de la familia Twik (*Tandem of P domain in a Weak Inward rectifying K⁺ channel*). Por otro lado, los tres

términos GO enriquecidos para la condición PRE tanto en presencia como ausencia de AT fueron “Organización de la matriz extracelular”, “Transporte transmembrana de ion sodio” y “Comportamiento exploratorio locomotor” (Figuras 13 y 14). En el caso del primero (Figura 15F), se encuentran siete secuencias codificantes de fibras de colágeno. En el caso de “Transporte transmembrana de ion sodio” (Figura 8H) se encuentran dos transportadores de ion cloro-bicarbonato dependientes de sodio, un transportador de sodio-bicarbonato electrogénico, el transportador de fosfato dependiente de sodio 2, un receptor de glutamato, y dos secuencias no nombradas (TsM_000209200 con dominio plicistina y TsM_000940900 con dominios de receptor ionotrópico de glutamato). En el caso de “Comportamiento exploratorio locomotor”, este término GO guarda alta redundancia con el término “Transporte transmembrana de ion sodio” y sus elementos son los dos transportadores de ion cloro-bicarbonato dependientes de sodio y un transportador de sodio-bicarbonato electrogénico mencionados anteriormente, además de una proteína dipeptidil-aminopeptidasa.

Al comparar EV y PRE únicamente en presencia de AT, en EV el término “Proceso catabólico de nucleótidos” guarda alta redundancia con “Regulación positiva de coagulación sanguínea”, “Comunicación celular” y “Proceso biosintético de pequeñas moléculas que contienen nucleobases”. Este último contiene cuatro transcritos que codifican a proteínas ectonucleótido pirofosfatasa-fosfodiesterasas (en particular *enpp-4*) identificados como genes altamente sobreexpresados en el análisis de expresión diferencial. Adicionalmente, dos transcritos del metabolismo de glucosa: piruvato quinasa y

fructosa bifosfato aldolasa. Además, un transcrito no nombrado, TsM_001059000, ortólogo de la ectonucleótido pirofosfatasa-fosfodiesterasa 5 (*enpp-5*) de *T. asiatica*. Otro término enriquecido y de baja redundancia es “Proceso catabólico xenobiótico” (Figura 15G), compuesto por la dihidropirimidina deshidrogenasa de NADP, beta ureidopropionasa, tioredoxina-glutation reductasa, uridina fosforilasa-1, acetilcolinaesterasa y la secuencia no nombrada TsM_001220200, ortólogo de la proteína con dominio CO-esterasa de *Mesocestoides corti*.

En el caso de los términos enriquecidos en PRE en presencia de AT, un grupo de términos GO de alta redundancia están representados en su mayoría por “Regulación positiva de locomoción” (Figura 14 y 15B), el cual contiene a dos ATPasas de transporte de calcio, un ortólogo de *unc-80*, heparan sulfato específico de la membrana basal, *fibulina-2*, la tirosina quinasa Fyn, y cuatro proteínas no nombradas (TsM_000475600, con dominios ATPasa de transporte de cationes; TsM_000521200 y TsM_000756300, con dominios tipo fibronectina-III; TsM_000903300, con dominio de tipo UNC80). Otros términos GO enriquecidos son “Detección de estímulos mecánicos” (constituido por tres proteínas *piezo-2*, dos proteínas *policistina-2*, y dos proteínas no nombradas: TsM_000480701 con dominio tipo PIEZO y TsM_000882800 ortólogo de RUN en *Echinococcus granulosus*), “Transporte de ácidos monocarboxílicos” (constituido por el transportador monocarboxilato-1, la serina-treonina quinasa PAK3, y cuatro proteínas no nombradas: TsM_000068500, compuesta por un único exón y ortólogo a una proteína secretada por *Hydatigera taeniformis*; TsM_001117600, con dominio de la superfamilia de transportadores MFS;

TsM_001232000, con dominio SKIP en un ortólogo en *Echinostoma caproni*; TsM_001232100 una proteína con un único segmento transmembrana predicho) y “Adhesión celular” (anosmina-1, proteína supresora de metastasis-1, la proteína de leucemia mieloide:linfoide o linaje mixto, y siete secuencias no nombradas: TsM_000594300, con ortólogo en *Hymenolepis nana* con dominio PDZ, TsM_000665000 con dominios anosmina-1 y fibronectina-III; TsM_000944100, sin ortólogos ni dominios funcionales anotados; TsM_000963100, con dominio PDZ; TsM_000963200, con dominio PDZ; TsM_000963300, con dominio Par3/HAL; TsM_001209500, con secuencia señal, dominios ricos en cisteína tipo Frizzled, segmento transmembrana y dominio quinasa).

Una categoría funcional común a la presencia de AT tanto en EV como en PRE es “Glicosilación de proteínas” (Figura 13, Figura 14, Figura 15D). Este contiene dos transportadores de cloro-bicarbonato dependientes de sodio, un transportador electrogénico de sodio-bicarbonato, dos glicoproteínas 3-alfa-L-fucosiltransferasa A, beta-13-n-galactosiltransferasa, la proteína Niemann-Pick-C1, la glucuronosiltransferasa *sqv-8*, y dos secuencias no nombradas: TsM_000164900, un ortólogo de hexosil-transferasa de *E. granulosus*; TsM_001220900, una secuencia de un único exón, ortólogo a una proteína con dominios Tnp4 en *H. taeniformis*.

Estas observaciones sugieren que, tomando únicamente en cuenta DEG, la transición de PRE a EV, independientemente de la presencia de AT, induce cambios en la organización de la matriz extracelular y expresión de

transportadores iónicos. Por otro lado, en la fase EV se favorece la expresión de proteínas asociadas al tráfico por microtúbulos y transportadores iónicos. El AT, en la transición de PRE a EV, induce la expresión de genes asociados a glicosilación de proteínas, locomoción y adhesión celular. Mientras tanto, en la fase EV en presencia de AT, se favorece la expresión de genes asociados al metabolismo de nucleótidos y carbohidratos, y estrés oxidativo.

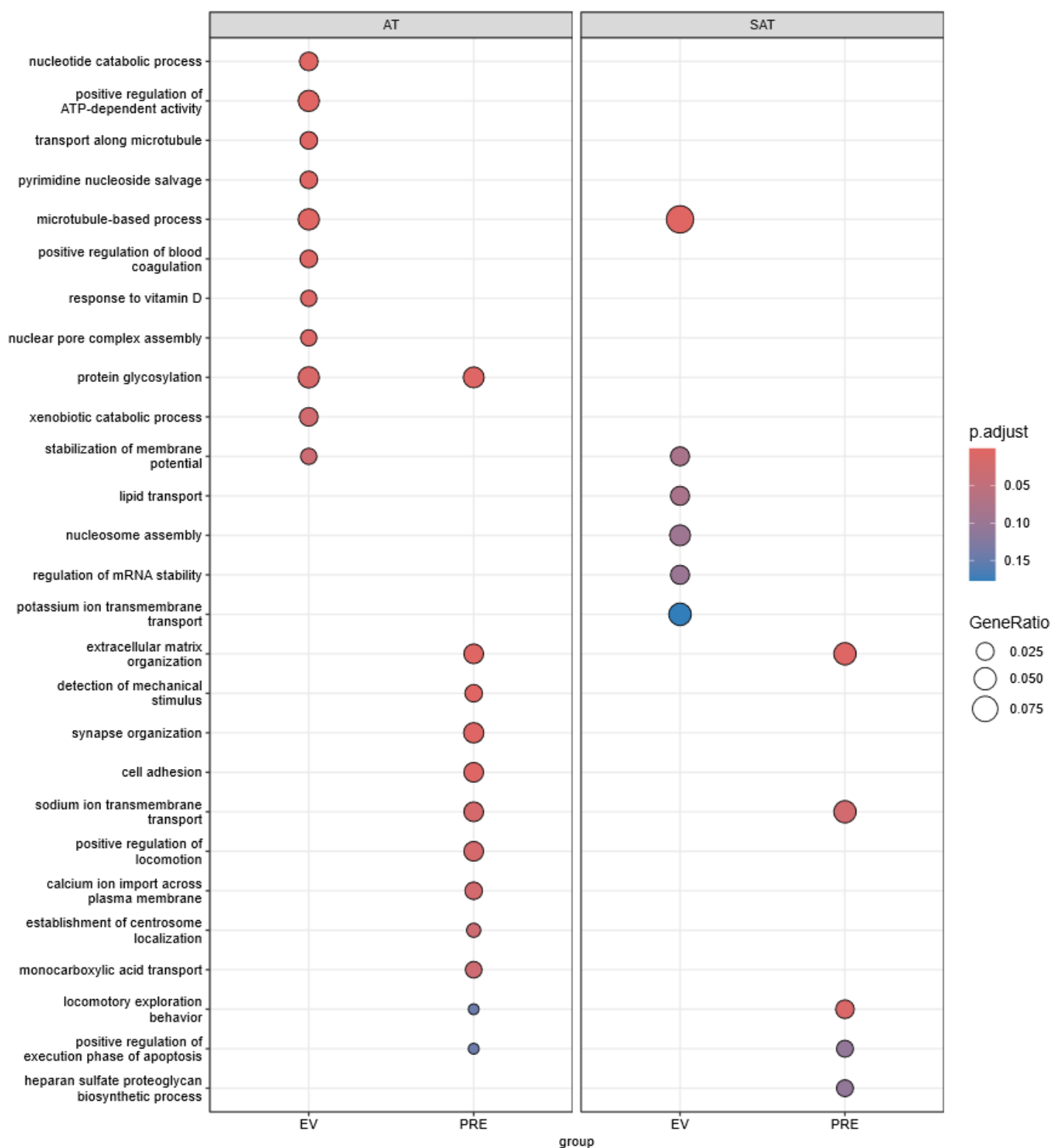


Figura 13. Enriquecimiento de categorías funcionales de ontología génica (GO) mediante análisis de sobrerepresentación (ORA) en DEGs obtenidos al comparar EV y PRE en presencia o ausencia de 0.1% AT. Gráfico de puntos de los términos GO con mayor representación en EV y PRE tanto en presencia como ausencia del inductor de evaginación (p -ajustado < 0.05).

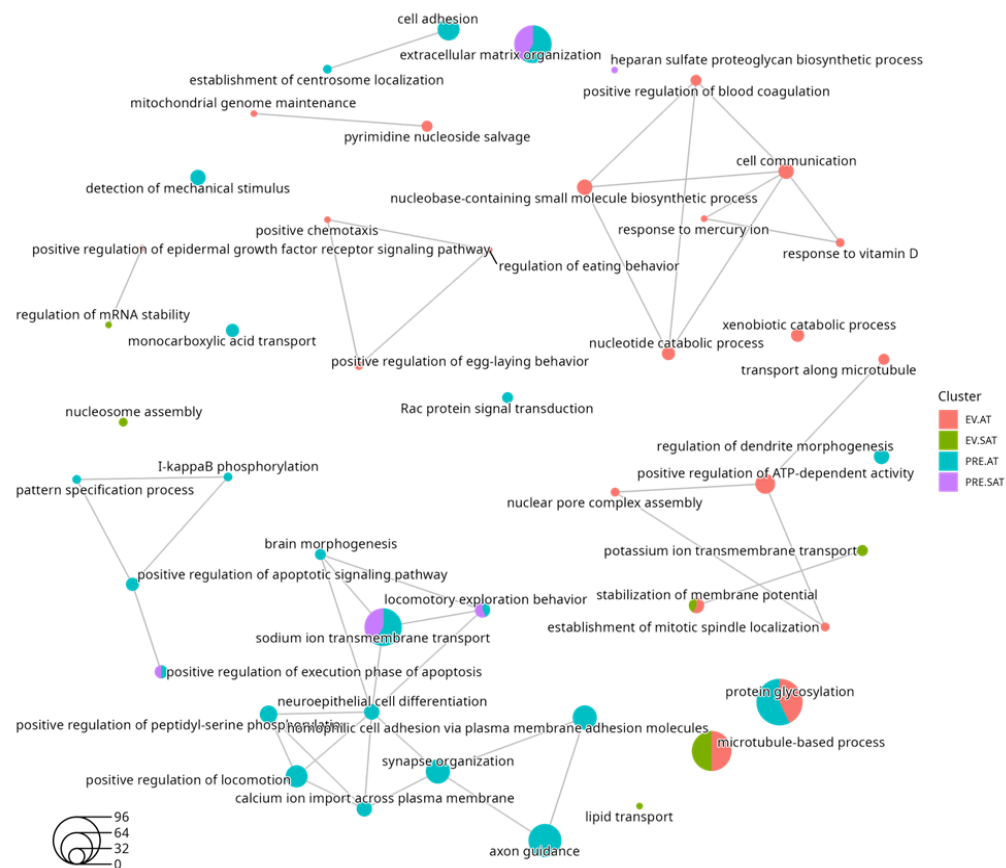


Figura 14. Enriquecimiento de categorías funcionales de ontología génica (GO) mediante análisis de sobrerrepresentación (ORA) en DEGs obtenidos al comparar EV y PRE en presencia o ausencia de 0.1% AT. Redes de elementos compartidos en términos GO enriquecidos en cada combinación de las variables fase de desarrollo y presencia de inductor. Cada nodo es un gráfico circular que representa una categoría funcional y la proporción de elementos correspondiente a la fase de desarrollo y presencia del inductor. Cada arista representa la similitud (genes compartidos) entre categorías funcionales.

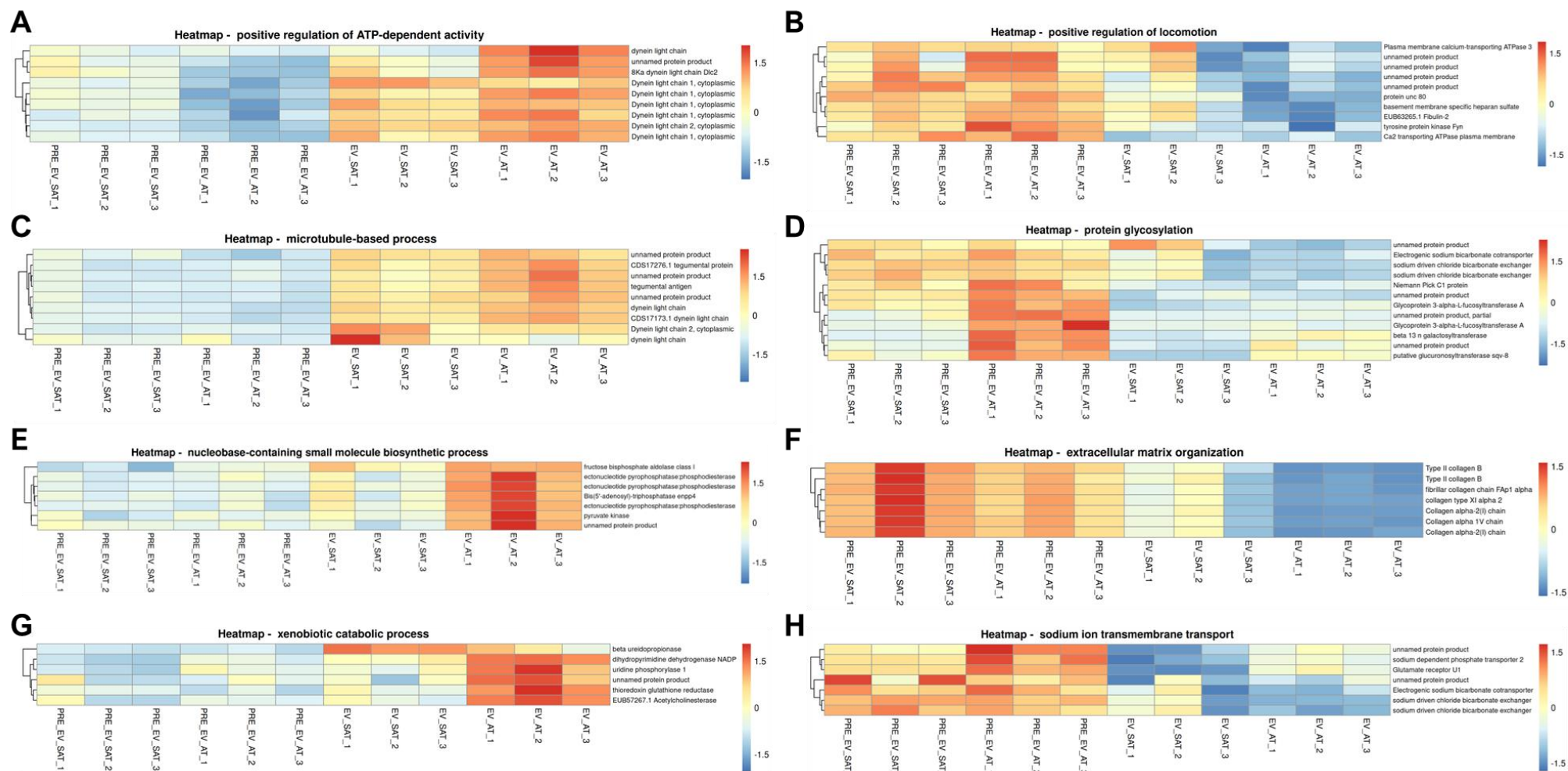


Figura 15. Mapas de calor de las categorías funcionales (A) Regulación positiva de actividad dependiente de ATP, (B) Regulación positiva de locomoción, (C) Procesos basados en microtúbulos, (D) Glicosilación de proteínas, (E) Procesos biosintéticos de moléculas pequeñas que contienen nucleobases, (F) Organización de la matriz extracelular, (G) Proceso catabólico de xenobióticos y (H) Transporte transmembrana de ion sodio. A, C, E y G son categorías funcionales cuyos genes se sobreexpresan en EV respecto a PRE, donde E y G son dependientes de AT. B, D, F y H son categorías funcionales cuyos genes se reprimen en EV respecto a PRE y responden al efecto del AT.

En el caso de los DEG en EV al comparar la presencia y ausencia de AT, y los DEG en PRE al comparar la presencia y ausencia de AT, no se encuentran términos GO comunes (Figura 16 y 17).

Los términos GO enriquecidos en EV en presencia de AT son “Glicosilación de proteínas” (Figura 18E, el cual contiene a Beta-1,3-galactosiltransferasa, dos beta-1,3-n-galactosiltransferasas, la proteína tipo exostosina-3, dos secuencias de UDP-GlcNAc:betaGal-beta-1,3-N-acetilflucosaminiltransferasa-9, la glucuronosiltransferasa *sqv-8*, y tres proteínas no nombradas: TsM_000550700, con un único dominio transmembrana; TsM_000099500 y TsM_000087900, de un único exón con dominios glicosiltransferasa), “Regulación positiva de la vía JNK” (el cual contiene Ankirina, Sintenina-1, serina-treonina quinasa N2, proteína de muerte celular programada-6, *dpak-1*, *gap-2*, *mapkkk-5*, el factor de crecimiento de tejido conectivo y TsM_000272100, con dominio PDZ y ortólogo de la guanilato-quinasa de membrana de *E. granulosus*) (Figura 7A) y “Transporte transmembrana de ion sodio”.

Los términos GO enriquecidos en EV en ausencia de AT son “Organización de la matriz extracelular” (Figura 18B), constituido por colágeno de tipo-II, colágeno de tipo-XI, dos secuencias de colágeno-alfa 2 de tipo I, una secuencia de colágeno de tipo-V y heparan sulfato de la membrana basal; “Morfogénesis de la epidermis” y “Adhesión homofílica celular mediante moléculas de adhesión de la membrana plasmática”.

En el caso de PRE en presencia de AT, se obtuvo “Vía intrínseca de apoptosis en respuesta a estrés en retículo endoplasmático” (Figura 18A) (Xaa-Pro-aminopeptidasa, *gap-2*, la secuencia no nombrada con un dominio transmembrana TsM_000733900, y TsM_001043600, un ortólogo de la proteína *gap-2* de *E. granulosus*). Esta categoría funcional posee alto grado de redundancia con un grupo grande de términos GO propios de PRE_AT (Figura 6B).

Adicionalmente, “Heterooligomerización de proteínas” (Figura 18F) no se encuentra dentro de este grupo de alta redundancia (Figura 10). Esta categoría funcional se compone de cuatro secuencias no nombradas: TsM_000222200 y TsM_000222300, con dominios CBS; TsM_000272100, con dominio PDZ y ortólogo de la guanilato-quinasa de membrana de *E. granulosus*; TsM_000859000, con un único exón, ortólogo de la GTP ciclohidrolasa de *T. asiatica*).

