

**META-ANÁLISIS DE LA EXPRESIÓN DE GENES ASOCIADOS A LA
DISFUNCIÓN ENDOTELIAL INDUCIDA POR VIRUS Y SU RELACIÓN
CON LA PLASTICIDAD CELULAR**

Helifonso Castellanos Méndez

Cód. 1.094.275.231

Universidad de Pamplona
Facultad de Ciencias Básicas
Departamento de Biología
Pamplona, Norte de Santander

2021

**META-ANÁLISIS DE LA EXPRESIÓN DE GENES ASOCIADOS A LA
DISFUNCIÓN ENDOTELIAL INDUCIDA POR VIRUS Y SU RELACIÓN
CON LA PLASTICIDAD CELULAR**

Trabajo de grado para optar por el título de:

Biólogo

Director:

Nelson Josué Fernández B. Sc. M. Sc. Ph. D

Codirectora:

Juliana Estefanía Arcila Galvis B. Sc. M. Sc. Ph. D.(c)

Universidad de Pamplona
Facultad de Ciencias Básicas
Departamento de Biología
Pamplona, Norte de Santander

23 de noviembre de 2021

Agradecimientos

Primero al Dios de amor y energía, el Dios de Spinoza.

A mi familia, principalmente a mi madre, Graciela, por ser lo más importante en mi vida, pero también por la entrega y sacrificios hechos por mí. A mi hermano, Daniel, por su voluntad, apoyo y confianza, a mi padre, mis demás hermanos y sobrinos.

A mis amigos, Mafe, Alex, Deisy, Angella, por ser apoyo emocional a lo largo de mi carrera y vida.

A mis compañeros de Universidad, por la cooperación y la amistad brindada.

Por último y no menos importante, a mi equipo de investigación, Paula, Juliana y Nelson, por tan invaluable apoyo, compañía y contribución en la misma.

RESUMEN

La disfunción endotelial es una de las causas más importantes en numerosas patologías, como aterosclerosis e hipertensión arterial, entre otras. Se caracteriza por la pérdida de uniones intercelulares en la barrera endotelial de los vasos sanguíneos, en algunos casos a raíz de infecciones virales. Investigaciones recientes muestran que los cambios fenotípicos y de expresión génica en la disfunción endotelial, tienen cierta similitud con el proceso de transición endotelial a mesenquimal (EndMT). Surge por tanto la hipótesis de que, si existe una alta asociación entre los genes y rutas de expresión génica entre los dos fenómenos, entonces la disfunción endotelial podría ser estudiada desde la perspectiva de la EndMT, particularmente en un contexto de infección viral. Esta declaración requiere del establecimiento de (i) modelos teóricos desde la literatura existente y (ii) su posterior experimentación en laboratorio. En este sentido, esta investigación se enfoca en el primer aspecto (punto i), y busca determinar los genes y/o moléculas relevantes de la disfunción y la transición endotelial. Para ello, se partió de un enfoque computacional que consideró la realización de una revisión sistemática, un metaanálisis, la descripción de categorías, análisis de redes de correlación y de interacción génica. Desde un cribado de literatura de 124 mil documentos, se hallaron 122 genes determinantes en la disfunción, 63 para la EndMT y 6 comunes. La comparación estadística de 2 datasets de ARNm, en disfunción endotelial, dio como resultado 42 transcritos diferencialmente expresados. En tanto que, para el análisis del único dataset de EndMT arrojó 4 y 87 genes comunes respectivamente con los datasets 1 y 2 de la disfunción. La suma de los genes del meta-análisis fue de 133. El enfoque metodológico de esta investigación permitió el establecimiento de un protocolo para metaanálisis, disponible para su aplicación y validación por la comunidad académica que trabaja en el área. Los hallazgos presentados conducen a generar conocimiento para el entendimiento de la disfunción y