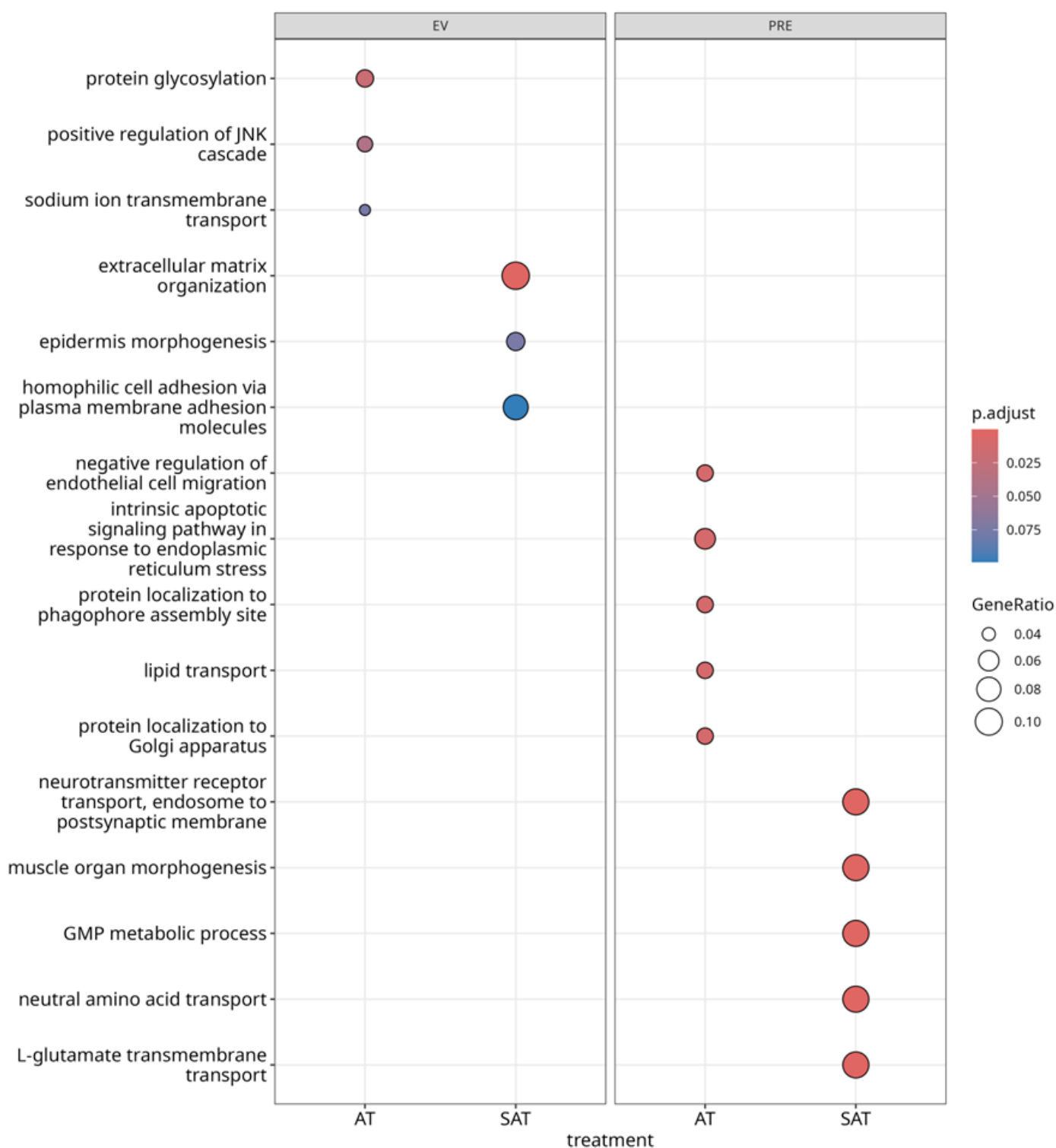


Figura 16. Enriquecimiento de categorías funcionales de ontología génica (GO) mediante análisis de sobrerrepresentación (ORA) en DEGs obtenidos al comparar el efecto de AT tanto en EV como en PRE. Gráfico de puntos de los términos GO con mayor representación en AT y SAT tanto en EV como PRE (p-ajustado < 0.05).

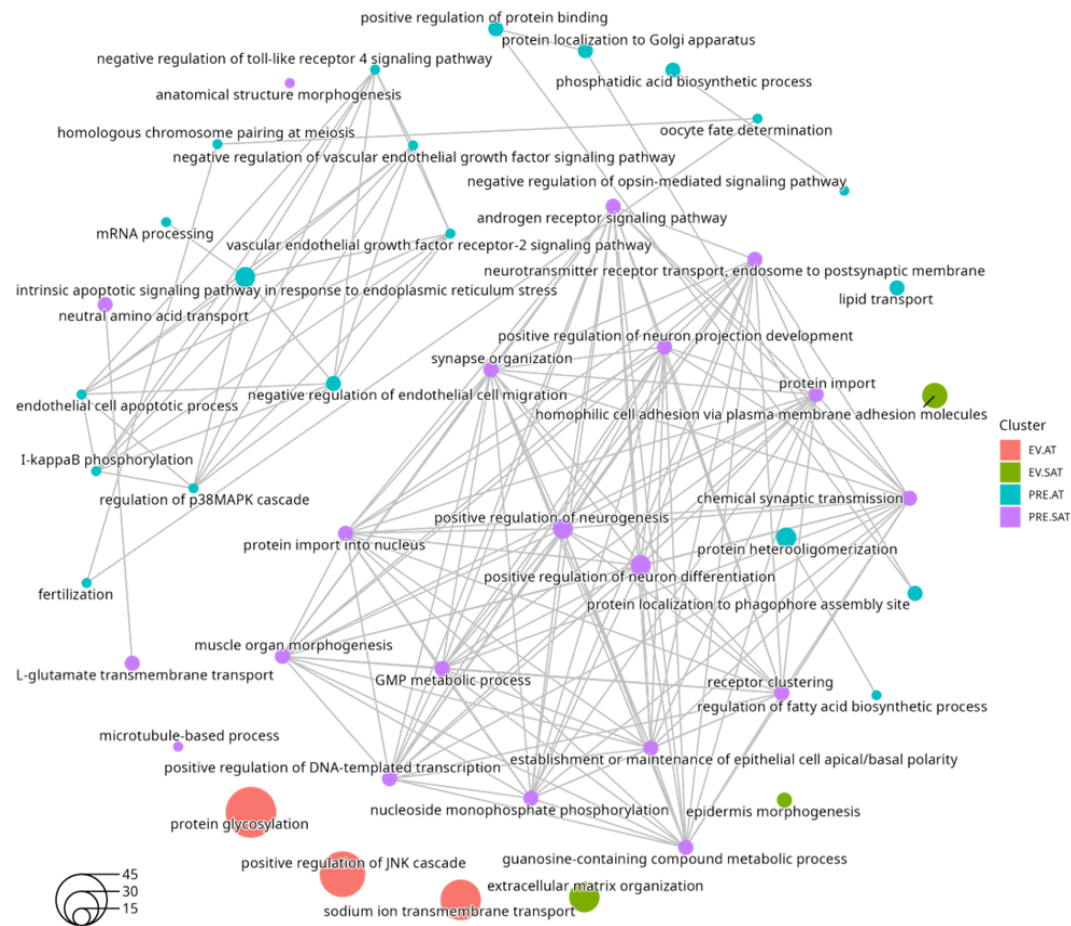


Figura 17. Enriquecimiento de categorías funcionales de ontología génica (GO) mediante análisis de sobrerrepresentación (ORA) en DEGs obtenidos al comparar el efecto de AT tanto en EV como en PRE. Redes de elementos compartidos en términos GO enriquecidos en cada combinación de las variables fase de desarrollo y presencia de inductor. Cada nodo es un gráfico circular que representa una categoría funcional y la proporción de elementos correspondiente a la fase de desarrollo y presencia del inductor. Cada arista representa la similitud (genes compartidos) entre categorías funcionales.

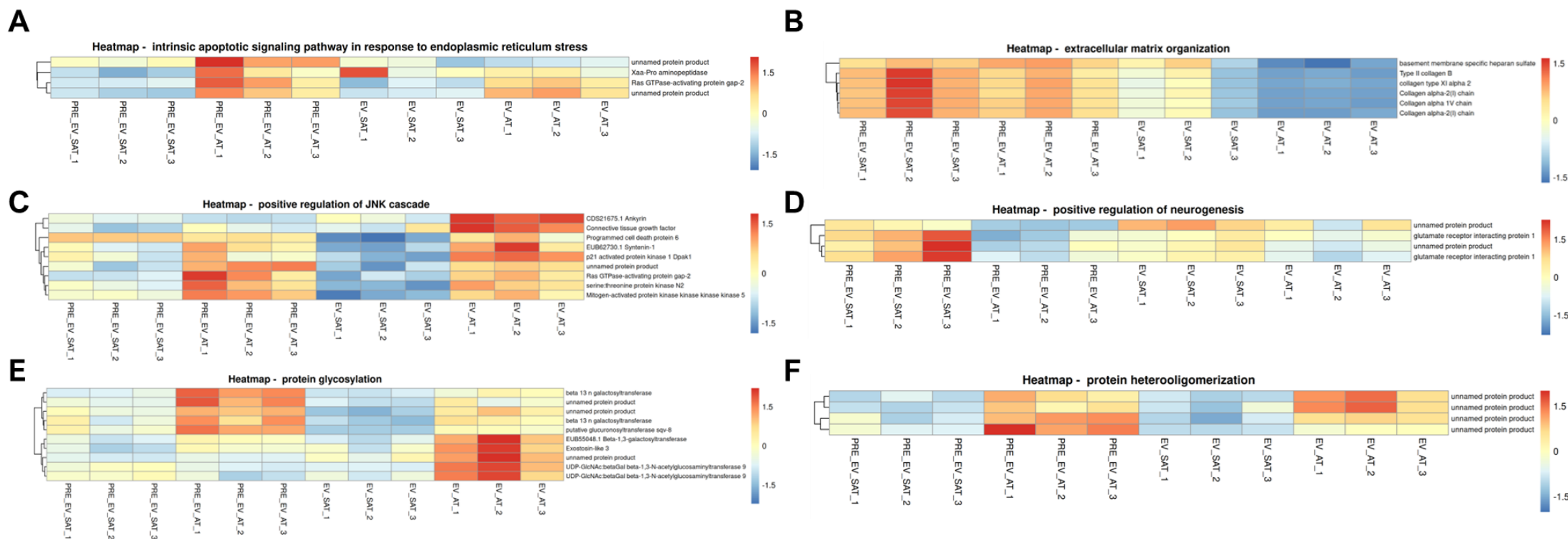


Figura 18. Mapas de calor de las categorías funcionales (A) Vía intrínseca de apoptosis en respuesta a estrés en retículo endoplasmático, (B) Organización de la matriz extracelular, (C) Regulación positiva de la vía JNK, (D) Regulación positiva de neurogénesis, (E) Glicosilación de proteínas y (F) Heterooligomerización de proteínas. A, C, E y F son categorías funcionales favorecidas por efecto de 0.1% AT en EV respecto a PRE. B y D son categorías funcionales cuyos genes se reprimen después de la evaginación en presencia del inductor.

5. El análisis de enriquecimiento de conjuntos de genes (GSEA) permitió identificar términos GO, cuyos genes poseen cambios de expresión sutiles, enriquecidos de acuerdo con la fase de desarrollo y la presencia del inductor de evaginación

A diferencia del análisis de sobrerrepresentación, el análisis de enriquecimiento de conjuntos de genes (*gene set enrichment analysis*, GSEA) utiliza los valores de expresión de todos los genes secuenciados, independientemente de si se encuentran diferencialmente expresados. Esto permite encontrar términos GO enriquecidos asociados a las condiciones experimentales que poseen genes cuya expresión cambia de manera discreta a lo largo del transcriptoma. Para ello se tomó el estadístico de Wald como métrica de ordenamiento, dada su proporcionalidad con la magnitud de expresión y la significancia estadística de cada gen.

Al comparar EV y PRE en presencia de AT (Figura 19) se encontró que en EV con respecto a PRE prevalecen términos asociados a regulación postranscripcional y postraducciona de la expresión de gene (“Corte y empalme de mRNA, vía spliceosoma”, “Proceso catabólico de proteínas dependiente de ubiquitina mediado por el proteosoma”, “Proceso catabólico de proteínas dependiente de ubiquitina mediado por el proteosoma”, “Proceso de biosíntesis de péptidos”, “Terminación de traducción mitocondria”, “Elongación de traducción mitocondria”, “Exportación de mRNA del núcleo”, “Deubiquitinación de proteínas”), además de procesos celulares amplios como “Regulación positiva de actividad dependiente de ATP” (conformado mayormente por cadenas ligeras de dineína), “Proceso catabólico celular”

(altamente genérico; agrupa genes asociados a vías del metabolismo de glucosa, nucleótidos y síntesis de proteínas), “Regulación positiva de actividad hidrolasa” (compuesto mayormente por cadenas ligeras de dineína), “Ensamblaje de cuerpos multivesiculares” (caracterizado por las proteínas endosomales Rab11a y SNF8 - componente del complejo de distribución endosomal ESCRTII), “Localización de complejos ribonucleoproteicos” (conformado por transcritos asociados al poro nuclear, proteínas de unión a RNA, proteínas asociadas a complejos ribonucleoproteicos del spliceosoma, proteínas Ran - GTPasas pequeñas asociadas al transporte entre núcleo y citoplasma y al establecimiento del huso mitótico), “Establecimiento de destino celular” (proteína de unión a Rho p21, Mab-21, SmadA, proteína del síndrome neural de Wiskott-Aldrich), “Localización de complejos que contienen proteínas” (con alta redundancia al término “Localización de complejos ribonucleoproteicos”), “Ensamblaje de nucleosoma” (caracterizado por proteínas de unión a histonas, paralogos de las histonas H2B y H4, además de una proteína con bromodominio adyacente a un dominio de dedo de zinc) y “Autofagia” (consistente de secuencias codificantes de la serina-treonina fosfatasa 2A PP2A y la proteína de fusión de vesículas Nsf). Por ende, en los términos GO enriquecidos en EV y no en PRE en presencia de AT prevalecen, en su mayoría, procesos anabólicos. Estos se caracterizan por la expresión de genes, exportación de RNA, splicing de mRNA y la síntesis, localización y reciclaje de proteínas. Además, genes del metabolismo de nucleótidos y glucosa se encuentran altamente expresados.

Por otro lado, en el estadio PRE con respecto a EV en presencia de AT (Figura 19) prevalecen procesos asociados a la actividad de elementos genéticos móviles

(“Proceso de biosíntesis de DNA mediada por RNA”, “Integración al DNA”), desarrollo del sistema nervioso (“Organización sináptica”, “Diferenciación de células neuroepiteliales”, “Transmisión sináptica colinérgica”), transporte a través de membrana (“Importación de ion calcio a través de membrana”, “Transporte de ácidos monocarboxílicos”, “Regulación del transporte transmembrana del ion sodio”), además de “Adhesión homofílica celular mediada por moléculas de adhesión de membrana”, “Determinación de simetría derecha/izquierda” (redundante con “Motilidad de espermatozoides flagelados” el cual presenta dineínas ciliares propias de células flageladas), “Organización de la matriz extracelular” (colágeno), “Detección de estímulos mecánicos” (proteínas PIEZO y Policistinas), “Homeostasis de estructuras anatómicas”, “Regulación negativa de peptidil-cisteina-S-nitrosilación”, “Regulación negativa de la vía de señalización mediada por opsinas”, “Regulación del ciclo celular en la transición de la fase G1/S” (caracterizado por ATPasas de la membrana plasmática de calcio, redundante con la categoría funcional “Importación de ion calcio a través de membrana”). Por ende, en PRE en presencia de AT y, a diferencia de EV, prevalece la activación de sistemas de regulación génica mediada por elementos genéticos móviles, aumento del transporte de iones y moléculas pequeñas, diferenciación de células espermáticas, formación de matriz extracelular y detección de movimiento.

En ausencia de AT, en EV con respecto a PRE (Figura 20) permanecen (según la presencia de AT) los términos asociados a la síntesis de proteínas (“Proceso biosintético de péptidos”, “Proceso de biosíntesis de amidas” (altamente redundante con el término anteriormente mencionado), “Traducción

citoplasmática”, “Iniciación de la traducción”, “Proceso catabólico de mRNA transcrito en el núcleo, degradación mediada por mutaciones sin sentido” (conteniendo proteínas asociadas al complejo ribonucleoproteico del ribosoma), “Localización de proteínas en membrana mediada por SRP”, “Ensamblaje de subunidad menor del ribosoma”, “Localización de complejos ribonucleoproteicos”) y el término altamente redundante, “Proceso catabólico celular” (el cual incluye en su mayoría genes pertenecientes a los términos anteriormente mencionados además de genes asociados al metabolismo de glucosa y nucleótidos). En este último, cabe mencionar cohesina, la subunidad del complejo promotor de la anafase, catepsina-B y la proteína Pelota.

En el caso del contraste en EV_SAT con respecto a PRE_SAT, los términos GO comunes a la comparación EV_AT con respecto a PRE_AT están relacionados al transporte transmembrana (“Regulación de la plasticidad sináptica” incluyendo ATPasas transmembrana de calcio, “Importación de ion calcio a través de membrana”).

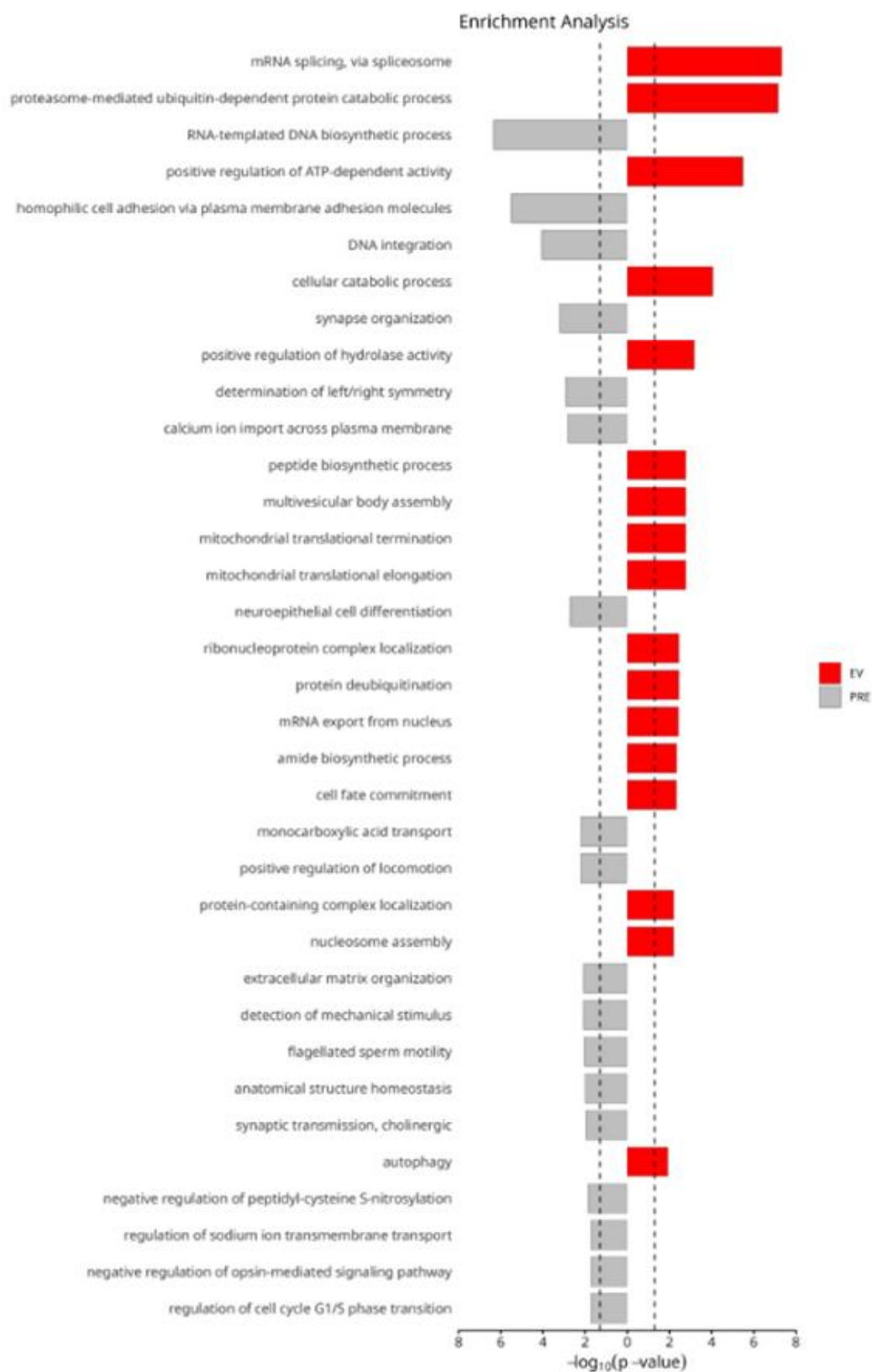


Figura 19. Análisis de enriquecimiento por GSE. Se muestran las categorías funcionales estadísticamente significativas (p -ajustado < 0.05). Comparación de categorías funcionales con un score de enriquecimiento positivo (EV, rojo) o negativo (PRE, gris) en presencia de 0.1% AT.

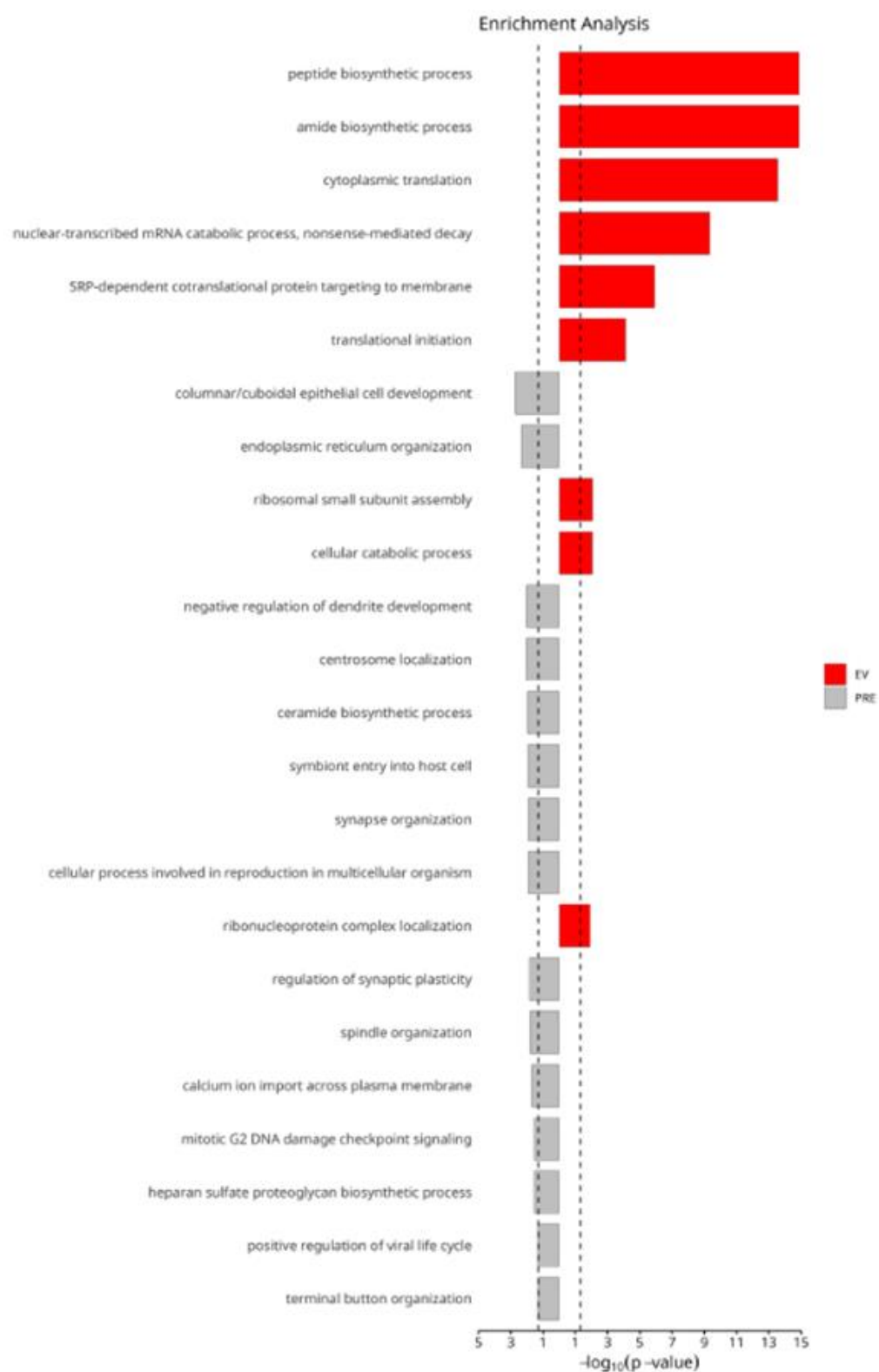


Figura 20. Análisis de enriquecimiento por GSE. Se muestran las categorías funcionales estadísticamente significativas (p -ajustado < 0.05). Comparación de categorías funcionales con un score de enriquecimiento positivo (EV, rojo) o negativo (PRE, gris) en ausencia de 0.1% AT.

Estas comparaciones permiten determinar las categorías funcionales enriquecidas en la transición entre PRE y EV, es decir, lo que ocurre antes y después de evaginar el escólex. En estos resultados, existe consistencia en los procesos que se dan en EV y no en PRE, tanto en presencia como ausencia de AT. Esto no ocurre al evaluar los procesos en PRE que no se dan en EV, tanto en presencia como ausencia de AT. Es posible que esto se deba a que el AT induce mayor variación funcional en el grupo PRE que en el grupo EV. Por ello, para determinar los procesos inducidos por el AT que se aprecian en el grupo EV y, particularmente, que afectan al grupo PRE, se determinaron los términos GO enriquecidos en presencia de AT con respecto a su ausencia en ambas fases de desarrollo (Figura 21 y 22).

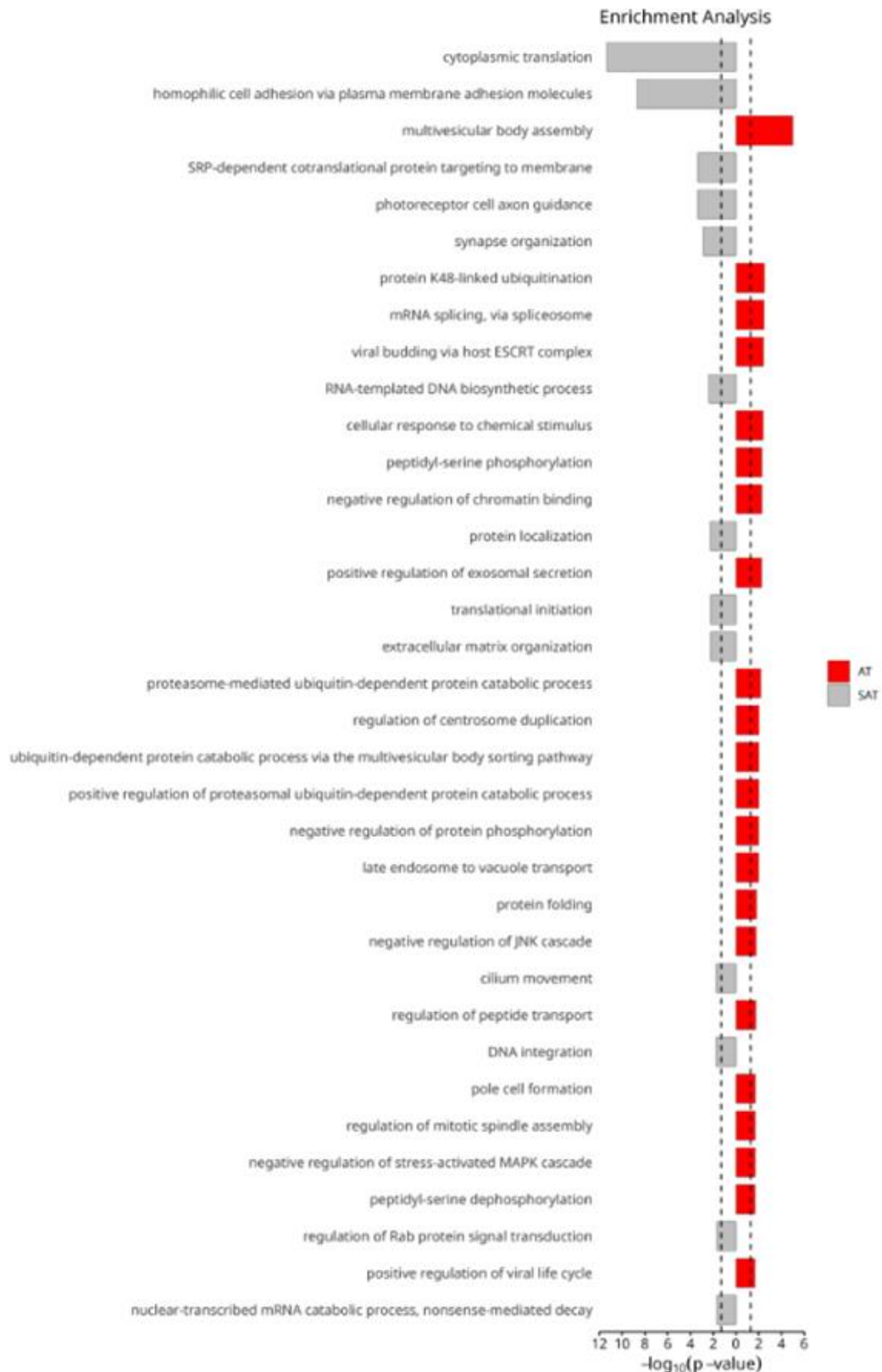


Figura 21. Análisis de enriquecimiento por GSE. Se muestran las categorías funcionales estadísticamente significativas (p -ajustado < 0.05). Comparación de categorías funcionales con un score de enriquecimiento positivo (+AT, rojo) o negativo (-AT, gris) en individuos EV.

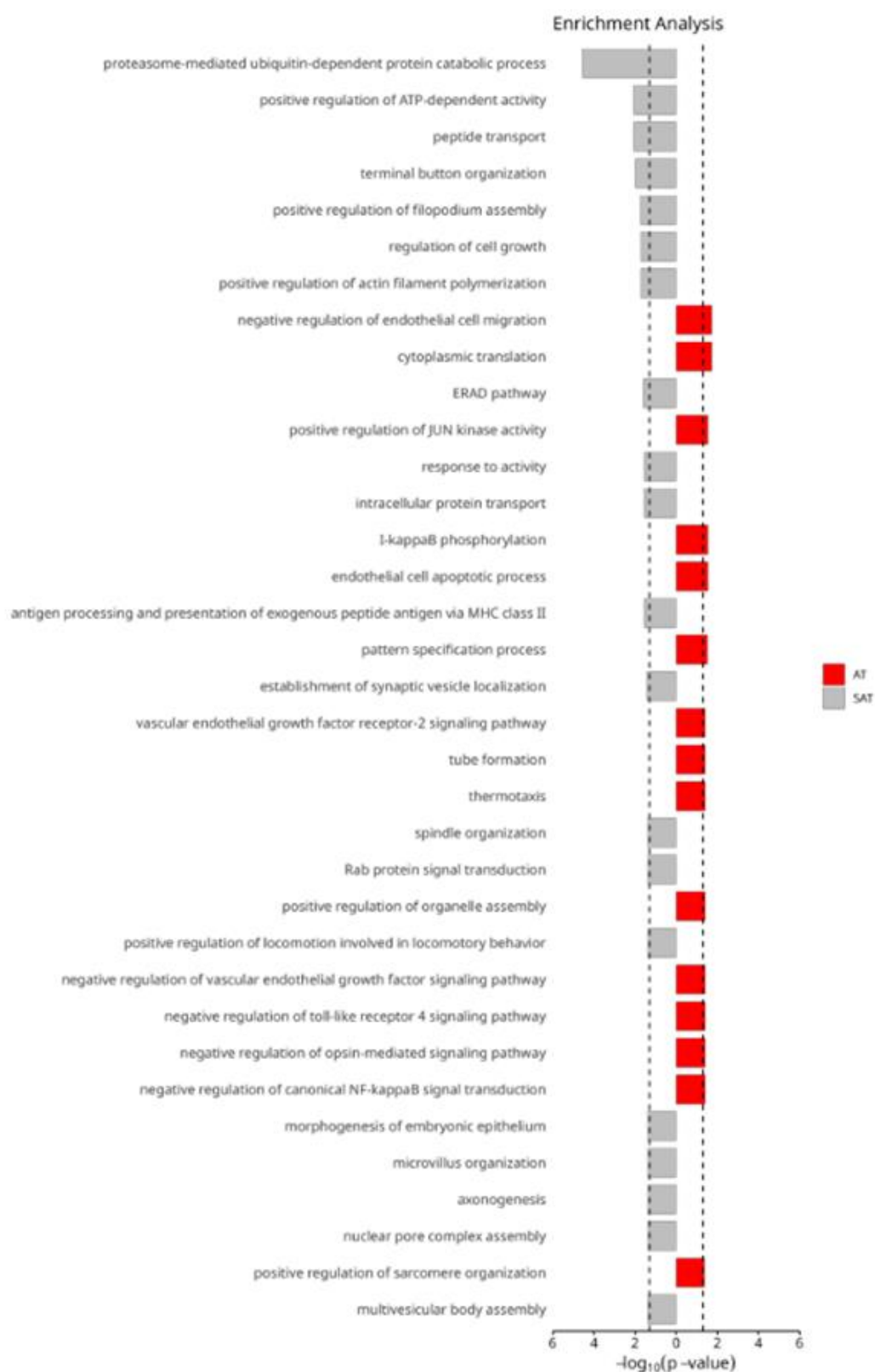


Figura 22. Análisis de enriquecimiento por GSE. Se muestran las categorías funcionales estadísticamente significativas (p -ajustado < 0.05). Comparación de categorías funcionales con un score de enriquecimiento positivo (+AT, rojo) o negativo (-AT, gris) en individuos PRE.

Al evaluar en EV las categorías funcionales enriquecidas solo en presencia de AT (Figura 21), y que por lo tanto estarían inducidas por AT con respecto a PRE, se encontraron términos asociados a secreción (“Ensamblaje de cuerpos multivesiculares”, “Salida de virus mediante el complejo endosomal de distribución requerido para transporte, ESCRT”, “Regulación positiva de secreción de exosomas”, “Regulación positiva del ciclo de vida viral” con secuencias que codifican proteínas de cuerpos multivesiculares endosomales, “Regulación de duplicación del centrosoma”), degradación de proteínas (“Ubiquitinación asociada a K48”, “Regulación positiva del proceso catabólico proteosomal dependiente de ubiquitina”), “Corte y empalme de mRNA mediado por el spliceosoma” (Potenciado por el inductor), “Plegamiento de proteínas” (caracterizado por proteínas de la familia HSP e isomerasas de puentes disulfuro).

Adicionalmente, algunos procesos enriquecidos en EV con respecto PRE independientemente de la presencia de AT (Figura 19 y 20), se encuentran enriquecidos en ausencia de AT con respecto a su presencia en individuos EV (Efecto represor del AT en EV vs. PRE). Estos procesos se asocian a la iniciación de la traducción y localización de proteínas (“Traducción citoplasmática”, “Localización de proteínas en membrana mediada por SRP”, “Localización de proteínas”, “Iniciación de la traducción”), además de “Organización de la matriz extracelular” y “Adhesión celular homofilica mediante moléculas de adhesión de la membrana plasmática”.

Independientemente de la transición de PRE a EV, en EV el AT reprime “Proceso biosintético de DNA con molde de RNA” y favorece la regulación negativa de las vías MAPK y JNK (“Regulación negativa de la cascada JNK”, “Regulación negativa de la cascada MAPK activada por estrés”; ambas categorías de alta redundancia con sobreexpresión de la GTPasa pequeña Ral, PP1B, *unc-97* y la serina-treonina quinasa RAC, también conocida como Akt).

En el caso de PRE, al evaluar los procesos reprimidos por AT (Figura 15), se encuentran “Transporte de péptidos” (además del término redundante “Transporte intracelular de proteínas”), “Regulación positiva de actividad dependiente de ATP” (compuesto mayoritariamente por dineínas), “Organización de botón terminal” (Neurexina-4, SMU-1, VAMP, subunidades beta de actina), “Regulación del crecimiento celular” (subunidad alfa-2 AMPK, PI3K, subunidad alfa de CAMK-II, PUF-60), “Transducción de señales por proteínas Rab” (Rab-3, 4A, 4B, 10, 11A, 14, 27A y un inhibidor de la disociación de GDP en Rab), procesos asociados al citoesqueleto de actina (“Regulación positiva de la polimerización de filamentos de actina”, “Morfogénesis del epitelio embrionario” y “Regulación positiva del ensamblaje de filopodios” con alta redundancia en los genes que codifican PKC, subunidad 2 de Arp2/3, Fascina-2, PAK p21 con unión a Rho, proteína de interacción con la quinasa Abl).

En el caso de los procesos inducidos por AT en PRE (Figura 22), se encuentran “Regulación negativa de la migración de células endoteliales”, “Traducción

citoplasmática” (conformado por proteínas ribosomales mayoritariamente), “Regulación positiva de la quinasa JUN” (MAPKKK-5, serina-treonina proteína quinasa N2, Ras gap-2, PAK-3, Syn-GAP, proteína con dominios Pleckstrina y FYVE, Axina; con redundancia con los términos “Fosforilación de I κ B”, “Apoptosis de células endoteliales”, “Proceso especificador de patrones” (este último incluye una proteína tipo *unc-4*), “Cascada de señalización VEGF-2”, “Formación de tubo”, “Regulación negativa de TLR-4”), “Regulación negativa de la cascada de señalización mediada por opsinas” (Proteínas tipo Tubby, DAG quinasa, fosfatidil-inositol-4,5-bifosfato fosfodiesterasa beta-4).

6. WGCNA permitió identificar módulos de genes cuyos patrones de expresión se encuentran correlacionados y se asocian a la fase del desarrollo EV o la presencia de AT

El GSEA permitió obtener información de conjuntos de genes asociados a una función biológica de acuerdo con sus términos GO. Sin embargo, esto no permite identificar genes cuya expresión se pueda estar regulando de manera conjunta independientemente de las funciones biológicas que asumen. Para ello, se utilizó WGCNA para identificar módulos de genes con patrones de expresión similares y se asociaron estos módulos mediante correlación a las fases de desarrollo y a la presencia del inductor.

Este análisis permitió identificar 24 módulos de genes identificados por colores (Figura 23A). Al correlacionar la expresión de cada módulo con el efecto del AT, se identificaron los módulos amarillo, negro, verde y salmón (Figura 23B). Al correlacionar la expresión de cada módulo con la fase de desarrollo EV, se encontró correlación positiva con los módulos azul y verde amarillento, y correlación negativa con los módulos, turquesa, marrón, verde y azul noche (Figura 23B). Para cada variable de diseño (fase de desarrollo y presencia de inductor) se analizaron en mayor detalle los tres módulos de mayor valor de correlación. En el caso del efecto del AT, se tomó el módulo verde en lugar del módulo salmón a pesar del mayor valor de correlación de este último, debido a que el verde presenta valores de correlación significativos en ambas variables de diseño.

Se evaluó la correlación de cada gen con la variable de diseño respecto a su pertenencia a cada módulo en los módulos amarillo (Figura 23C), verde (Figura 23D) y negro (Figura 23E) en el caso de asociación con AT, y los módulos azul (Figura 23F), marrón (Figura 23G) y turquesa (Figura 23H), en asociación con EV. En el caso de una correlación positiva, los genes en el cuadrante superior derecho del gráfico (alta pertenencia al módulo y asociación con la condición experimental) representan genes de alta interconectividad. Mientras tanto, en el caso de una correlación negativa, los genes en el cuadrante inferior derecho (asociación negativa con EV y alta pertenencia al módulo) representan genes de alta interconectividad.

Para investigar a mayor profundidad estos módulos, se realizó un ORA. Se obtuvo información funcional solamente para los módulos amarillo (asociado con AT, figura 24) y azul (asociado con la fase de desarrollo EV, figura 26). En el caso del módulo amarillo, se obtuvo enriquecimiento funcional de categorías altamente redundantes, principalmente asociadas a cascadas de quinasas MAPK y JNK (Figura 25). Algunos genes comunes a la mayoría de estas categorías funcionales son *dpak-1*, *gap-2*, *s6k*, la subunidad beta 1 de la AMPK, *fyn*, *lamp2*, *ral*, *fer*, el transportador de zinc *zip-8* y la proteína de interacción con disabled de tipo 2. En el caso del módulo azul, se encontraron los términos redundantes asociados a la traducción “Iniciación de la traducción”, “Proceso biosintético de amidas”, “Localización de proteínas a membrana mediante SRP”, “Traducción citoplasmática”, “Proceso biosintético de péptidos”, “Proceso catabólico de mRNA transcrito en el núcleo, degradación *non-sense*”, además de la categoría funcional “Regulación negativa del transporte intracelular” (Representados por

los genes *Rap-1b*, la HSP *p36*, *sinaptotagmina* y *map-2*). Los genes más representativos corresponden a proteínas asociadas al ribosoma, factores de iniciación de la traducción y la proteína Pelota (Figura 27).

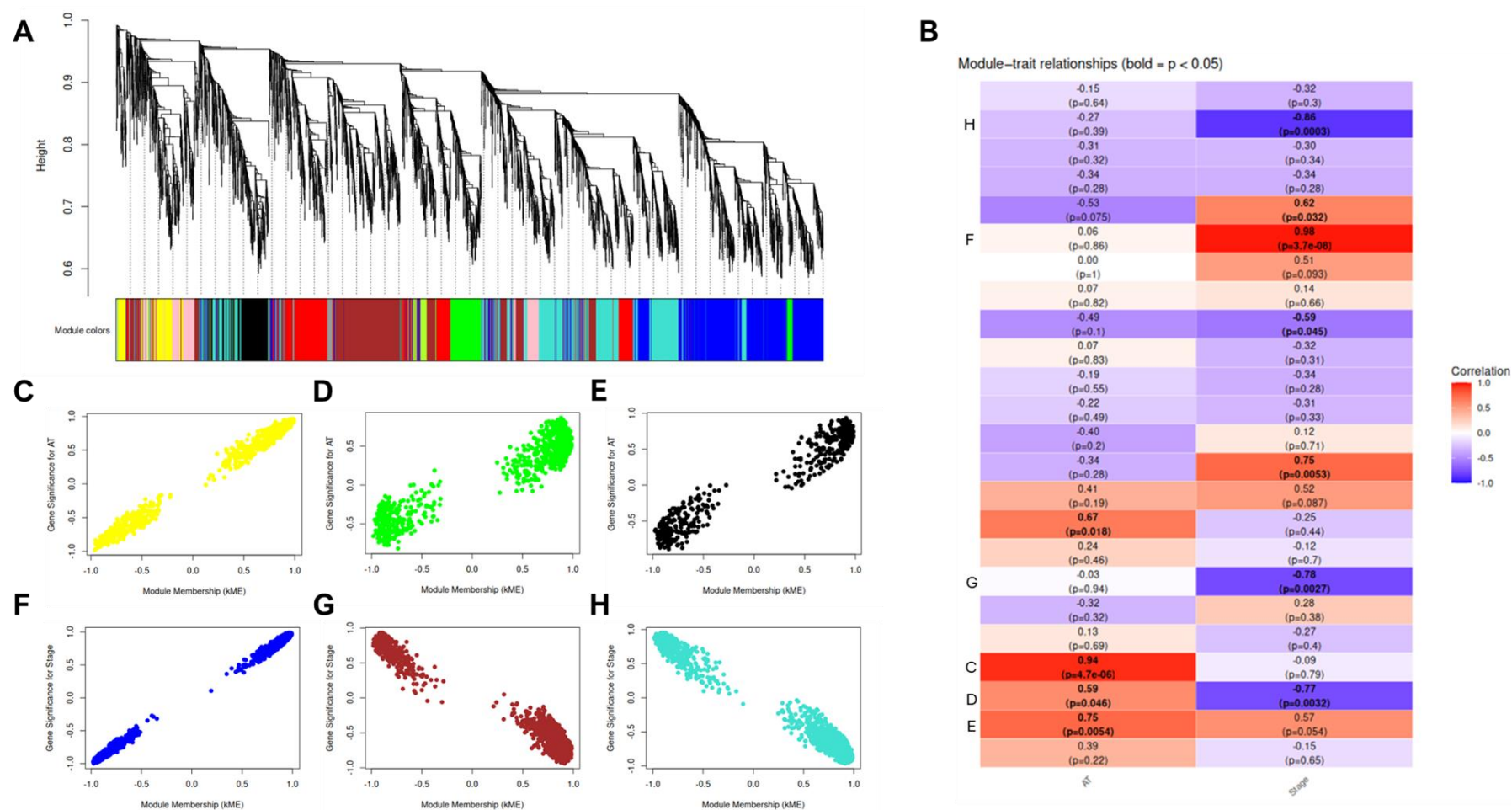


Figura 23. WGCNA del 75% de genes con mayor varianza. (A) Dendrograma de los módulos de genes con niveles de expresión correlativa identificados con al menos 30 genes representados por colores. (B) Mapa de calor de la asociación entre clústeres y condiciones experimentales. Se muestra en cada celda el valor de correlación de Pearson para la asociación del valor propio de cada módulo de genes (*eigengenes*) y la condición experimental (presencia de AT o fase de desarrollo), y el valor de significancia estadística. (C), (D) y (E) muestran un diagrama de dispersión del valor de asociación de cada gen con la presencia de AT y el valor de pertenencia de cada gen al módulo respectivo. Los genes en el cuadrante superior derecho representan genes nodo (*hub*) en la red de correlación de expresión de genes. (F), (G) y (H) muestran un diagrama de dispersión del valor de asociación de cada gen con la fase de desarrollo EV y el valor de pertenencia de cada gen al módulo respectivo.