la transición en términos de infección viral, así como su posterior ampliación en investigaciones experimentales. En este aspecto, esta propuesta es la base de una investigación traslacional con miras a la generación e implementación de metodologías y estrategias para la apropiación, transferencia y uso de nuestros resultados a nivel glocal.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	13
2. MARCO DE REFERENCIA	14
2.1. MARCO INSTITUCIONAL	14
2.2. TRANSICIÓN ENDOTELIAL AL MESENQUIMAL (ENDMT)	15
2.3. DISFUNCIÓN ENDOTELIAL	17
2.4. ANÁLISIS DE EXPRESIÓN DIFERENCIAL	19
2.5. REVISIÓN SISTEMÁTICA Y METAANÁLISIS	21
2.5.1. ImaGEO	23
2.5.2. IRIS-EDA	25
3. OBJETIVOS	27
3.1. OBJETIVO GENERAL	27
3.2. OBJETIVOS ESPECÍFICOS	27
4. METODOLOGÍA	27
4.1. DEFINICIÓN DE DATASETS Y ESTUDIOS RELEVANTES	28
4.1.1. Rastreo y selección de artículos científicos	28
4.1.2. Selección de datasets para metaanálisis	31
4.2. PROTOCOLO PARA METAANÁLISIS	31
4.2.1. Datasets de prueba	32
4.2.2. Análisis de softwares	32
4.3. IDENTIFICACIÓN DE MOLÉCULAS	32
4.3.1. Revisión sistemática	32

	7
4.3.2. Metaanálisis	33
4.3.2.1. Metaanálisis ImaGEO	33
4.3.2.2. Análisis de datasets por GEO2R	33
4.3.2.3. Comparación de resultados de GEO2R	34
4.4. ANÁLISIS DE VÍAS DE EXPRESIÓN DE LAS MOLÉCULAS DIFERENCIALMENTE EXPRESADAS	35
4.4.1. Construcción de red de asociación de moléculas	35
4.4.2. Identificación de moléculas relevantes en los fenómenos	35
5. RESULTADOS	35
5.1. CRIBADO DE ARTÍCULOS Y DATASETS	35
5.2. ANÁLISIS DE SOFTWARES	40
5.2.1. NetworkAnalyst 3.0	40
5.2.2. Análisis de ImaGEO	44
5.2.3. Análisis de IRIS-EDA (An Integrated RNA-Seq Interpretation System for gene Expression Data Analysis)	46
5.2.4. Árbol de decisión	46
5.3. IDENTIFICACIÓN DE MOLÉCULAS	50
5.3.1. Moléculas de la revisión sistemática	50
5.3.2. Moléculas del metaanálisis	53
6. DISCUSIÓN	61
7. CONCLUSIONES	72
8. REFERENCIAS BIBLIOGRÁFICAS	73

ÍNDICE DE TABLAS

Tabla 1. Clasificación del nivel de evidencia y fuerza de la recomendación , tomado de (M ^a et al., 2008).	22
Tabla 2. Ecuaciones de búsqueda . Los términos en negrita comprenden la ecuación base de la búsqueda y los demás términos aluden a los excluyentes utilizados. Se empleó el comodín (*) para mayor exactitud de la búsqueda.	28
Tabla 3. Síntesis porcentual del cribado en los repositorios científicos para la revisión sistemática. De acuerdo a las ecuaciones sintéticas o basales se indica, el número de estudios resultantes de la búsqueda sin restricción (SR), el número de estudios resultantes de la ecuación sintética con restricciones (CR), el estado porcentual de estudios seleccionados (%SRS) y de los estudios excluidos (%CREx). Los documentos finalmente seleccionados (FS) son producto de la aplicación de filtros a CR desde cada repositorio, indicando el porcentaje de estudios excluidos (%FSEx) por los mismos y el porcentaje de seleccionados (%FS).	36
Tabla 4. Síntesis porcentual del cribado de datasets para el metaanálisis . De 31 datasets preseleccionados, se excluye el 90.32% y finalmente, solo el 9.68% es seleccionado.	38
Tabla 5. Parámetros usados para el tratamiento de los datos en NetworkAnalyst .	41
Tabla 6. Resultados del metaanálisis ejecutado en NetworkAnalyst . Descripción de los genes diferencialmente expresados y el tiempo de ejecución en cada modelo.	43
Tabla 7. Resultados del metaanálisis realizado en ImAGEO . Se indica el número de genes diferencialmente expresados en cada dataset, la conjugación promedio de los 3 datasets y el tiempo de ejecución tardado en cada modelo.	46
Tabla 8. Propiedades de las moléculas integradas desde la literatura . A) disfunción endotelial en la infección del virus del dengue se reportaron 165 moléculas con diferente	