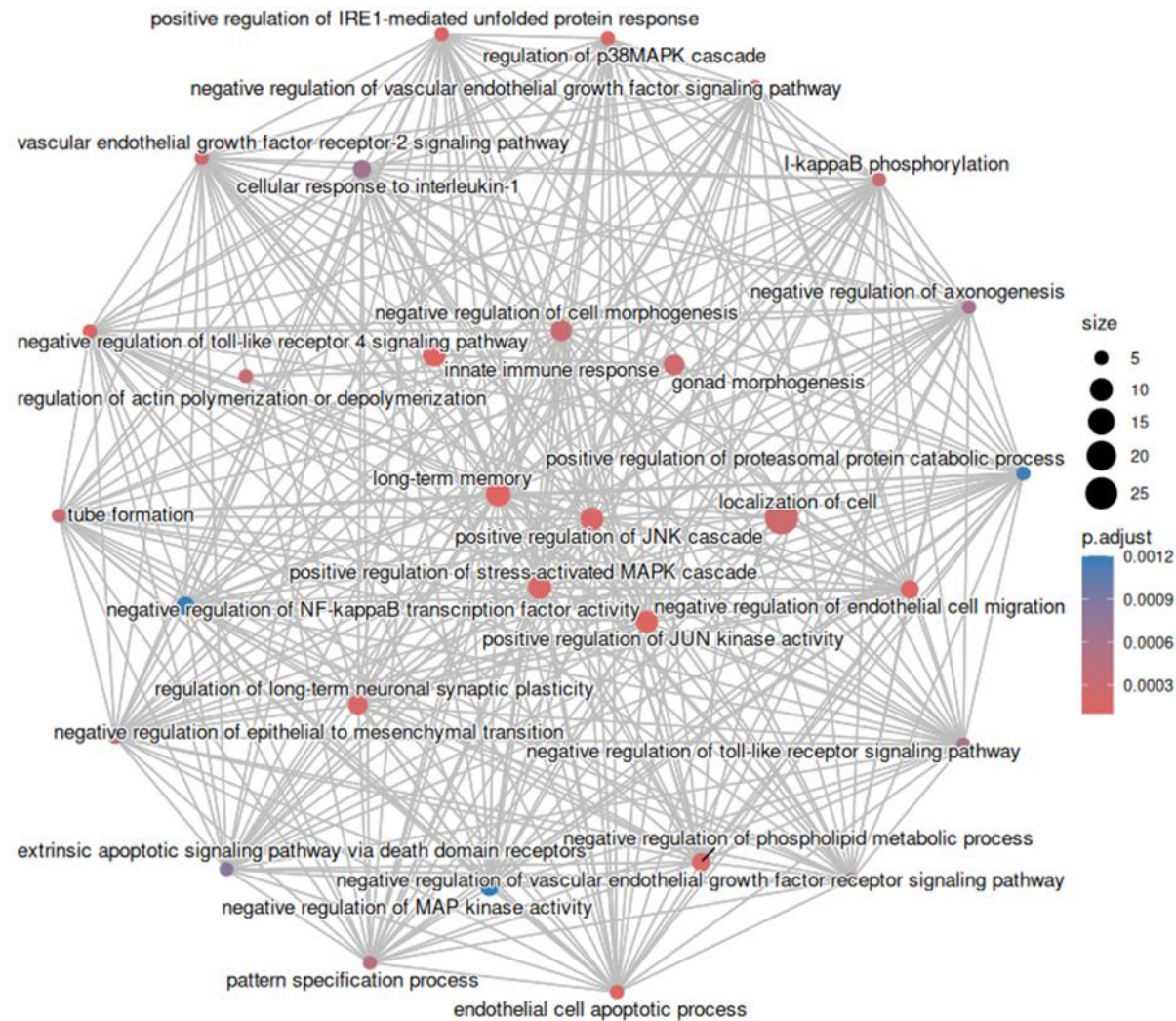


Figura 24. ORA de categorías funcionales enriquecidas en el módulo amarillo asociado al AT. Redes de elementos compartidos en términos GO enriquecidos.

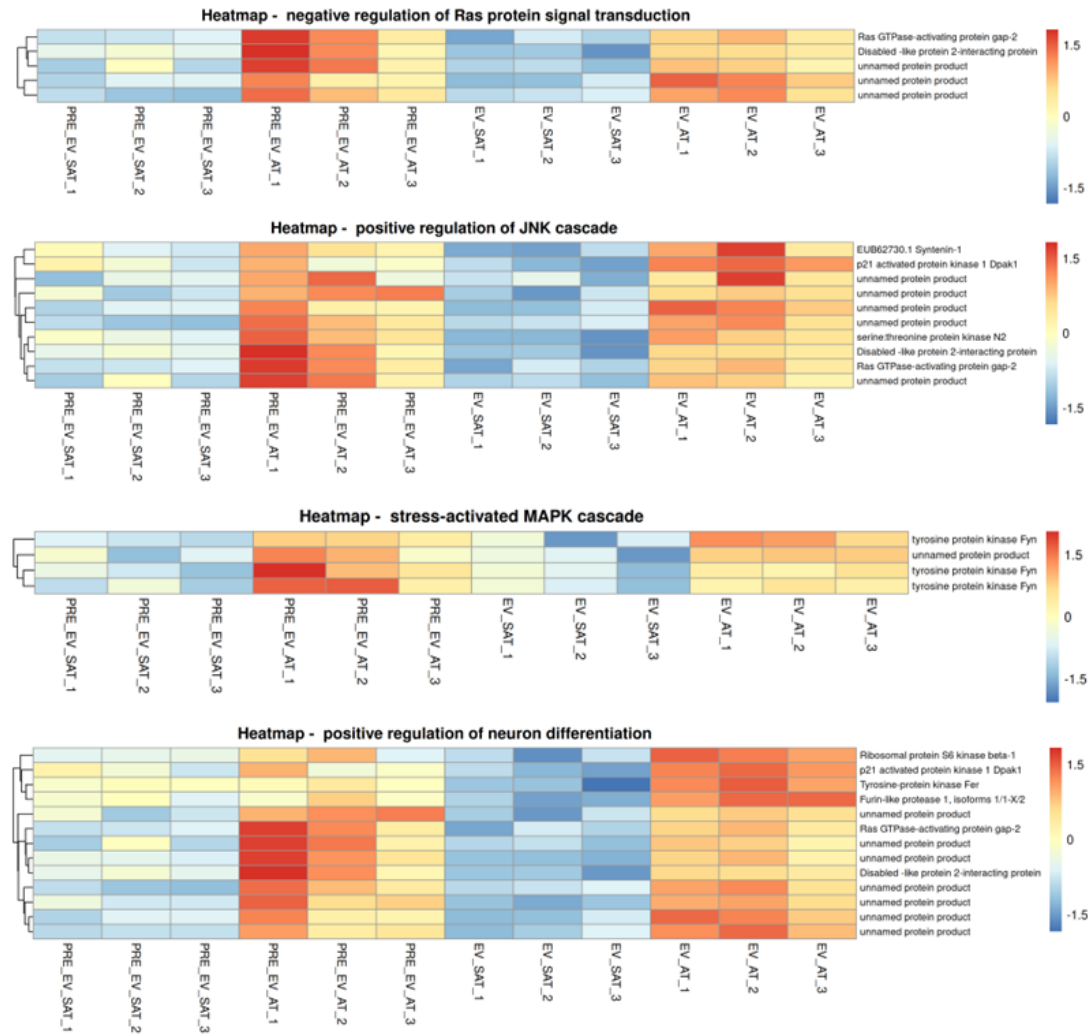


Figura 25. Mapas de calor de las categorías funcionales asociadas al módulo amarillo. “Regulación negativa de la transducción de señales de la proteína Ras”. “Regulación positiva de la vía JNK”, “Vía MAPK activada por estrés” y “Regulación positiva de la diferenciación neuronal”.

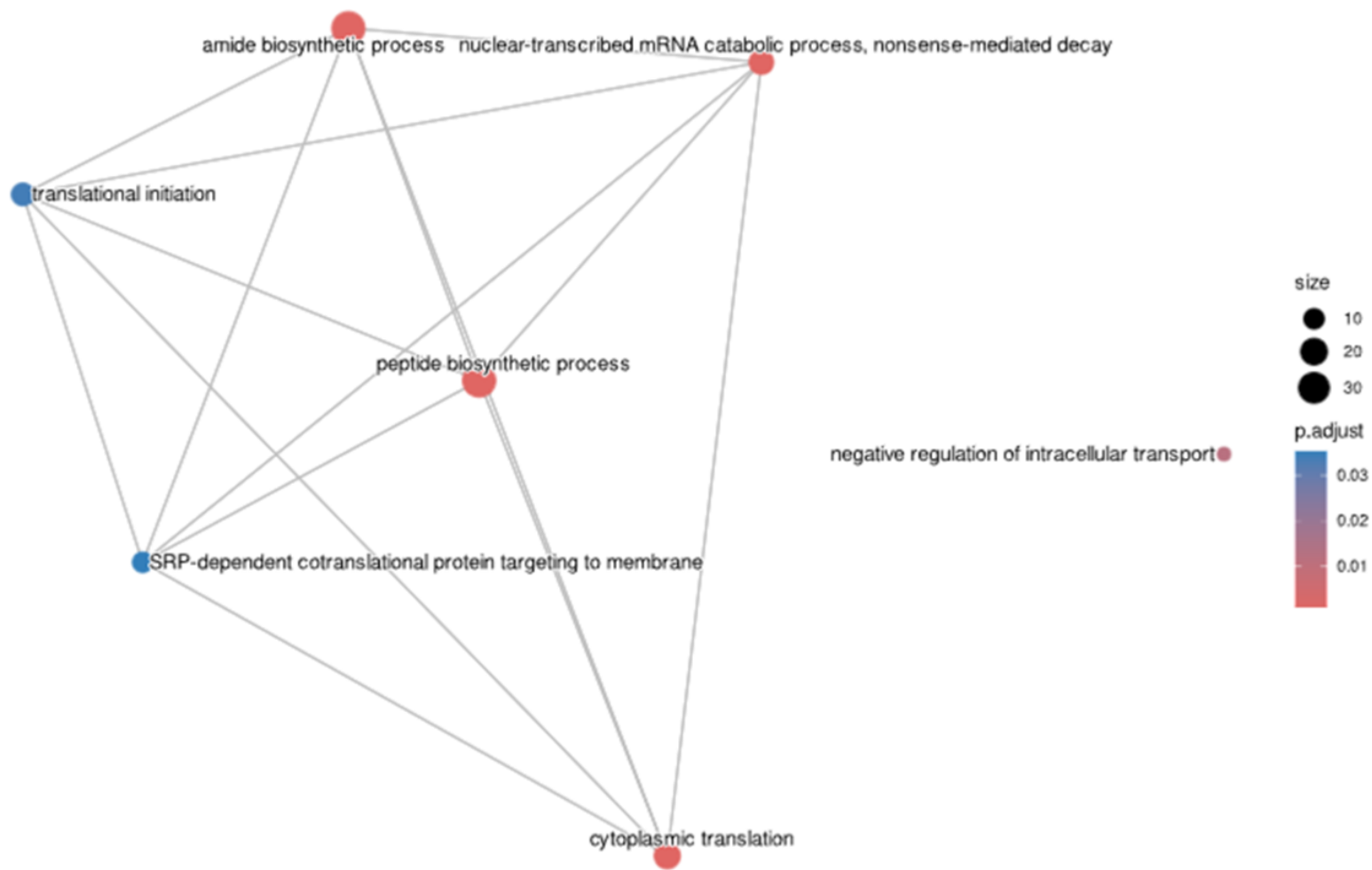


Figura 26. ORA de categorías funcionales enriquecidas en el módulo azul, asociado a EV. Redes de elementos compartidos en términos GO enriquecidos.

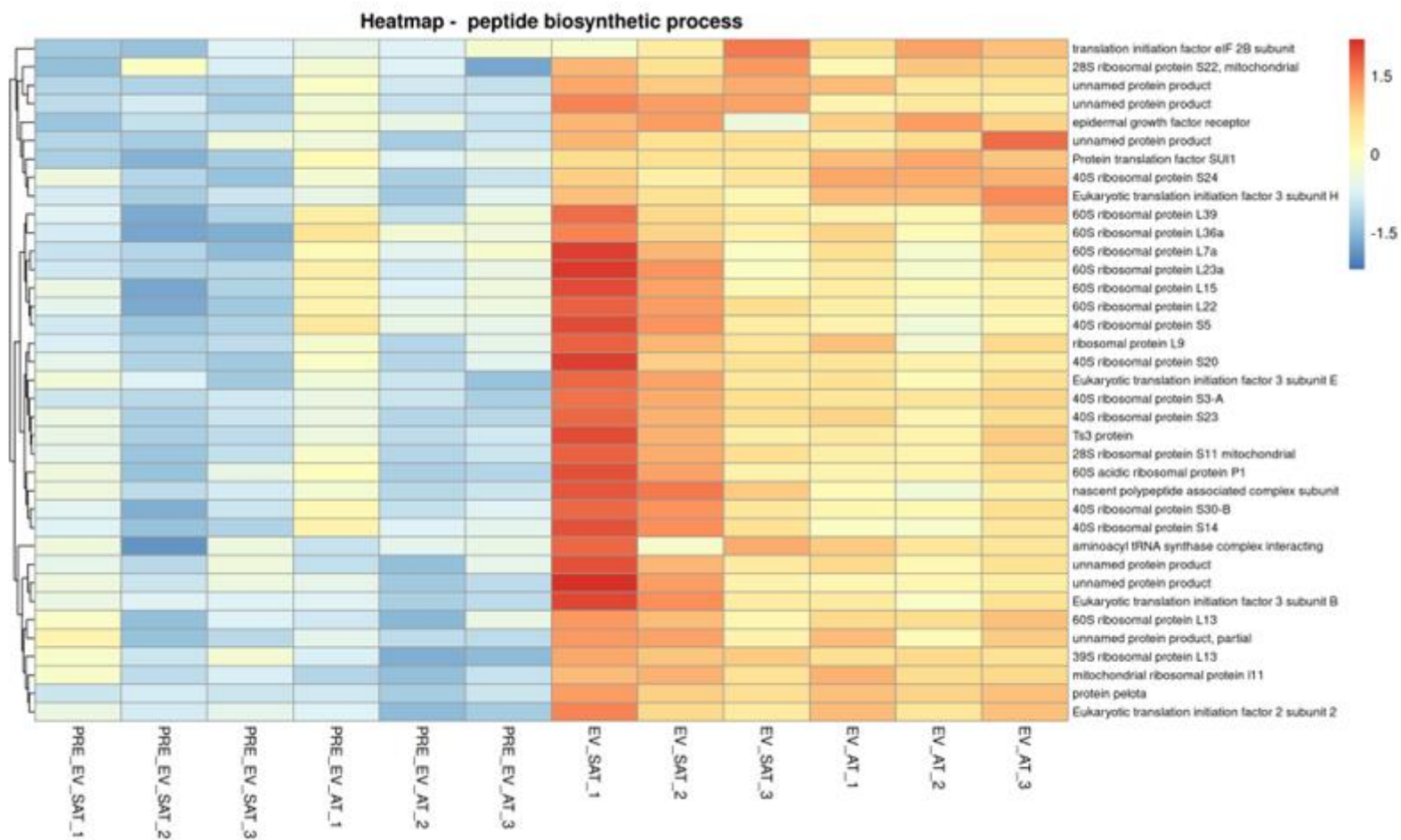


Figura 27. Mapa de calor de la categoría funcional “Proceso biosintético de péptidos” asociado al módulo azul.

7. Las redes de asociación física o funcional STRING permiten validar a nivel de dominios funcionales proteicos las vías representadas en los módulos de genes obtenidos mediante WGCNA

Se utilizó la base de datos de interacción y asociación proteína-proteína STRING para determinar interacción funcional o física entre los elementos de los módulos amarillo y azul. Para ello, a partir de las secuencias de aminoácidos obtenidas desde WormBase ParaSite se construyó una nueva base de datos en la plataforma de STRING. Posteriormente, se evaluó la interacción entre los genes de alta interconectividad de cada módulo y se realizó un enriquecimiento de términos Reactome (una base de datos enfocada en vías de señalización), GO (específicamente los términos asociados a función molecular), COMPARTMENTS (una base de datos de localización subcelular) y palabras clave anotadas en la base de datos UniProt.

Para el módulo amarillo se obtuvo la red de interacción mostrada en la figura 28A. Las vías enriquecidas (Figura 28B) por Reactome fueron “Influjo de Zinc por la familia de genes SLC39”, “Transportadores de Zinc”, “Transportadores de iones metálicos SLC” y “Transporte de sales biliares, ácidos orgánicos, iones metálicos y aminos”. Este último conformado por las secuencias TsM_000927200 (secuencia no-nombrada con dos dominios transmembrana), TsM_000579500 (ortólogo al transportador de zinc, SLC39A1) y TsM_000066900 (ortólogo del transportador SLC39A4).

Adicionalmente, se encontró enriquecimiento de términos de la base de datos COMPARTMENTS de “Fibras paralelas” y “Complejo proteína quinasa

activado por nucleótidos”. El minado de palabras clave de la base de datos UniProt encontró el término “Quinasas”. Por otra parte, los términos GO (en la ontología de función molecular) enriquecidos corresponden a “Unión a receptores de señalización”.

En el caso del módulo azul, se obtuvo una red altamente interconectada de proteínas (Figura 29A). Al investigar las vías Reactome (Figura 29B) a las cuales esta red se asocia, se obtuvo enriquecimiento de vías involucradas en la traducción de proteínas. Las vías con mayor grado de representación fueron “Traducción dependiente de CAP”, “Inicio de la traducción en eucariotas”, “Formación de subunidades 40S libres”, “Silenciamiento de la expresión de Ceruloplasmina mediada por L13a”, “Hidrólisis de GTP y unión de la subunidad 60S”, “Decaimiento mediado por secuencias *nonsense*”, “Localización de proteínas a membrana mediado por SRP” y “Traducción”. En el caso de los términos de ontología génica asociados a función molecular (Figura 22C), se encontró sobrerrepresentados los términos “Constituyente estructural del ribosoma” y “Actividad de moléculas estructurales”. Asimismo, las palabras claves de UniProt con mayor representación (Figura 29D) fueron “Ribonucleoproteína” y “Proteína ribosomal”.

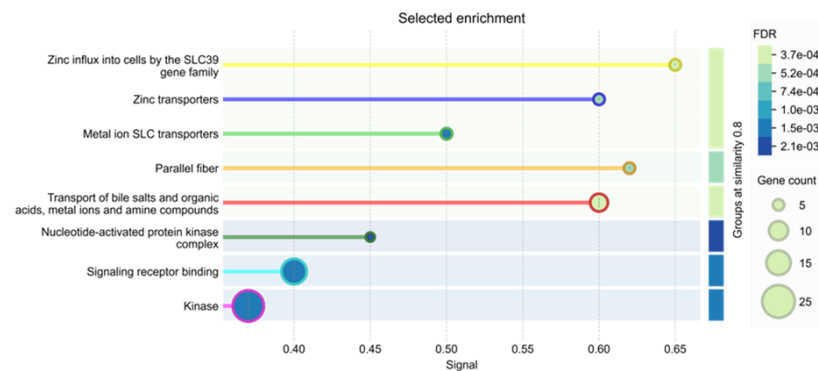
A**B**

Figura 28. Análisis de interacción y asociación proteína-proteína mediante la base de datos de STRING y enriquecimiento de vías anotadas en Reactome. (A) Red de asociación de proteínas del módulo amarillo ($n = 248$) representando las secuencias con asociación funcional o interacción física predicha. En colores se muestran las proteínas asociadas a términos sobrerrepresentados. (B) Análisis de enriquecimiento de Reactome, GO (Función molecular), COMPARTMENTS (localización celular) y palabras clave anotadas (UniProt) para las proteínas del módulo amarillo representadas en la red de asociación e interacción.

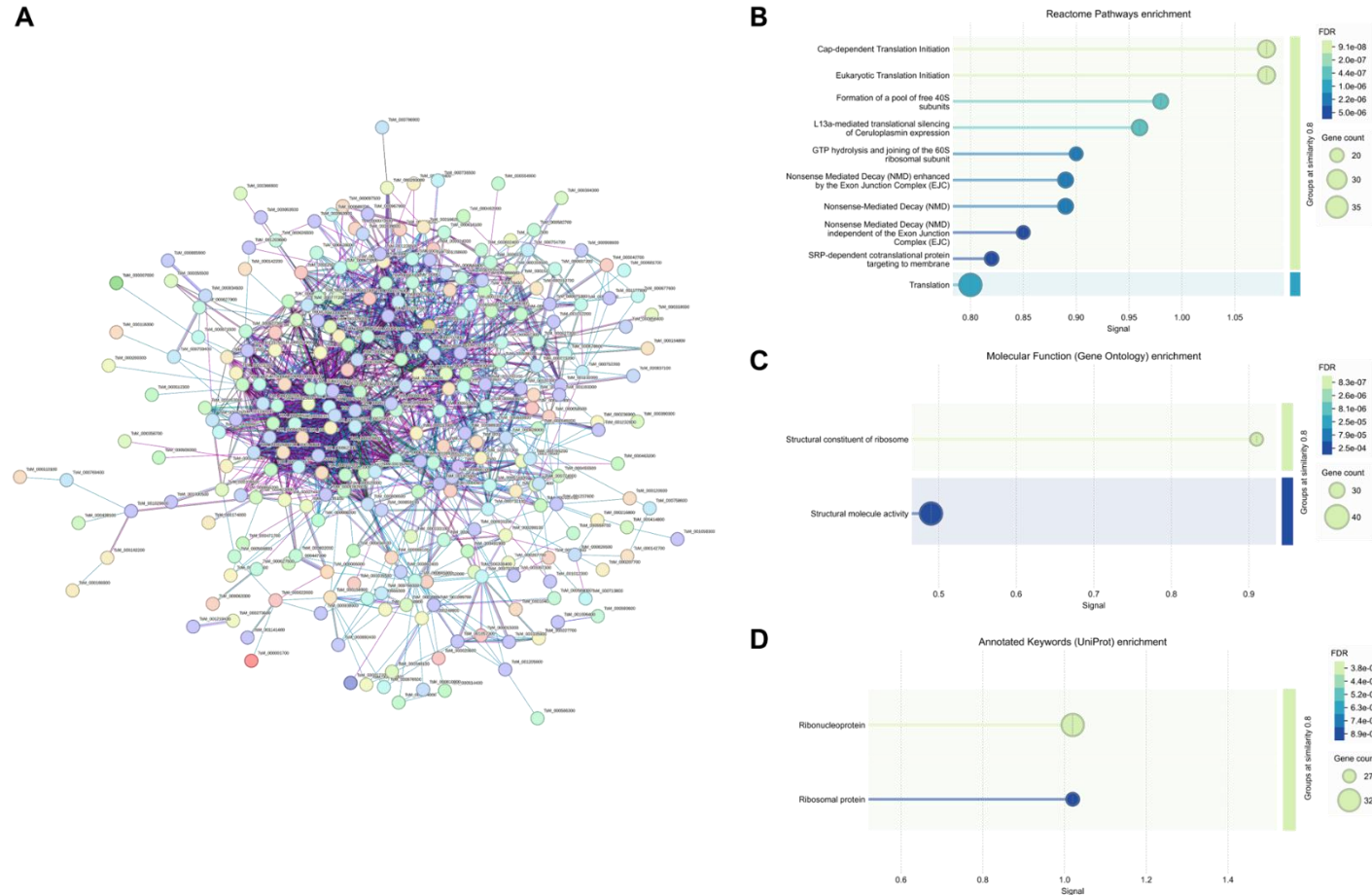


Figura 29. Análisis de interacción y asociación proteína-proteína mediante la base de datos de STRING y enriquecimiento de vías anotadas en Reactome para el módulo azul. (A) Red de asociación de proteínas del módulo azul ($n = 572$) representando asociación funcional o interacción física. Análisis de enriquecimiento de GO (función molecular) (B), vías Reactome (C) y palabras clave anotadas en UniProt (D), para las proteínas del módulo azul con interacción directa o indirecta.

8. Los módulos asociados al ácido taurocólico y la fase EV contienen genes que codifican factores de transcripción

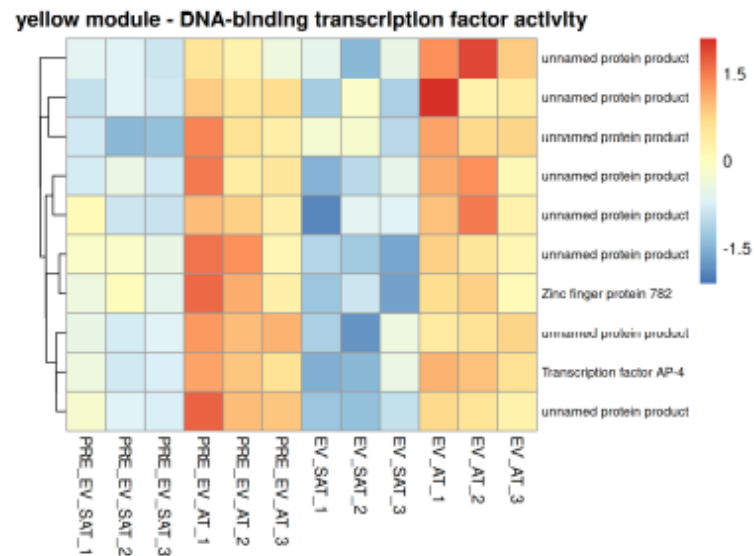
Los genes de los módulos amarillo (asociado a AT) y azul (asociado a la fase EV) por su alto grado de correlación a nivel de expresión génica, pueden estar coregulados por uno o varios factores de transcripción comunes. En el primer caso se trataría de factores de transcripción activados por efecto del AT y asociados a la vía MAPK, mientras que en el segundo se esperaría encontrar factores de transcripción que orquestan la expresión de genes asociados a la traducción y biogénesis ribosomal en el estadio EV.

Para determinar su presencia, se filtraron los genes asociados a las categorías funcionales dentro del dominio de función molecular “Actividad de factores de transcripción de unión a DNA, específico a RNA polimerasa-II” (GO:0000981) y “Actividad activadora de factores de transcripción de unión a DNA, específico a RNA polimerasa-II” (GO:0001228).

En el caso del módulo amarillo asociado a AT, se encontraron 8 secuencias sin descripción. Dentro de estas, “TsM_001194500” carece de dominios identificables. “TsM_000985500”, “TsM_000798900”, “TsM_000571400” poseen dominios tipo dedos de zinc C2H2. “TsM_001154800” posee dominios tipo POU. “TsM_001194300” corresponde a una descripción genérica de un factor de transcripción con dominio tipo dedos de zinc C2H2. “TsM_000067700” es una secuencia ortóloga a *cog-1* de *C. elegans*. Finalmente, “TsM_000643400” corresponde a una secuencia homóloga a TFAP-4 con dominios tipo bHLH (*basic Helix-Loop-Helix*) (Figura 30A).

En el caso del módulo azul asociado a EV, se encontraron los factores de transcripción SmadA y SmadC, la proteína que interactúa con MLX ortóloga de *mml-1* de *C. elegans*, la proteína homeobox tipo BarH-2, la secuencia ortóloga de *hrpk-1* de *C. elegans*, la proteína inducida por ecdisona 74EF, y el factor de transcripción nuclear NFY-A. Dentro de las secuencias sin descripción, los ortólogos de *C. elegans* *daf-19*, *dro-1*, *fkh-9* y *gldi-8* (Figura 30B).

A



B

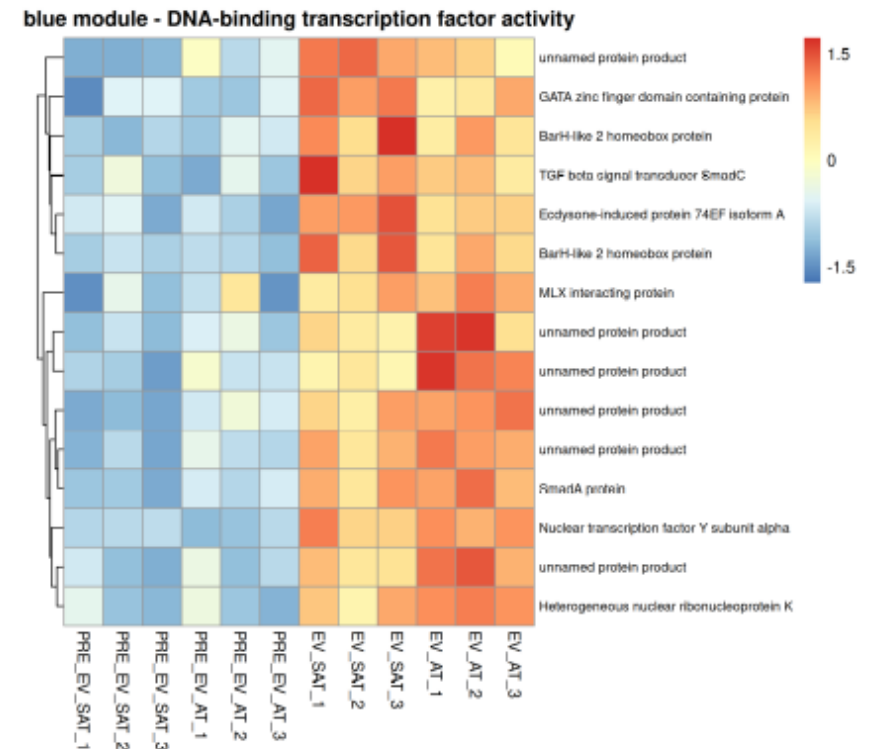


Figura 30. Mapas de calor de elementos de GO:0000981 y GO:0001228 pertenecientes al módulo amarillo asociado con AT (A) y al módulo azul asociado con la fase EV (B).

VI. DISCUSIONES

La evaginación del escólex es un evento esencial en el desarrollo del estadio adulto de *T. solium*. Los resultados obtenidos en este estudio permitieron caracterizar, por primera vez y a nivel transcriptómico, las diferencias entre cisticercos antes (PRE) y poco después de la evaginación del escólex (EV), y permitieron indagar el efecto molecular de las sales biliares (en el presente caso, el AT) en el proceso de desarrollo del parásito.

De acuerdo con los análisis realizados, resulta evidente que la evaginación del escólex no es solamente un proceso mecánico, sino que involucra una reprogramación molecular compleja que puede ser modulada por el AT. Para ello, herramientas para el análisis funcional e integral de los datos de expresión (como lo son el enriquecimiento funcional de términos GO (ORA y GSEA) y las redes de coexpresión génica (WGCNA) e interacción proteína-proteína), permitieron identificar rutas y procesos característicos de cada fase de desarrollo y también procesos inducidos por AT.

La presente discusión se centra en las implicaciones técnicas de los resultados obtenidos, los procesos biológicos característicos de cada fase de desarrollo y el papel del AT (Figura 31). Asimismo, se expande la información incluyendo genes candidatos para futuros ensayos a nivel funcional, que permitan identificar los elementos clave en el desarrollo hacia el estadio adulto, y redefinir a las sales biliares (concretamente, al AT) como un determinante programado (es decir, un elemento del hospedero aprovechado evolutivamente por el parásito como una señal

específica) para la maduración del parásito y no solo como un elemento adicional del ambiente dentro del hospedero.

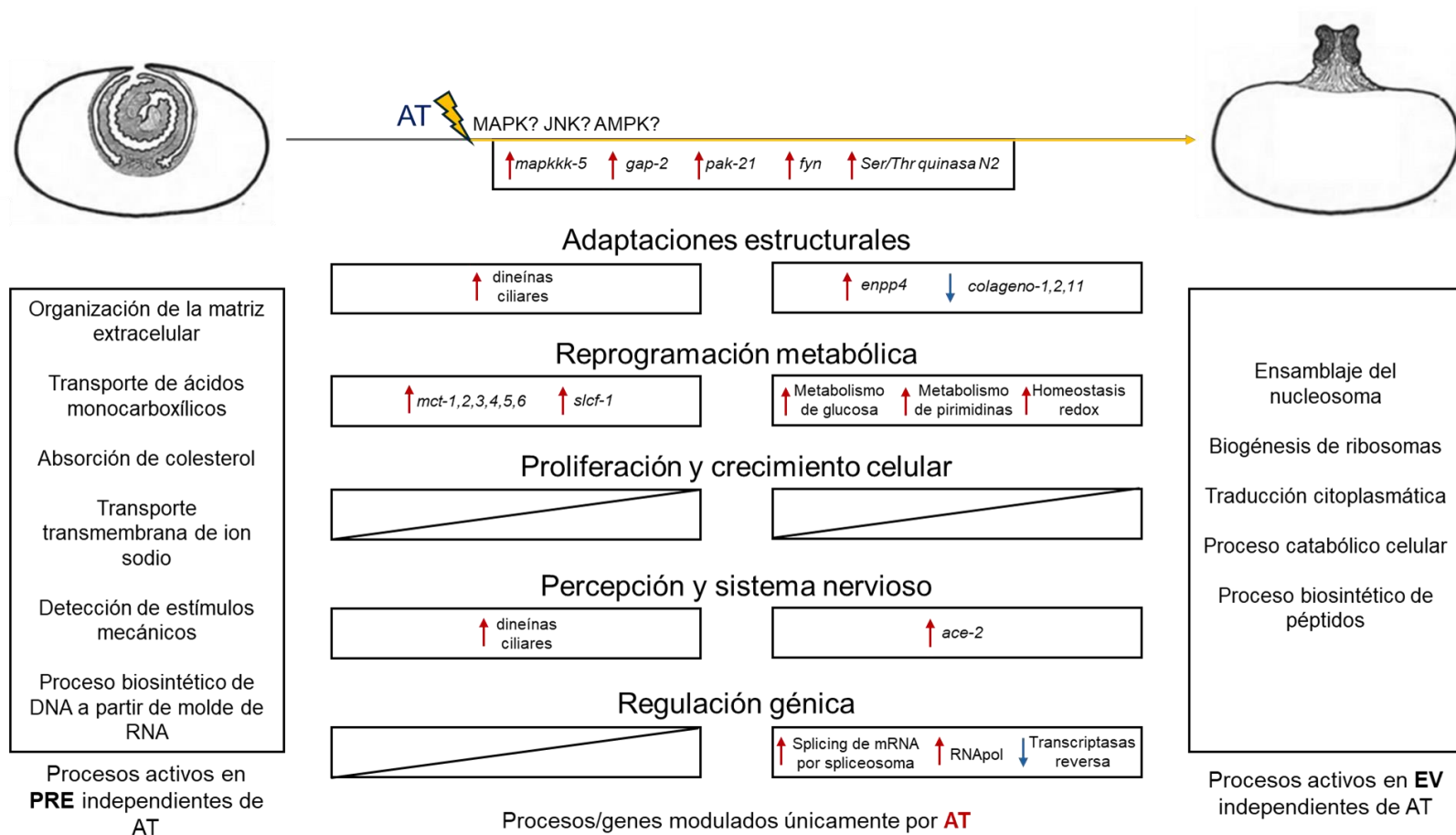


Figura 31. Resumen de los procesos activos en los grupos PRE y EV en presencia y ausencia de AT. La línea amarilla en la parte superior representa la activación y crecimiento inicial de *T. solium in vitro*, destacando el papel del AT. Se ilustran los procesos activos en cada grupo independientes de AT, así como también los procesos/genes modulados por AT en las cinco grandes categorías propuestas.

1. Consideraciones del cultivo y su efecto en el perfil transcripcional de *T. solium*

En el análisis exploratorio no-supervisado inicial, las únicas muestras que no tienen un agrupamiento consistente de acuerdo con la condición experimental son los POST (cisticercos en cultivo durante 120 horas, con el escólex evaginado) tanto en presencia como ausencia de AT.

La causa de esta heterogeneidad puede ser atribuida a que los organismos no están sincronizados en el tiempo de evaginación del escólex a pesar de iniciar el cultivo al mismo tiempo. Esto se refleja en que los dos outliers (un AT y un SAT) se agrupan con los EV SAT tanto en el PCA como en el agrupamiento jerárquico (Figura 9).

La pregunta por responder es cuál es el mínimo tiempo para apreciar una respuesta al inductor en el cultivo a nivel transcripcional. Es evidente la diferencia entre los individuos no cultivados y los PRE en PCA y agrupamiento jerarquizado, sin embargo, es aparente que 6 horas no es tiempo suficiente para inducir una respuesta fuerte por AT antes de la evaginación del escólex. Esto se refleja en la corta distancia entre las muestras PRE AT y PRE SAT por PCA y agrupamiento jerárquico. Sin embargo, este es tiempo suficiente en cultivo para encontrar diferencias transcripcionales con respecto a individuos EV.

En correlación al periodo de cultivo de individuos EV de 24 y 48 horas hasta la evaginación, en *E. granulosus* en cultivo por 24 horas con o sin bilis se pueden establecer cambios en el perfil transcripcional con respecto a las 0 horas (68).

Asimismo, otro reporte en el mismo parásito establece diferencias en cultivo con AT a las 12 horas de cultivo, aunque en este caso se realiza el cultivo en medio bifásico (52).

Se sugiere por ende, en caso de repetir el experimento, considerar el tiempo de evaginación del escólex como criterio de inclusión en el caso de evaluar cisticercos en cultivo por encima de las 48 horas, o en su defecto, incrementar el número de réplicas biológicas para captar de forma representativa réplicas atípicas y realizar análisis de expresión diferencial.

2. Evidencias de la regulación transcripcional durante la evaginación del escólex y la estimulación por AT

El análisis exploratorio inicial evidencia diferencias en los perfiles transcripcionales de cisticercos antes y después de evaginar el escólex. Adicionalmente, estas diferencias son más marcadas en presencia de AT (Figura 9). Bajo las características de este experimento, que se realizó con RNA extraído de organismos enteros, solo retirando la vesícula, es imposible determinar la expresión localizada de los transcritos. Sin embargo, la información disponible de ortólogos permite dilucidar los procesos activos en la transición de PRE a EV y el efecto del AT.

Dentro del conjunto de genes con mayor diferencia en su expresión al pasar de PRE a EV se encuentran el gen *enpp-4*, una fosfatasa alcalina que se sobreexpresa en presencia de AT; y *gp50*, que se reprime, independientemente de AT

Koziol y colaboradores evidenciaron la expresión localizada de la fosfatasa alcalina *alp-1/2* en las células del tegumento de la capa germinativa del quiste hidatídico y *alp-3* en el sistema excretor de los protoscólices, inducida por la activación del parásito *E. multilocularis* (24). La fosfatasa alcalina TsM_000911600, sobreexpresada en EV por efecto de AT, es un ortólogo de *alp-2* de *E. multilocularis*. ¿Qué función cumple *alp-2* en *T. solium*? Debido a su localización en *E. multilocularis*, se puede especular que también en *T. solium* se encuentre a nivel del tegumento del parásito. Su sobreexpresión en una región con capacidad proliferativa como lo es la capa germinativa del quiste hidatídico sugiere la formación de tegumento en una región con capacidad proliferativa en *T. solium*, como lo es el cuello (11,90). Es necesario estudiar los genes *alp* en *T. solium* a nivel genómico, como también anatómico y funcional (hibridación *in situ* de montaje completo, WMISH; caracterización bioquímica).

El gen *enpp-4* codifica a la ectonucleótido fosfodiesterasa pirofosfatasa 4, una enzima de membrana no estudiada en platelmintos involucrada en el metabolismo de nucleótidos trifosfato y diadenosina polifosfato (91). Su función en el desarrollo de céstodos es desconocida. Sin embargo, en modelos de biología del desarrollo de cordados, como *X. laevis*, se ha encontrado que su sobreexpresión en embriones está relacionada a la formación de tejidos renales ectópicos, por lo que tendría un rol en la formación de tejidos renales como un receptor (92). Es necesario estudiar la localización de estas proteínas en *T. solium* para conocer las implicancias de la expresión de este gen en la morfogénesis del tejido renal del parásito.

Finalmente, el antígeno GP50 es una glicoproteína de membrana anclada por GPI que es parte del antígeno reconocido por lectina de lenteja utilizada para el diagnóstico de cisticercosis (93). En este caso, se puede sugerir que no sería un blanco adecuado para el diagnóstico de teniasis dado que el presente trabajo encontró reducción en la expresión de la secuencia que codifica dicha proteína en EV.

Adicionalmente, el AT induce en PRE la expresión de un ortólogo de una proteína del tegumento anotada en *E. granulosus* y *E. multilocularis*.

Por otra parte, independientemente de AT, se observa represión del factor de transcripción TEF-3 en EV. Este es un enhancer (también conocido como TEAD) asociado a la proteína TAZ, la cual se inhibe al activar la vía de señalización Hippo. El complejo TEF/TAZ induce procesos como proliferación celular y transición epitelio-mesénquima (94). La función de los factores de transcripción TEF en platelmintos es aún desconocida. Sin embargo, en planarias, se ha evidenciado que el silenciamiento del gen *hippo* disminuye la tasa de apoptosis e incrementa la tasa de células entrando en el ciclo celular. A pesar de ello, las células se encontraron detenidas en fase M y adicionalmente se redujo el número de células diferenciadas (95). La disminución en su expresión podría ser un indicio de incremento en la actividad proliferativa propia del desarrollo y morfogénesis hacia el estadio adulto del parásito.

Adicionalmente, se detectó represión de Enolasa-C, la enzima de la vía glicolítica que cataliza la transformación de 2-fosfoglicerato a fosfoenolpiruvato. El ortólogo de esta secuencia en *S. mansoni* (Smp_024110) es un producto de trans-splicing (96). En *T. solium* se desconoce si esta secuencia sufre trans-splicing y si este mecanismo es responsable de la represión mediada por AT.

3. Los procesos biológicos asociados a la evaginación del escólex y el efecto del AT

El ORA permitió encontrar funciones sobrerrepresentadas en el conjunto de genes diferencialmente expresados y el GSEA permitió encontrar funciones sobrerrepresentadas cuyos genes poseen cambios transcripcionales sutiles y generalizados. WGCNA permitió agrupar genes de acuerdo con su perfil transcripcional cuyas funciones fueron validadas posteriormente en redes de interacción proteína-proteína. Cada una de estas herramientas ofrece información única que es corroborada entre sí, lo cual permite establecer un panorama claro de los cambios transcripcionales ocurridos en el parásito antes y después de la evaginación del escólex en presencia o ausencia de AT. Esto permite describir los procesos y rutas activas, además de identificar genes clave cuya función resulta crucial en condiciones fisiológicas para la transición hacia el estadio adulto de *T. solium*.

3.1. Adaptaciones estructurales: la formación de tegumento en PRE

A pesar de la ausencia de estudios en céstodos, se sabe que el tegumento es una estructura que actúa como captadora de nutrientes apoyada

principalmente sobre células musculares y componentes de la matriz extracelular, que componen una membrana basal (1).

Los cisticercos PRE, tanto en presencia como ausencia de ácido taurocólico, exhiben sobreexpresión de genes asociados al término GO “Organización de la matriz extracelular”, específicamente, Colágeno II tipo B, Colágeno I, Colágeno XI y heparán sulfato de la membrana basal. En EV, el AT modula negativamente la expresión de estos genes.

La sobreexpresión de estos elementos es un indicador del inicio del crecimiento del tegumento en cisticercos cuyo escólex aún no ha evaginado. La formación de esta estructura es un reflejo del crecimiento del organismo y su preparación para la captación de nutrientes después de la evaginación del escólex.

3.2. Reprogramación metabólica hacia la forma adulta

Dado que el cisticerco y el adulto se hospedan en tejidos con ambientes diferenciados (músculo esquelético e intestino, respectivamente) es razonable que la expresión de genes asociados al metabolismo esté sujeta a la variación en la disponibilidad de nutrientes. La caracterización transcriptómica de este trabajo permitió encontrar que el transporte de ácidos monocarboxílicos y absorción de colesterol son propios de la fase PRE; y el metabolismo de glucosa, nucleótidos y homeostasis redox de la fase EV.

3.2.1. La expresión de transportadores de monocarboxilato (MCT) en PRE

Dentro de los transportadores identificados, se encuentran secuencias ortólogas a los *mct-1,2,3,4,5,6* y *slcf-1* de *C. elegans*. Este último, al

ser reprimido, induce longevidad con una respuesta metabólica similar a la observada en restricción calórica. Este fenotipo se caracteriza por una alta producción de piruvato, es dependiente de la actividad de la piruvato deshidrogenasa y evidencia un aumento en el metabolismo mitocondrial reflejado en mayor producción de especies reactivas de oxígeno (97).

La represión de *slcf-1* en EV con respecto a PRE podría significar el inicio de un cambio en el programa metabólico del parásito caracterizado por un aumento en el metabolismo mitocondrial hacia el estadio adulto. Este depende del ciclo de ácidos tricarboxílicos o dismutación del malato, como se ha caracterizado anteriormente (1,16,19). Los cambios en la expresión de estos transportadores son potenciados por el AT. La relación entre AT y la expresión de MCT se desconoce.

3.2.2. La expresión de NPC1 y el transporte de colesterol en PRE

Niemann-Pick C1 (NPC1) es una proteína lisosomal que exporta colesterol del lisosoma a otras organelas. Un estudio previo sobre ortólogos de la proteína en secuencias de céstodos permitió identificar a TsM_000003800, una secuencia ortóloga a NPC1L1 de humanos. Esta es una secuencia con dominios similares a NPC1 que se expresa a nivel de membrana y es responsable de la absorción de colesterol en el intestino (48). En el presente trabajo se determinó que esta secuencia se encuentra sobreexpresada en el estadio PRE en comparación a EV. Adicionalmente, otras secuencias homólogas muestran el mismo perfil

transcripcional (TsM_000591700, TsM_000245600, TsM_000659300). Es necesario evaluar su expresión a nivel espacial para poder confirmar su función. Dado que los céstodos son incapaces de sintetizar colesterol de novo (19), resulta inesperada la disminución de su expresión posterior a la evaginación.

3.2.3. El AT como inductor metabólico en EV

En cuanto al metabolismo de glucosa, el AT induce la sobreexpresión en EV de las enzimas piruvato quinasa, fructosa-1,6-bifosfato aldolasa y galactosa-1-fosfato uridiltransferasa. En cuanto al metabolismo de nucleótidos, en EV se encuentra sobreexpresada la 3-ureidopropionasa. Esta enzima cataliza la degradación de pirimidinas y se localiza en células musculares estriadas de las paredes del nemátodo *C. elegans*, y se especula que puede permitir la síntesis de dipéptidos, como carnosina propios de tejidos musculares (98). De la misma forma, la NADP-dihidropirimidina deshidrogenasa y citidina desaminasa, enzimas de la biosíntesis y salvataje de pirimidinas.

Las enzimas tioredoxina-glutation reductasas son selenoproteínas que se encargan de mantener el balance redox de la célula. Esta enzima se encuentra sobreexpresada en EV únicamente en presencia de AT. En *E. granulosus* se ha comprobado que esta proteína se encuentra a nivel citosólico y mitocondrial. Ambas isoformas son derivadas de trans-splicing, donde solo una posee la secuencia señal mitocondrial (99). Se desconoce si el mismo mecanismo se da en *T. solium* y la secuencia del SL-RNA. La ausencia de secuencia señal mitocondrial permite

establecer que el gen anotado corresponde a la forma citosólica. Adicionalmente, otras enzimas relacionadas al balance redox se sobreexpresan en una forma hiperproliferativa del parásito, en comparación con cisticercos comunes (47).

¿A qué se debe la sobreexpresión de estos elementos en EV en presencia del AT? Se podría especular que el AT es una señal que indica la presencia de nutrientes, lo cual induce la sobreexpresión de enzimas asociadas al metabolismo de carbohidratos, vías que se encuentran íntegras en el genoma de *T. solium* (19). Adicionalmente, representaría la transición de un estadio quiescente (cisticerco) a un estadio metabólicamente activo (adulto). El aumento en el catabolismo celular requiere un control del balance redox, especialmente a nivel mitocondrial. Esto se reflejaría en la expresión de la tioredoxina-glutation reductasa. Esta relación causal, sin embargo, requiere ser probada. Por otra parte, el aumento en la expresión de enzimas del metabolismo de pirimidinas puede ser evidencia de un aumento en el requerimiento y reciclaje de nucleótidos, producto de proliferación celular.

3.3. Proliferación y crecimiento celular en EV

En el contexto del desarrollo al estadio adulto, los céstodos se valen de la proliferación celular para el crecimiento y morfogénesis. Especialmente, posterior a la evaginación y en el intestino del humano, *T. solium* genera miles de proglótides en el proceso de estrobilación

(incluyendo tegumento). Cada proglótide contiene aparatos reproductores masculinos y femeninos, así como también células germinales. La proliferación celular y síntesis de proteínas representarían un punto de partida para este proceso.

3.3.1. Proliferación de células somáticas y germinales en EV

Es conocido que los transcritos de histonas se expresan y acumulan en la fase S del ciclo celular en metazoos (100). Adicionalmente, *h2b* es un marcador de células germinativas en proliferación en *E. multilocularis* (25) e *H. diminuta* (11). En el presente trabajo se ha encontrado en el término GO “Ensamblaje del nucleosoma” sobreexpresión de parálogos de *h2b* en el estadio EV respecto a PRE. Esto resulta en evidencia para proponer un aumento en la tasa de proliferación celular en EV.

Por otra parte, TsM_000696800, es una secuencia no nombrada ortóloga a *asfl-1* y *unc-85* de *C. elegans* que se encuentra sobreexpresada en EV respecto a PRE. Mutantes en ambas secuencias en este nematodo son estériles. *asfl-1* cumple un rol importante en la replicación a nivel de líneas germinales, mientras que *unc-85* y *asfl-1* en conjunto regulan la replicación en células somáticas (101). Determinar la localización de este transcrito puede abrir una nueva ruta en la investigación de los mecanismos de proliferación celular, espermatogénesis y oogénesis de *T. solium*. Respecto a la categoría funcional “Transporte transmembrana de ion sodio”, el transportador de sodio-fosfato, TsM_000458000, es ortólogo de 6 secuencias en *C.*

elegans correspondientes a los genes *pitr-1,2,3,4,5,6*. Concretamente *pitr-1* se expresa en células germinales y es esencial para la producción de vitelogenina-2, necesario para la formación de embriones en este organismo (102).

3.3.2. La síntesis de proteínas en EV

En ausencia de AT, procesos favorecidos en EV con respecto a PRE son la biogénesis de ribosomas y la traducción en el citoplasma (Figura 20, 21). El WGCNA revela que el patrón de expresión de secuencias pertenecientes a categorías funcionales asociadas a la traducción sigue la misma tendencia tanto en AT como en SAT y guardan correlación con el estadio EV (Figura 17B, Figura 23F, Figura 23D).

El aumento en la expresión de proteínas ribosomales y factores de inicio/elongación de la traducción son indicio de mayor biogénesis ribosomal y podrían indicar mayor síntesis de proteínas. Esto sería necesario para el crecimiento del parásito al estadio adulto en EV. RNAseq de protoescolices de *E. granulosus* inducidos con AT por 12 y 24h evidencia enriquecimiento de la categoría funcional “Actividad catalítica” la cual contiene un número elevado de elementos asociados con la biogénesis ribosomal (52).

Recientemente, se ha reportado que el albendazol, un fármaco de benzimidazol usado para el tratamiento de infecciones por céstodos, tiene como blanco los microtúbulos que conectan los citones al tegumento distal de *M. corti*. La traducción se da a nivel de citones (evidenciada por la localización del RNA 18S) y las proteínas

sintetizadas viajan a través de microtúbulos al tegumento distal. Adicionalmente, el tratamiento con albendazol reprime la traducción en los citones independientemente de la respuesta de proteínas no-plegadas (UPR) (103). Esta evidencia podría sugerir la biogénesis de ribosomas en el desarrollo de nuevos citones. Sin embargo, la cinética ribosomal en céstodos se desconoce. Estudios de localización por WMISH podrían esclarecer si la biogénesis de ribosomas se da únicamente a nivel de los citones tegumentarios.

3.4. Relación con el entorno, sistema nervioso y señalización

La adaptación al entorno, en concreto el intestino del hospedero, es clave para el desarrollo del estadio adulto del parásito. En este sentido, la expresión de genes vinculados a diferenciación del sistema nervioso, respuesta sensorial y comunicación celular se asocian al desarrollo y fisiología del parásito.

En este trabajo concretamente, la diferenciación de neuronas glutamatérgicas, GABAérgicas y motoras podría estar ocurriendo tanto antes como durante y después de la evaginación del escólex. Por otro lado, la detección de movimiento y el movimiento en sí se estarían favoreciendo en la fase de desarrollo PRE.

Asimismo, respuestas celulares coordinadas a estímulos requieren sincronización que a su vez está comandada por cascadas de señalización. En este caso, el AT es un fuerte candidato a inductor de estas vías.

3.4.1. Transportadores iónicos asociados al sistema nervioso en PRE

De forma independiente a la adición de AT, en PRE se encuentra enriquecido el término “Transporte transmembrana de ion sodio”. Los cisticercos PRE expresan 2 transportadores iónicos de cloro-bicarbonato dirigidos por sodio, sodio-fosfato además de 2 receptores ionotrópicos de glutamato. En el caso de los transportadores cloro-bicarbonato dirigidos por sodio, TsM_000090300, es un ortólogo de *abts-1* de *C. elegans* sobreexpresado durante el desarrollo de sinapsis a nivel de sistema nervioso y la señalización inhibitoria GABA (104). El otro transportador cloro-bicarbonato dirigido por sodio, TsM_001143100, posee un ortólogo en *E. multilocularis* y *E. granulosus*, que no ha sido estudiado. Por otro lado, el receptor de glutamato, TsM_000940800, es ortólogo de *glr-6* en *C. elegans*, expresado en la neurona termosensorial RIA y es importante para la termotaxis en este nemátodo (105).

3.4.2. Las dineínas ciliares en el estadio PRE

Las dineínas ciliares, a diferencia de las dineínas intracelulares, son responsables del movimiento de cilios y flagelos. Son importantes para la motilidad y a nivel reproductivo. La categoría funcional “Determinación de simetría izquierda-derecha” está compuesta por secuencias que codifican elementos del complejo de dineínas ciliares. Estas se encuentran mayormente expresadas en PRE y el AT potencia su expresión en este estadio. No han sido estudiadas a profundidad en

platelmintos, sin embargo, el knockout de *Smed-dc2* en *Schmidtea mediterranea* induce un fenotipo de planarias con capacidad motora disminuida y ausencia de cilios (106). Se desconoce si las microtrichas de los céstodos están asociadas a esta clase de proteínas.

3.4.3. La sobreexpresión de PIEZO en PRE

Adicionalmente, como parte del término “Detección de estímulos mecánicos” existe sobreexpresión de genes *piezo-2*. Las proteínas de la familia PIEZO son canales iónicos activados por movimiento. En el caso de *C. elegans* se ha demostrado que permiten el movimiento faríngeo durante la alimentación (107). En céstodos, que carecen de un sistema digestivo, su función es desconocida. Probablemente, puedan estar asociados a neuronas sensoras con protrusiones dendríticas, positivas a tubulina acetilada, ubicadas en el tegumento conocidas como sencilla o bulbo sensorial (21). Es necesario ensayos de WMISH para determinar la localización precisa de estas proteínas PIEZO.

3.4.4. ACE2 como indicador del desarrollo de neuronas motoras en EV

Adicionalmente, de forma dependiente de AT, en EV se sobreexpresa una acetilcolinaesterasa ortóloga a *ace-2* de *C. elegans*. Esta se expresa en neuronas motoras principalmente (108). Esto refleja mayor diferenciación de neuronas motoras producto del crecimiento del parásito.

3.4.5. La expresión dependiente del AT de elementos de vías de señalización MAPK

La vía de señalización MAPK se encuentra conservada en metazoos y se ha estudiado en el contexto de transmisión de señales como por ejemplo citoquinas, factores de crecimiento y hormonas. En el presente trabajo se encontró sobreexpresión de *mapkkk-5* de manera dependiente a AT. El rol de esta proteína no se ha evaluado anteriormente en céstodos, sin embargo, un estudio en *E. granulosus* donde se inhibió el ortólogo a esta enzima presentó pérdida completa de células en la capa germinativa del quiste hidatídico, lo cual evidenciaría su función en células germinativas (109). En el caso del efecto del AT, se ha demostrado anteriormente en *Taenia crassiceps* cómo otras moléculas derivadas del colesterol (en este caso 17- β -estradiol) tienen un efecto en la activación de esta vía y la proliferación celular (110). Por otra parte, la progesterona induce la evaginación del escólex de *T. solium* (28), aunque se desconoce si de forma dependiente de la vía MAPK. Adicionalmente, se ha demostrado también que esta vía se encuentra activa en una forma hiperproliferativa del cisticerco de *T. solium* (111).

Otra secuencia asociada a esta vía sobreexpresada en presencia de AT es *gap-2*, una enzima que promueve la actividad GTPasa de Ras, inhibiendo la vía MAPK. En *C. elegans* existen al menos 9 isoformas con localización diferencial (112). En platelmintos no existen estudios a la fecha de la función de esta proteína.

Otras secuencias de la vía MAPK sobreexpresadas son *pak-21*, la secuencia que codifica a la serina-treonina quinasa N2 y la proteína quinasa Fyn. Su actividad es dependiente de la abundancia de

transcritos, por lo que son candidatos a estudios funcionales que verifiquen su localización y activación.

3.5. Regulación génica

Es plausible que la adaptación al entorno del hospedero y la transición entre estadios involucre la reprogramación molecular del parásito. Por ello, resulta importante el rol de la regulación génica a nivel pre y post-transcripcional. La regulación compleja y dinámica se caracteriza por la activación de elementos genéticos móviles en PRE. Adicionalmente, en EV la diversidad funcional del transcriptoma del parásito es regulada por corte y empalme alternativo, así como también a nivel epigenético.

3.5.1. Activación de elementos genéticos móviles en PRE

Los céstodos se caracterizan por la ausencia de los genes *piwi* y *vasa*, involucrados en la represión de la actividad de elementos genéticos móviles. Se desconoce cómo los céstodos reprimen la expresión de estos elementos para mantener su integridad genética (57). 26 secuencias sobreexpresadas en PRE poseen dominios transcriptasa reversa y 4 poseen secuencias de transcriptasa reversa derivado de retrotransposones. Debido a la pobre anotación del genoma de *T. solium* se desconoce cuántas de estas secuencias pueden encontrarse activas o corresponder a pseudogenes. Al visualizar su localización en el genoma, estos elementos se encuentran dentro de secuencias

anotadas como LTR-1 de la familia 4, flanqueadas por elementos LTR-1 de la familia 3.

Por otra parte, se han identificado retrotransposones en miniatura no autónomos denominados ta-TRIM propios de taenidos. En *T. solium* se han encontrado al menos 5 con secuencia completa. Sin embargo, la calidad fragmentada del genoma disponible impide identificarlos por completo. Estos elementos se encuentran expresados en gran magnitud en estadios tempranos del desarrollo de protoescólices de *E. multilocularis* y son característicos de células germinativas, el linaje totipotente de los céstodos (113). La elevada expresión y actividad de estos elementos, reflejada en la sobreexpresión de transcriptasas reversas, podría ser un indicio de un incremento en la población de células germinativas en individuos PRE. Sin embargo, hace falta identificar las células en las cuales estas enzimas se expresan para verificar esta afirmación. Las secuencias obtenidas en este proyecto podrían ser útiles para terminar de identificar los elementos ta-TRIM de *T. solium*.

Inesperadamente, la expresión de estos genes es reprimida en individuos EV, especialmente por efecto del AT. La relación entre la represión de la expresión de transcriptasas reversas y el AT se desconoce.

3.5.2. Regulación epigenética en el estadio EV

La categoría funcional “Ensamblaje del nucleosoma”, muestra la presencia de un ortólogo de *hat-1* (TsM_000505900) y la detección de

una secuencia con bromodominio unido a dedos de zinc (ambas asociadas a remodelamiento de la cromatina) evidencian regulación epigenética de la expresión de genes en el estadio EV.

3.5.3. Corte y empalme alternativo en EV

Transcritos asociados a la maduración del RNA y splicing se encuentran sobreexpresados en EV con respecto a PRE. Este efecto es potenciado por el AT. En este caso, las diferencias en la expresión de estos genes entre EV y PRE en presencia de AT son más pronunciadas que en ausencia de AT. Estas secuencias en su mayoría corresponden a proteínas Sm de unión a snRNAs U1, U2, U4, U5 y U6, factores de splicing, la poli(A) polimerasa alfa y las subunidades de RNA polimerasa RPB2,3,4,11 y RPABC5. Estas últimas evidencian un incremento en la expresión de genes en EV con respecto a PRE (y más pronunciado por efecto de AT).

Por otra parte, la expresión de factores de splicing y proteínas de unión a snRNAs evidencian un posible cambio de isoformas hacia el desarrollo al estadio adulto con respecto al cisticerco antes de evaginar el escólex. Este fenómeno se da por ejemplo en lactato deshidrogenasa en *H. microstoma* al comparar estadios adulto y oncosfera (16). También se ha evidenciado en *E. granulosus* donde el cultivo en 12, 24h de protoescólices en AT induce splicing alternativo en múltiples transcritos (52). La relación entre AT y splicing se desconoce. Adicionalmente, el genoma de referencia disponible en WormBase ParaSite carece de anotaciones de isoformas (19). Los datos de

secuenciación de este trabajo podrían ser utilizados para descubrir nuevas isoformas y evaluar splicing alternativo.

3.5.4. Reguladores transcripcionales maestros asociados al AT y la fase EV

El análisis de WGCNA permitió determinar conjuntos de genes y funciones con posible regulación conjunta. Sin embargo, los enriquecimientos funcionales mediante ORA y STRING no permitieron capturar secuencias codificantes de factores de transcripción. Probablemente, dado que este análisis toma en cuenta el RNA proveniente de organismos completos, la expresión local de factores de transcripción, en regiones anatómicas determinadas, podría estar enmascarada. La identificación de estos permitiría determinar reguladores maestros transcripcionales que puedan explicar los perfiles de expresión correlativos en cuanto al conjunto de genes de la vía MAPK asociados al AT, y al conjunto de genes asociados a la biogénesis ribosomal y traducción del módulo correlacionado con la fase EV.

En cuanto a factores de transcripción asociados al AT, los únicos identificados resultan TFAP-4 y la secuencia ortóloga a *cog-1*. TFAP-4 es un factor de transcripción estudiado principalmente en el contexto del cáncer que reprime la expresión del inhibidor de quinasas dependientes de ciclina, p21, de forma dependiente de c-Myc. Por ende, la actividad de ambas proteínas en conjunto actúa como

inductores de proliferación celular (114). Por otro lado, en asociación con la proteína Snail, TFAP-4 activa la transición epitelio-mesénquima, un proceso típico del desarrollo embrionario y que en planarias está asociado a la regeneración (115, 116). Por otro lado, en *C. elegans cog-1* es requerido para la formación de órganos del aparato reproductor, en particular la vulva (117). Se desconoce si su expresión depende de la vía MAPK. Sin embargo, esta vía se asocia al desarrollo de dicho órgano (118).

En cuanto a los factores de transcripción identificados en el conjunto de genes asociados a la fase EV, resultan interesantes SmadA/C, la secuencia ortóloga a *mml-1* y *nfya-1/2*. SmadA/C son factores de transcripción asociados a la vía de señalización TGF- β . En *T. solium* y *T. crassiceps* se identificaron dos receptores de esta molécula. Su adición *in vitro* resulta en un incremento en el tamaño del cisticerco de *T. crassiceps* y mejora la viabilidad de cisticercos de *T. solium* en cultivo (119). Por otra parte, recientemente se vio que esta vía de señalización permite la replicación *in vitro* de células germinativas de *T. solium* (120). En *E. multilocularis*, esta vía se asocia a la formación de protoescólices (26) y, en cestodos en general, a estrobilación (29). Estos resultados guardan relación con marcadores moleculares de proliferación celular activos en la fase EV.

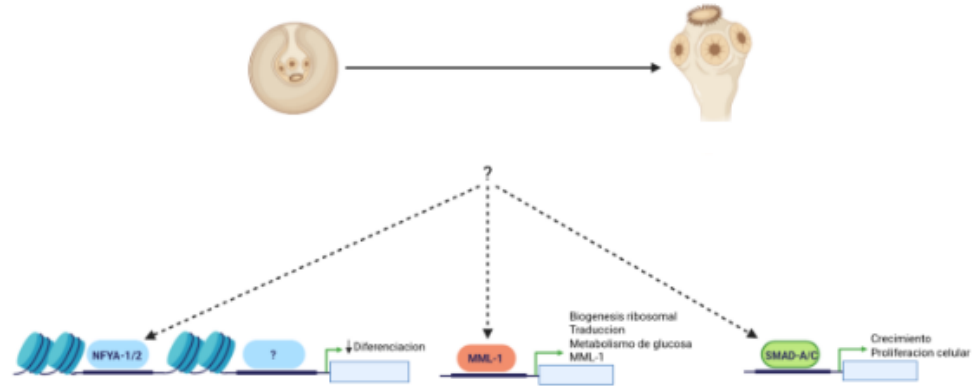
Por otra parte, *nfya-1/2* en *C. elegans* es un sustrato de la vía MAPK-ERK, y su fosforilación inhibe el desarrollo de la vulva en este organismo al reducir su afinidad por el DNA (121). Adicionalmente, su expresión se asocia a la diferenciación de un tipo de neuronas sensoriales-motoras (122). Fuera de sus funciones en el desarrollo de *C. elegans*, se ha caracterizado su función como factor de transcripción pionero al facilitar el acceso de reguladores maestros a la cromatina. Este es esencial en el mantenimiento del estado de células madre embrionarias al permitir la unión de los factores Oct-4, Sox2, Nanog y Prdm14, característicos de este linaje celular (123).

Finalmente, MML-1 es una proteína de interacción con MLX, parte de la superfamilia de factores de transcripción Myc. Los platelmintos, al igual que los nemátodos, son los únicos metazoos que carecen de ortólogos de Myc (124). En *C. elegans* el complejo MondoA (*mml-1*)/MLX cumple funciones análogas a las que cumple el complejo Myc/Max en la mayoría de metazoos como regulador maestro del metabolismo de carbohidratos (125, 126). En *C. elegans*, la pérdida de *mml-1* resulta en reducción de la longevidad y tanto su transcripción como localización a nivel nuclear se asocia a menor actividad de la vía de señalización mediada por proteínas tipo-insulina o a ausencia de nutrientes (127). En otros organismos, la función de Myc se asocia a mayor biogénesis ribosomal, síntesis de proteínas y modulación de la actividad de las RNA polimerasas I y III (128, 129).

Estas aproximaciones en otros organismos sugieren un panorama del proceso de evaginación del escólex de *T. solium* y el efecto del AT (Figura 32A). En ausencia de AT, la expresión de *nfya-1/2*, *mml-1* y *smad-A/C* inducen la evaginación del escólex al promover el crecimiento celular y la síntesis y traducción de proteínas. El mecanismo por el cual se estimula la expresión de estos reguladores es desconocido.

En presencia de AT, se incrementa la actividad de la vía MAPK, la cual modula negativamente a NFYA-1/2. Esto favorece la diferenciación de algunos linajes celulares imposibles de determinar mediante la aproximación experimental de este estudio hecho con RNA proveniente de organismos completos. La vía MAPK a su vez modula la expresión de MML-1, lo cual favorece la síntesis y traducción de proteínas y el metabolismo de glucosa. El AT a su vez modula mediante un mecanismo desconocido la expresión de TFAP-4, que en conjunto con MML-1, reprimen la expresión de P21 favoreciendo en mayor medida la proliferación celular. SMAD-A/C promueve el crecimiento y proliferación celular de modo independiente a AT (Figura 32B).

A



B

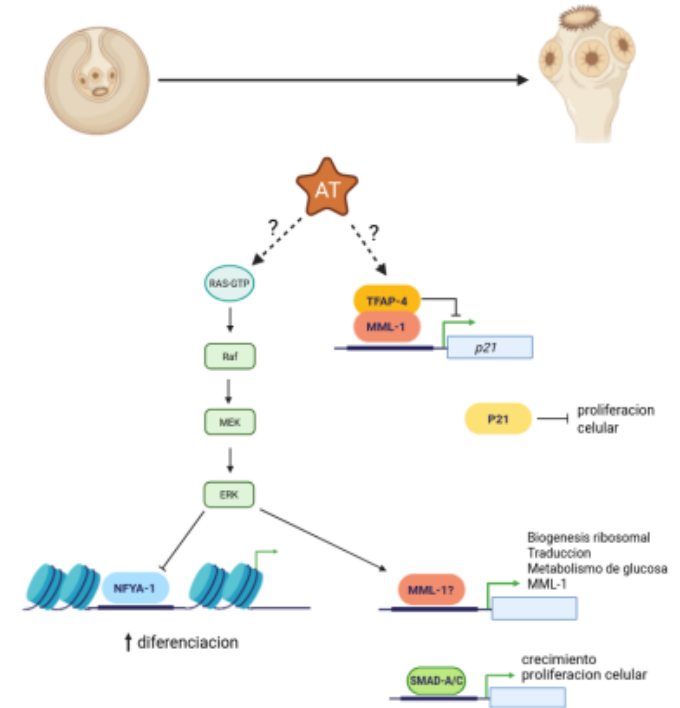


Figura 32. Mecanismo propuesto de inducción de la evaginación del escólex de *T. solium* en ausencia (A) y presencia (B) de ácido taurocólico. En este último se muestra la vía MAPK representada por RAS-GTP, Raf, MEK, ERK.

VII. LIMITACIONES Y RECOMENDACIONES

El diseño experimental del presente trabajo está sujeto a que hasta el momento no existe una descripción morfológica de la larva metacéstode y su desarrollo hacia el estadio adulto no está estudiado. Debido a ello, el criterio de inclusión de las muestras se basó en el tiempo de cultivo y el estado de evaginación del escólex. Se sugiere por ende evaluar marcadores morfológicos o moleculares que permitan definir el grupo POST de manera más precisa, para lo cual se requieren nuevos cultivos, sistemáticos, de etapas posteriores a las 48 h.

Adicionalmente, se utilizaron individuos completos para la extracción de RNA. Esto puede enmascarar la expresión posicional de algunos genes (particularmente factores de transcripción) asociados a ejes del desarrollo o tejidos altamente especializados. Para evaluar su expresión localizada, se sugiere utilizar WMISH o realizar RT-qPCR de regiones específicas del parásito (en el caso de cisticercos con el escólex ya evaginado; antes de esto, no se pueden seccionar las estructuras en el interior de la vesícula).

No existen nuevos genomas anotados de *T. solium* desde el año 2013, y este último es bastante fragmentado (reflejado en un N50 de 68.0 kb) (15). No existen isoformas anotadas para las secuencias identificadas desde la publicación de dicho genoma. La información obtenida del transcriptoma analizado en este trabajo se puede beneficiar y enriquecer con nuevas anotaciones basadas en genomas de alta calidad y transcriptómica de lecturas largas mediante tecnologías de secuenciación de tercera generación (PacBio, Nanopore).

La información transcriptómica obtenida puede sugerir, mas no comprobar, la actividad de algunas vías de señalización, por ejemplo MAPK. Es necesaria una comprobación funcional del nivel de actividad de la vía inducida por AT. Por ejemplo, es posible evaluar los sustratos fosforilados en la cascada mediante western blot, comparando fases de evaginación y presencia del inductor. Complementar la información obtenida con un análisis proteómico puede dar una vista general mucho más detallada y precisa.

Adicionalmente, los genes identificados en el presente trabajo son fuertes candidatos a estudios de localización por WMISH. Esto permitirá comprender su función y rol en la morfogénesis de los distintos tejidos del parásito. La validación de la expresión de estos genes debe realizarse mediante RT-qPCR.

VIII. CONCLUSIONES

1. Mediante secuenciación por RNAseq se identificaron aspectos clave de la evaginación del escólex de *T. solium* y el efecto del AT.
2. Los individuos PRE se caracterizan por sobreexpresar genes asociados a formación de tegumento, desarrollo del sistema nervioso, movimiento y detección de movimiento, retrotransposones, reprogramación del metabolismo y absorción de colesterol. Los individuos EV se caracterizan por expresar genes del metabolismo de glucosa, balance redox, metabolismo de nucleótidos, diferenciación de neuronas motoras, proliferación celular, splicing, biogénesis ribosomal y traducción. Estos procesos estarían dirigidos por los factores de transcripción MML-1, SmadA/C, NFYA-1/2.
3. El AT potencia la expresión transportadores de la ácidos monocarboxílicos y dineínas ciliares asociadas a movimiento en PRE. Asimismo, en EV potencia la expresión de genes asociados a splicing, metabolismo de glucosa, nucleótidos y balance redox. Además, reprime la expresión de transcriptasas reversas en esta fase de desarrollo. Por otra parte, tanto en PRE como en EV, promueve la expresión del represor TFAP-4, asociado a proliferación celular.
4. El efecto del AT es evidente en PRE y EV en la expresión de genes asociados a vías de señalización MAPK/ERK y MAPK/JNK. Demostrar que el AT induce estas vías requiere de una validación funcional.
5. En cada gran categoría donde se apreciaron las diferencias descritas, se encontraron genes candidatos a futuros estudios por qPCR o WMISH:
 - En adaptaciones estructurales: *alp-2*, *enpp-4*; *colágeno-1,2,11* y *heparan sulfato*; *mct-1-6* y *slcf-1*.

- En reprogramación metabólica: *npc1l-1*; *enolasa-c*, *piruvato quinasa*, *fructosa-1,6-bifosfato aldolasa*, *galactosa-1-fosfato uridiltransferasa*, *3-ureidopropionasa*, *NADPH-dihidropirimidina deshidrogenasa*, *citidina deaminasa*, *tioredoxina glutathione reductasa*.
- En proliferación y crecimiento celular: *h2b*, *asf1-1*, *unc-85*, *pitr-1-6*, genes codificantes de proteínas ribosomales y factores de traducción.
- En percepción y sistema nervioso: *abts-1*, *glr-6*; *piezo-2*, dineínas ciliares, *ace-2*.
- En vías de señalización moduladas por AT: *mapkkk-5*, *gap-2*, *pak-21*, *Ser-Thr quinasa N2*, *fyn*, *tef-3*.
- En regulación génica: genes codificantes de factores de splicing, transcriptasas reversa y RNA polimerasas; los factores de transcripción *mml-1*, *nfya-1/2*, *smadA/C*, *tfap-4*.

IX. REFERENCIAS BIBLIOGRÁFICAS

1. Smyth JD, McManus DP. The Physiology and Biochemistry of Cestodes. Cambridge: Cambridge University Press; 1989.
2. García HH, Gonzalez AE, Evans CA, Gilman RH. *Taenia solium* cysticercosis. The Lancet. 2003;362(9383):547-56.
3. Donadeu, M. et al. WHO *Taenia Solium* Endemicity Map – 2022 Update. <https://www.who.int/publications/i/item/who-wer9717-169-172> (2022).
4. Hossain S, Shafqat Shabir, Toye PG, Thomas LF, Falcone FH. Insights into the diagnosis, vaccines, and control of *Taenia solium*, a zoonotic, neglected parasite. Parasites & Vectors. 2023;16(1).
5. WHO Estimates of the Global Burden of Foodborne Diseases: Foodborne Disease Burden Epidemiology Reference Group 2007-2015. <https://iris.who.int/handle/10665/199350> (World Health Organization, Geneva, 2015).
6. Garcia HH, Nash TE, Del Brutto OH. Clinical symptoms, diagnosis, and treatment of neurocysticercosis. The Lancet Neurology. 2014;13(12):1202–15.
7. Stevenson P. Observations on the hatching and activation of fresh *Taenia saginata* eggs. Ann Trop Med Parasitol. 1983;77(4):399-404.
8. Conn DB, Świdorski Z. A standardised terminology of the embryonic envelopes and associated developmental stages of tapeworms (Platyhelminthes: Cestoda). Folia Parasitol (Praha). 2008;55(1):42-52.

9. Sciutto E, Fragoso G, Fleury A, Laclette JP, Sotelo J, Aluja A, et al. Taenia solium disease in humans and pigs: an ancient parasitosis disease rooted in developing countries and emerging as a major health problem of global dimensions. Microbes Infect. 2000;2(15):1875-90.
10. Palma S, Chile N, Carmen-Orozco RP, Trompeter G, Fishbeck K, Cooper V, et al. *In vitro* model of postoncosphere development, and in vivo infection abilities of Taenia solium and Taenia saginata. Dalton JP, editor. PLoS Negl Trop Dis. 2019;13(3):e0007261.
11. Rozario T, Quinn EB, Wang J, Davis RE, Newmark PA. Region-specific regulation of stem cell-driven regeneration in tapeworms. eLife. 2019;8:e48958.
12. Rabiela M. Evagination of Taenia solium Cysticerci A Histologic and Electron Microscopy Study. Arch Med Res. 2000;31(6):605-7.
13. Šlais J. The Morphology and Pathogenicity of the Bladder Worms. Dordrecht: Springer Netherlands; 1970.
14. Basika T, Paludo GP, Araujo FM, Salim AC, Pais F, Maldonado L, et al. Transcriptomic profile of two developmental stages of the cestode parasite Mesocostoides corti. Mol Biochem Parasitol. 2019;229:35-46.
15. Camargo De Lima J, Monteiro KM, Basika Cabrera TN, Paludo GP, Moura H, Barr JR, et al. Comparative proteomics of the larval and adult stages of the model cestode parasite Mesocostoides corti. J Proteomics. 2018;175:127-35.

- 16.Preza M, Calvelo J, Langleib M, Hoffmann F, Castillo E, Koziol U, et al. Stage-specific transcriptomic analysis of the model cestode *Hymenolepis microstoma*. *Genomics*. 2021;113(2):620-32.
- 17.Castaneda-Carpio D, Gutierrez-Loli R, Maravi-Jaime J, Del Aguila SW, Villar-Davila V, Moyano LM, et al. Transcriptome of *Taenia solium* during *in vitro* cyst activation and initial growth into the tapeworm stage. *Sci Data*. 2025;12(1):808.
- 18.Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Disponible en: www.bioinformatics.babraham.ac.uk/projects/fastqc/
- 19.Tsai IJ, Zarowiecki M, Holroyd N, Garcarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 2013;496(7443):57-63.
- 20.Blume La Torre J. La inhibición de la gamma secretasa altera la expresión de genes de la vía de señalización Notch en el cisticerco de *Taenia solium* y reprime la evaginación del escólex [Tesis de Maestría]. Universidad Peruana Cayetano Heredia; 2020. Disponible en: <https://repositorio.upch.edu.pe/handle/20.500.12866/8940>
21. Kamenetzky L, Maldonado LL, Cucher MA. Cestodes in the genomic era. *Parasitology research*. 2021;121(4):1077–89.
- 22.Rozario T, Newmark PA. A confocal microscopy-based atlas of tissue architecture in the tapeworm *Hymenolepis diminuta*. *Exp Parasitol*. 2015;158:31-41.

23. Galindo M, Paredes R, Marchant C, Miño V, Galanti N. Regionalization of DNA and protein synthesis in developing stages of the parasitic platyhelminth *Echinococcus granulosus*. J Cell Biochem. 2003;90(2):294-303.
24. Koziol U, Rauschendorfer T, Zanon Rodríguez L, Krohne G, Brehm K. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*. EvoDevo. 2014;5(1):10.
25. Koziol U, Jarero F, Olson PD, Brehm K. Comparative analysis of Wnt expression identifies a highly conserved developmental transition in flatworms. BMC Biol. 2016;14(1).
26. Kaethner M, Epping K, Bernthaler P, Rudolf K, Thomann I, Leitschuh N, et al. Transforming growth factor- β signalling regulates protoscolex formation in the *Echinococcus multilocularis* metacestode. Front Cell Infect Microbiol. 2023;13.
27. Rothman AH. Studies on the excystment of tapeworms. Exp Parasitol. 1959;8(4):336-64.
28. Escobedo G, Camacho-Arroyo I, Hernández-Hernández OT, Ostoa-Saloma P, García-Varela M, Morales-Montor J. Progesterone Induces Scolex Evagination of the Human Parasite *Taenia solium*: Evolutionary Implications to the Host-Parasite Relationship. J Biomed Biotechnol. 2010;2010:1-10.
29. Jarero F, Baillie A, Riddiford N, Montagne J, Koziol U, Olson PD. Muscular remodeling and anteroposterior patterning during tapeworm segmentation. Dev Dyn. 2024;253(11):998-1023.

30. Paludo GP, Thompson CE, Miyamoto KN, Guedes RLM, Zaha A, De Vasconcelos ATR, et al. Cestode strobilation: prediction of developmental genes and pathways. *BMC Genomics*. 2020;21(1).
31. Tielens AGM, Hellemond JJV. Unusual aspects of metabolism in flatworm parasites. En: Maule AG, Marks NJ, editores. *Parasitic flatworms: molecular biology, biochemistry, immunology and physiology*. 1.^a ed. UK: CABI; 2006. p. 387-407.
32. Amahong K, Yan M, Li J, Yang N, Liu H, Bi X, et al. EgGLUT1 Is Crucial for the Viability of *Echinococcus granulosus sensu stricto* Metacestode: A New Therapeutic Target? *Front Cell Infect Microbiol*. 2021;11:747739.
33. Rodríguez-Contreras D, Skelly PJ, Landa A, Shoemaker CB, Laclette JP. Molecular and functional characterization and tissue localization of 2 glucose transporter homologues (TGTP1 and TGTP2) from the tapeworm *Taenia solium*. *Parasitology*. 1998;117(6):579-88.
34. Rosen R, Uglem GL. Localization of facilitated diffusion and active glucose transport in cysticercoids of *Hymenolepis diminuta* (Cestoda). *Int J Parasitol*. 1988;18(5):581-4.
35. Castañeda Carpio D. Cambios en la expresión de transportadores de glucosa de *Taenia solium* asociados al proceso de evaginación [Tesis de Licenciatura]. Universidad Peruana Cayetano Heredia; 2021.
36. Saidur Rahman M, Mettrick DF. Carbohydrate intermediary metabolism in *Hymenolepis microstoma* (Cestoda). *Int J Parasitol*. 1982;12(2-3):155-62.

37. Parkinson J, Wasmuth JD, Salinas G, Bizarro CV, Sanford C, Berriman M, et al. A Transcriptomic Analysis of *Echinococcus granulosus* Larval Stages: Implications for Parasite Biology and Host Adaptation. Jones MK, editor. PLoS Negl Trop Dis. 2012;6(11):e1897.
38. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science. 2009;324(5930):1029-33.
39. Tennessen JM, Baker KD, Lam G, Evans J, Thummel CS. The Drosophila Estrogen-Related Receptor Directs a Metabolic Switch that Supports Developmental Growth. Cell Metab. 2011;13(2):139-48.
40. Agathocleous M, Love NK, Randlett O, Harris JJ, Liu J, Murray AJ, et al. Metabolic differentiation in the embryonic retina. Nat Cell Biol. 2012;14(8):859-64.
41. Bennet EM, Heath PA, Bryant C. The effects of changes in the definitive host environment on the metabolism of *Hymenolepis diminuta* during growth and maturation. Int J Parasitol. 1993;23(1):57-68.
42. Frayha GJ. Comparative metabolism of acetate in the taeniid tapeworms *Echinococcus granulosus*, *E. multilocularis* and *Taenia hydatigena*. Comp Biochem Physiol Part B Comp Biochem. 1971;39(1):167-70.
43. Alvite G, Esteves A. Lipid binding proteins from parasitic platyhelminthes. Front Physiol. 2012;3.

- 44.Lee E, Kim S, Bae Y, Chung J, Suh M, Na B, et al. A hydrophobic ligand-binding protein of the *Taenia solium* metacestode mediates uptake of the host lipid: Implication for the maintenance of parasitic cellular homeostasis. *PROTEOMICS*. 2007;7(21):4016-30.
- 45.Illescas O, Carrero JC, Bobes RJ, Flisser A, Rosas G, Laclette JP. Molecular characterization, functional expression, tissue localization and protective potential of a *Taenia solium* fatty acid-binding protein. *Mol Biochem Parasitol*. 2012;186(2):117-25.
- 46.Montero E, Gonzalez LM, Bonay P, Rosas G, Hernandez B, Sciutto E, et al. *Taenia solium*: Identification and preliminary characterization of a lipid binding protein with homology to the SEC14 catalytic domain. *Exp Parasitol*. 2007;116(3):191-200.
- 47.Orrego MA, Szczesniak MW, Vasquez CM, Verastegui MR, Bustos JA, Garcia HH, et al. Transcriptomic analysis of subarachnoid cysts of *Taenia solium* reveals mechanisms for uncontrolled proliferation and adaptations to the microenvironment. *Sci Rep*. 2024;14(1):11833.
- 48.Kulakowski Corá R, Prado Paludo G, Andrade Paes J, Bunselmeyer Ferreira H. In silico comparative analysis of cestode and human NPC1 provides insights for ezetimibe repurposing to visceral cestodiasis treatment. *Sci Rep*. 2024;14(1).
- 49.Surgan MH, Roberts LS. Adsorption of Bile Salts by the Cestodes, *Hymenolepis diminuta* and *H. microstoma*. *J Parasitol*. 1976;62(1):78.

50. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, et al. The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat Genet*. 2013;45(10):1168-75.
51. Sargan MH, Roberts LS. Effect of Bile Salts on the Absorption of Glucose and Oleic Acid by the Cestodes, *Hymenolepis diminuta* and *H. microstoma*. *J Parasitol*. 1976;62(1):87.
52. Mohammadi MA, Harandi MF, McManus DP, Mansouri M. Genome-wide transcriptome analysis of the early developmental stages of *Echinococcus granulosus* protoscoleces reveals extensive alternative splicing events in the spliceosome pathway. *Parasit Vectors*. 2021;14(1).
53. Regev A, Lamb MJ, Jablonka E. The Role of DNA Methylation in Invertebrates: Developmental Regulation or Genome Defense? *Mol Biol Evol*. 1998;15(7):880-91.
54. Rosado Fantappiè M, Rodrigues Pereira Gimba E, Rumjanek FD. Lack of DNA Methylation in *Schistosoma mansoni*. *Exp Parasitol*. 2001;98(3):162-6.
55. Geyer KK, Chalmers IW, MacKintosh N, Hirst JE, Geoghegan R, Badets M, et al. Cytosine methylation is a conserved epigenetic feature found throughout the phylum Platyhelminthes. *BMC Genomics*. 2013;14(1).
56. Aalto AP, Pasquinelli AE. Small non-coding RNAs mount a silent revolution in gene expression. *Curr Opin Cell Biol*. 2012;24(3):333-40.
57. Fontenla S, Rinaldi G, Smircich P, Tort JF. Conservation and diversification of small RNA pathways within flatworms. *BMC Evol Biol*. 2017;17(1).

58. Blumenthal T. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* 1995;11(4):132-6.
59. Rajkovic A, Davis RE, Simonsen JN, Rottman FM. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci.* 1990;87(22):8879-83.
60. Davis RE, Singh H, Botka C, Hardwick C, Ashraf el Meanawy M, Villanueva J. RNA trans-splicing in *Fasciola hepatica*. Identification of a spliced leader (SL) RNA and SL sequences on mRNAs. *J Biol Chem.* 1994;269(31):20026-30.
61. Brehm K, Jensen K, Frosch M. mRNA Trans-splicing in the Human Parasitic Cestode *Echinococcus multilocularis*. *J Biol Chem.* 2000;275(49):38311-8.
62. Davis RE. Surprising diversity and distribution of spliced leader RNAs in flatworms. *Mol Biochem Parasitol.* 1997;87(1):29-48.
63. Pajuelo MJ, Eguiluz M, Dahlstrom E, Requena D, Guzmán F, Ramirez M, et al. Identification and Characterization of Microsatellite Markers Derived from the Whole Genome Analysis of *Taenia solium*. Brehm K, editor. *PLoS Negl Trop Dis.* 2015;9(12):e0004316.
64. Olson PD, Tracey A, Baillie A, James K, Doyle SR, Buddenborg SK, et al. Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. *BMC Biol.* 2020;18(1).

65. Almeida CR, Stoco PH, Wagner G, Sincero TC, Rotava G, Bayer-Santos E, et al. Transcriptome analysis of *Taenia solium* cysticerci using Open Reading Frame ESTs (ORESTES). *Parasit Vectors*. 2009;2(1).
66. Wang X, Song W, Ji G, Song Y, Liu X, Luo X, et al. Regulation of DNA methylation on key parasitism genes of *Cysticercus cellulosae* revealed by integrative epigenomic-transcriptomic analyses. *Hereditas*. 2021;158(1).
67. Korhonen PK, Kinkar L, Young ND, Cai H, Lightowlers MW, Gauci C, et al. Chromosome-scale *Echinococcus granulosus* (genotype G1) genome reveals the Eg95 gene family and conservation of the EG95-vaccine molecule. *Commun Biol*. 2022;5(1).
68. Bai Y, Zhang Z, Jin L, Zhu Y, Zhao L, Shi B, et al. Dynamic Changes in the Global Transcriptome and MicroRNAome Reveal Complex miRNA-mRNA Regulation in Early Stages of the Bi-Directional Development of *Echinococcus granulosus* Protoscoleces. *Front Microbiol*. 2020;11:654.
69. Liu Z li, Wang S, Cai X peng, Zeng Q ying. Transcriptome profiling of *Cysticercus Pisiformis* provides insight into responses to host bile acids. *Parasitol Int*. 2021;81:102246.
70. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. 2017;215:2–10.

71. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, et al. InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* 2025;53(D1):D444-56.
72. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674-6.
73. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30(9):1236-40.
74. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Tamura K, editor. *Mol Biol Evol.* 2021;38(12):5825-9.
75. Ringnér M. What is principal component analysis? *Nat Biotechnol.* 2008;26(3):303-4.
76. Holmes S, Huber W. *Modern statistics for modern biology.* Cambridge, United Kingdom: Cambridge University Press; 2019.
77. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci.* 1998;95(25):14863-8.
78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

79. Agresti A. Foundations of linear and generalized linear models. Hoboken, New Jersey: John Wiley & Sons Inc; 2015. 1 p.
80. The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Baryshnikova A, editor. GENETICS. 2023;224(1).
81. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25-9.
82. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141.
83. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics. 2004;20(18):3710-5.
84. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005;102(43):15545-50.
85. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.
86. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol. 12 de enero de 2005;4(1).

- 87.Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-13.
- 88.Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019;47(8):e47-e47.
- 89.Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. Prlic A, editor. *PLoS Comput Biol.* 2013;9(8):e1003118.
- 90.Orrego-Solano MÁ, Cangalaya C, Nash TE, Guerra-Giraldez C. Identificación de células proliferativas en quistes de *Taenia solium*. *Rev Peru Med Exp Salud Pública.* 2014;31(4).
- 91.Borza R, Salgado-Polo F, Moolenaar WH, Perrakis A. Structure and function of the ecto-nucleotide pyrophosphatase/phosphodiesterase (ENPP) family: Tidying up diversity. *J Biol Chem.* 2022;298(2):101526.
- 92.Massé K, Bhamra S, Paroissin C, Maneta-Peyret L, Boué-Grabot E, Jones EA. The enpp4 ectonucleotidase regulates kidney patterning signalling networks in *Xenopus* embryos. *Commun Biol.* 2021;4(1).
- 93.Hancock K. Characterization and cloning of GP50, a *Taenia solium* antigen diagnostic for cysticercosis. *Mol Biochem Parasitol.* 2004;133(1):115-24.

94. Zhang H, Liu CY, Zha ZY, Zhao B, Yao J, Zhao S, et al. TEAD Transcription Factors Mediate the Function of TAZ in Cell Growth and Epithelial-Mesenchymal Transition. *J Biol Chem*. 2009;284(20):13355-62.
95. De Sousa N, Rodríguez-Esteban G, Rojo-Laguna JI, Saló E, Adell T. Hippo signaling controls cell cycle and restricts cell plasticity in planarians. Tapon N, editor. *PLOS Biol*. 2018;16(1):e2002399.
96. Boroni M, Sammeth M, Gava SG, Jorge NAN, Macedo AM, Machado CR, et al. Landscape of the spliced leader trans-splicing mechanism in *Schistosoma mansoni*. *Sci Rep*. 2018;8(1).
97. Mouchiroud L, Molin L, Kasturi P, Triba MN, Dumas ME, Wilson MC, et al. Pyruvate imbalance mediates metabolic reprogramming and mimics lifespan extension by dietary restriction in *Caenorhabditis elegans*. *Aging Cell*. 2011;10(1):39-54.
98. Janowitz T, Ajonina I, Perbandt M, Woltersdorf C, Hertel P, Liebau E, et al. The 3-ureidopropionase of *Caenorhabditis elegans*, an enzyme involved in pyrimidine degradation. *FEBS J*. 2010;277(19):4100-9.
99. Agorio A, Chalar C, Cardozo S, Salinas G. Alternative mRNAs Arising from Trans-splicing Code for Mitochondrial and Cytosolic Variants of *Echinococcus granulosus* Thioredoxin Glutathione Reductase. *J Biol Chem*. 2003;278(15):12920-8.
100. Jaeger S, Barends S, Giegé R, Eriani G, Martin F. Expression of metazoan replication-dependent histone genes. *Biochimie*. 2005;87(9-10):827-34.

101. Grigsby IF, Rutledge EM, Morton CA, Finger FP. Functional redundancy of two *C. elegans* homologs of the histone chaperone Asf1 in germline DNA replication. *Dev Biol.* 2009;329(1):64-79.
102. Balklava Z, Rathnakumar ND, Vashist S, Schweinsberg PJ, Grant BD. Linking Gene Expression in the Intestine to Production of Gametes Through the Phosphate Transporter PITR-1 in *Caenorhabditis elegans*. *Genetics.* 2016;204(1):153-62.
103. Guarnaschelli I, Koziol U. Albendazole specifically disrupts microtubules and protein turnover in the tegument of the cestode *Mesocostoides corti*. Williams DL, editor. *PLOS Pathog.* 2025;21(6):e1013221.
104. Bellemer A, Hirata T, Romero MF, Koelle MR. Two types of chloride transporters are required for GABA_Areceptor-mediated inhibition in *C. elegans*: Two chloride transporters control GABA signalling. *EMBO J.* 2011;30(9):1852-63.
105. Brockie PJ, Madsen DM, Zheng Y, Mellem J, Maricq AV. Differential Expression of Glutamate Receptor Subunits in the Nervous System of *Caenorhabditis elegans* and Their Regulation by the Homeodomain Protein UNC-42. *J Neurosci.* 2001;21(5):1510-22.
106. Kyuji A, Patel-King RS, Hisabori T, King SM, Wakabayashi KI. Cilia Loss and Dynein Assembly Defects in Planaria Lacking an Outer Dynein Arm-Docking Complex Subunit. *Zoolog Sci.* 2020;37(1):7.
107. Park YJ, Yeon J, Cho J, Kim DY, Bai X, Oh Y, et al. PIEZO acts in an intestinal valve to regulate swallowing in *C. elegans*. *Nat Commun.* 2024;15(1).

108. Combes D, Fedon Y, Toutant JP, Arpagaus M. Acetylcholinesterase genes in the nematode *Caenorhabditis elegans*. *International review of cytology*. 2001;207–39.
109. Zhang C, Li J, Aji T, Li L, Bi X, Yang N, *et al*. Identification of Functional MKK3/6 and MEK1/2 Homologs from *Echinococcus granulosus* and Investigation of Protoscolecidal Activity of Mitogen-Activated Protein Kinase Signaling Pathway Inhibitors *In vitro* and *In vivo*. *Antimicrob Agents Chemother*. 019;63(1).
110. Escobedo G, Soldevila G, Ortega-Pierres G, Chávez-Ríos JR, Nava K, Fonseca-Liñán R, *et al*. A New MAP Kinase Protein Involved in Estradiol-Stimulated Reproduction of the Helminth Parasite *Taenia crassiceps*. *J Biomed Biotechnol*. 2010;2010:1-11.
111. Orrego MA, Verastegui MR, Vasquez CM, Garcia HH, Nash TE. Proliferative cells in racemose neurocysticercosis have an active MAPK signalling pathway and respond to metformin treatment. *Int J Parasitol*. 2022;52(6):377-83.
112. Hayashizaki S, Iino Y, Yamamoto M. Characterization of the *C. elegans gap-2* gene encoding a novel Ras-GTPase activating protein and its possible role in larval development. *Genes Cells*. 1998;3(3):189-202.
113. Koziol U, Radio S, Smircich P, Zarowiecki M, Fernández C, Brehm K. A Novel Terminal-Repeat Retrotransposon in Miniature (TRIM) Is Massively Expressed in *Echinococcus multilocularis* Stem Cells. *Genome Biol Evol*. 2015;7(8):2136-53.

114. Jung P, Menssen A, Mayr D, Heiko Hermeking. AP4 encodes a c-MYC-inducible repressor of p21. *Proceedings of the National Academy of Sciences*. 2008;105(39):15046–51.
115. Abnave P, Aboukhatwa E, Kosaka N, Thompson J, Hill MA, Aboobaker AA. An X-ray shielded irradiation assay reveals EMT transcription factors control pluripotent adult stem cell migration in vivo in planarians. *Development*. 2017.
116. Jackstadt R, Roeh S, Neumann J, Jung P, Hoffmann RW, Horst D, et al. AP4 is a mediator of epithelial–mesenchymal transition and metastasis in colorectal cancer. *Journal of Experimental Medicine*. 2013;210(7):1331–50.
117. Palmer RE, Inoue T, Sherwood DR, Jiang LI, Sternberg PW. *Caenorhabditis elegans cog-1* Locus Encodes GTX/Nkx6.1 Homeodomain Proteins and Regulates Multiple Aspects of Reproductive System Development. *Developmental Biology*. 2002;252(2):202–13.
118. Tan PB, Lackner MR, Kim SK. MAP Kinase Signaling Specificity Mediated by the LIN-1 Ets/LIN-31 WH Transcription Factor Complex during *C. elegans* Vulval Induction. *Cell*. 1998;93(4):569–80.
119. Adalid-Peralta L, Rosas G, Arce-Sillas A, Bobes RJ, Cárdenas G, Hernández M, et al. Effect of Transforming Growth Factor- β upon *Taenia solium* and *Taenia crassiceps* Cysticerci. *Scientific Reports*. 2017;7(1).
120. Orrego MA, Del OH, Vasquez CM, Nash TE, Garcia HH. In vitro system for *Taenia solium* cysts culture and growth: progress toward a neurocysticercosis model. *Future Microbiology*. 2025;1–7.

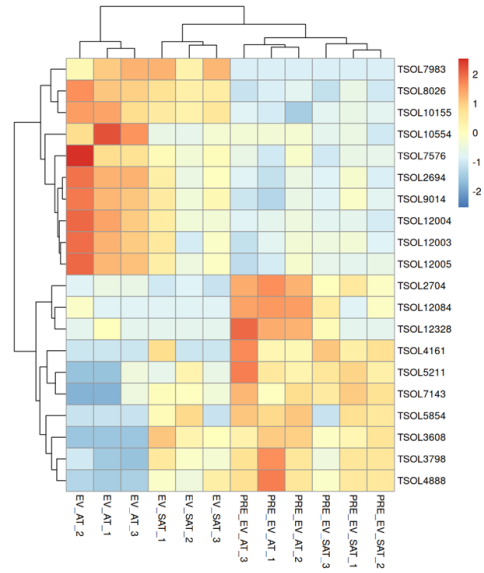
121. Segen Aklilu, Krakowiak M, Frempong A, Wilson K, Powers C, Fantz D. Nfya-1 functions as a substrate of ERK-MAP kinase during *Caenorhabditis elegans* vulval development. *Cells and Development*. 2021;169:203757–7.
122. Heo W, Hwang H, Kim J, Oh SH, Yu Y, Lee JH, et al. The CCAAT-box transcription factor, NF-Y complex, mediates the specification of the IL1 neurons in *C. elegans*. *BMB Reports*. 2023;56(3):153–9.
123. Oldfield Andrew J, Yang P, Conway Amanda E, Cinghu S, Freudenberg Johannes M, Yellaboina S, et al. Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. *Molecular Cell*. 2014;55(5):708–22.
124. Wasik K, Gurtowski J, Zhou X, Ramos OM, Delás MJ, Battistoni G, et al. Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. *Proceedings of the National Academy of Sciences*. 2015;112(40):12462–7.
125. Pickett CL, Breen KT, Ayer DE. A *C. elegans* Myc-like network cooperates with semaphorin and Wnt signaling pathways to control cell migration. *Developmental Biology*. 2007;310(2):226–39.
126. Sans CL, Satterwhite DJ, Stoltzman CA, Breen KT, Ayer DE. MondoA-Mlx Heterodimers Are Candidate Sensors of Cellular Energy Status: Mitochondrial Localization and Direct Regulation of Glycolysis. *Molecular and Cellular Biology*. 2006;26(13):4863–71.

127. Johnson DW, Llop JR, Farrell SF, Yuan J, Stolzenburg LR, Samuelson AV. The *Caenorhabditis elegans* Myc-Mondo/Mad Complexes Integrate Diverse Longevity Signals. Ashrafi K, editor. PLoS Genetics. 2014;10(4):e1004278.
128. Campbell KJ, White RJ. MYC Regulation of Cell Growth through Control of Transcription by RNA Polymerases I and III. Cold Spring Harbor perspectives in medicine. 2014;4(5):a018408–8.
129. Destefanis F, Manara V, Bellosta P. Myc as a Regulator of Ribosome Biogenesis and Cell Competition: A Link to Cancer. International Journal of Molecular Sciences. 2020;21(11):4037.

X. ANEXOS

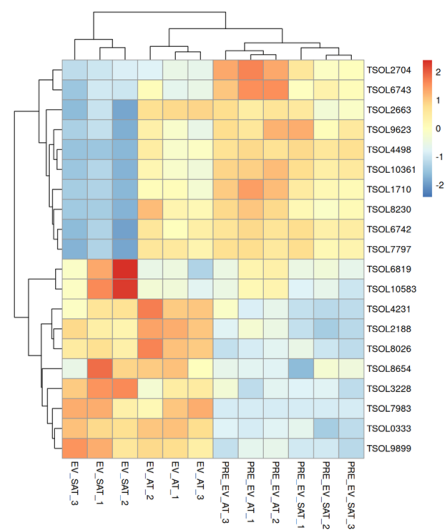
1. Genes con mayor y menor valor de *fold change* al comparar EV respecto a PRE en presencia de AT

Accession Number	Symbol	Description
TsM_000767500	TSOL7576	GLIPR1 protein 1
TsM_001212900	TSOL12004	ectonucleotide pyrophosphatase:phosphodiesterase
TsM_000275000	TSOL2694	intestinal type alkaline phosphatase 1
TsM_000911600	TSOL9014	Alkaline phosphatase
TsM_000808300	TSOL7983	hypothetical protein ECG_00274
TsM_001213000	TSOL12005	ectonucleotide pyrophosphatase:phosphodiesterase
TsM_000812600	TSOL8026	---NA---
TsM_001212800	TSOL12003	Bis(5'-adenosyl)-triphosphatase enpp4
TsM_001026200	TSOL10155	unnamed protein product
TsM_001066300	TSOL10554	unnamed protein product
TsM_000368100	TSOL3608	diagnostic antigen gp50
TsM_001245400	TSOL12328	Glycoprotein 3-alpha-L-fucosyltransferase A
TsM_000276000	TSOL2704	zinc finger, C4 type
TsM_000498700	TSOL4888	---NA---
TsM_001220900	TSOL12084	unnamed protein product, partial
TsM_000595500	TSOL5854	---NA---
TsM_000531100	TSOL5211	---NA---
TsM_000387400	TSOL3798	diagnostic antigen GP50
TsM_000425900	TSOL4161	unnamed protein product
TsM_000724200	TSOL7143	unnamed protein product



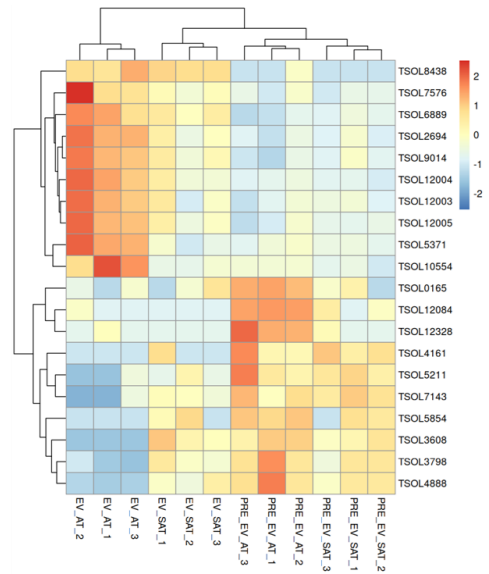
2. Genes con mayor y menor valor de *fold change* al comparar EV respecto a PRE en ausencia de AT

Accession Number	Symbol	Description
TsM_000808300	TSOL7983	hypothetical protein ECG_00274
TsM_000691800	TSOL6819	unnamed protein product
TsM_001069300	TSOL10583	unnamed protein product
TsM_001000500	TSOL9899	unnamed protein product
TsM_000033900	TSOL0333	hydrophobic ligand binding protein 1
TsM_000433000	TSOL4231	unnamed protein product
TsM_000329600	TSOL3228	unnamed protein product
TsM_000222900	TSOL2188	unnamed protein product
TsM_000875400	TSOL8654	unnamed protein product
TsM_000812600	TSOL8026	---NA---
TsM_000972300	TSOL9623	unnamed protein product
TsM_000833000	TSOL8230	unnamed protein product
TsM_000276000	TSOL2704	zinc finger, C4 type
TsM_000684200	TSOL6743	unnamed protein product
TsM_000459800	TSOL4498	unnamed protein product
TsM_000173200	TSOL1710	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2
TsM_000684100	TSOL6742	unnamed protein product
TsM_000271900	TSOL2663	unnamed protein product
TsM_000789600	TSOL7797	unnamed protein product
TsM_001046800	TSOL10361	diagnostic antigen gp50



3. Genes con mayor y menor valor de *fold change* al comparar la presencia de AT respecto a su ausencia en la fase de desarrollo EV

Accession Number	Symbol	Description
TsM_000767500	TSOL7576	GLIPR1 protein 1
TsM_001212900	TSOL12004	ectonucleotide pyrophosphatase:phosphodiesterase
TsM_000275000	TSOL2694	intestinal type alkaline phosphatase 1
TsM_000911600	TSOL9014	Alkaline phosphatase
TsM_001213000	TSOL12005	ectonucleotide pyrophosphatase:phosphodiesterase
TsM_001212800	TSOL12003	Bis(5'-adenosyl)-triphosphatase enpp4
TsM_001066300	TSOL10554	unnamed protein product
TsM_000853800	TSOL8438	---NA---
TsM_000698800	TSOL6889	EUB58500.1 Arginase-2
TsM_000547100	TSOL5371	unnamed protein product
TsM_000368100	TSOL3608	diagnostic antigen gp50
TsM_001245400	TSOL12328	Glycoprotein 3-alpha-L-fucosyltransferase A
TsM_000498700	TSOL4888	---NA---
TsM_001220900	TSOL12084	unnamed protein product, partial
TsM_000595500	TSOL5854	---NA---
TsM_000531100	TSOL5211	---NA---
TsM_000387400	TSOL3798	diagnostic antigen GP50
TsM_000425900	TSOL4161	unnamed protein product
TsM_000724200	TSOL7143	unnamed protein product
TsM_000016800	TSOL0165	Ankyrin repeat domain containing protein



4. Genes con mayor y menor valor de *fold change* al comparar la presencia de AT respecto a su ausencia en la fase de desarrollo PRE

Accession Number	Symbol	Description
TsM_000808300	TSOL7983	hypothetical protein ECG_00274
TsM_000812600	TSOL8026	---NA---
TsM_001026200	TSOL10155	unnamed protein product
TsM_000022000	TSOL0215	Cytidine deaminase
TsM_000566900	TSOL5569	CDS17276.1 tegumental protein
TsM_000953600	TSOL9436	unnamed protein product
TsM_001051600	TSOL10409	DAZ-associated protein 1
TsM_000161500	TSOL1596	unnamed protein product
TsM_000626100	TSOL6161	unnamed protein product
TsM_000566700	TSOL5567	tegumental antigen
TsM_000276000	TSOL2704	zinc finger, C4 type
TsM_000684200	TSOL6743	unnamed protein product
TsM_000605000	TSOL5950	unnamed protein product
TsM_000447100	TSOL4371	enolase C
TsM_000768200	TSOL7583	replication association protein
TsM_000210300	TSOL2066	Transcriptional enhancer factor TEF-3
TsM_001033100	TSOL10224	unnamed protein product
TsM_000183500	TSOL1808	unnamed protein product
TsM_000778800	TSOL7689	fibronectin-domain containing protein
TsM_000982300	TSOL9720	unnamed protein product