comportamiento, B) disfunción endotelial en infección viral, se reportaron 24 moléculas y C) fenómeno de Transición endotelial a mesenquimal, reporta 122 moléculas; la frecuencia de los cambios por molécula (Change type) donde Up, se refiere a las moléculas sobre expresadas, Down, las moléculas subexpresadas, Present, se refiere a las moléculas que tiene influencia en el fenómeno sin cambios de expresión, Inhibited, alude a las moléculas inhibidas, P+, aquellas moléculas que se fosforilaron. La frecuencia por los tejidos más relevantes menciona, U.C. (cordón umbilical), Derm. (dermis), Lung (pulmón), Card. (cardiaco), Brain (cerebral), Int. (intestinal), LEC (células endoteliales linfáticas), Other (otros tejidos involucrados); las moléculas expresadas a nivel mRNA (ARN mensajero), protein (proteína) o both (ARN mensajero y proteína), N.I. (No identificado); los tipos de molécula reportados, Qc (quimiocinas), Enz (enzimas), NF (factor de necrosis), Gp (Glucoproteína), GF (factor de crecimiento).

50

Tabla 9. **Genes diferencialmente expresados en disfunción endotelial.** Se indican los genes con diferencia estadísticamente significativa, resultante de la comparación de GSE34628 y GSE142762 por medio de la plataforma ImaGEO.

54

Tabla 10. **Anexo #1.** Tabla de ecuaciones de búsqueda utilizada en los repositorios científicos.

86

Tabla 11. **Anexo #2.** Tabla de cribado y selección de datasets para el metaanálisis.

88

ÍNDICE DE FIGURAS

- Figura 1. **Transición endotelio-mesénquima y Moléculas Involucradas Conocidas**
(Adaptado de (Piera-Velazquez & Jimenez, 2019)). Se parte de una morfología endotelial con características de monocapa estrechamente unida, que durante el proceso va disminuyendo la expresión de productos como CD31/ moléculas de adhesión entre células endoteliales y plaquetas (PECAM-1), cadherina endotelial vascular (VE-cadherin), COL4, receptor de factor de crecimiento epidérmico-vascular (VEGRF) y el factor von Willebrand. Esto da lugar a la nueva morfología en las células con aumento en la expresión de moléculas como, actina de músculo liso- α (SMA- α), COL1, COL3, fibronectina (FN), dominio extra de FN-A (EDA)/N-cadherina y proteína específica de fibroblastos-1 (FSP-1). 16
- Figura 2. Ilustración de mapeo y cuantificación de genes expresados diferencialmente.
Adaptado de (Costa et al., 2010). 21
- Figura 3. **Flujo de trabajo de ImGEO**: 1. Entrada de los datos, 2. Control de calidad de los datos ingresados, 3. Meta-análisis de los datos suministrados, 4. Análisis funcionales (opcional de acuerdo al objetivo del estudio), 5. Reporte de resultados. (Toro-Domínguez et al., 2019). 25
- Figura 4. **Flujo de trabajo de IRIS-EDA**: sección de ingreso de datos propios desde servidor Input Data (Submit and QC), así como desde un repositorio por medio del ID. GEO en Data Submission (GEO). Una sección dedicada al análisis de los datos, Discovery-Driven Analyses. Y una sección dedicada al análisis de expresión diferencial, Differential Gene Expression (DGE Analysis) (<http://bmb1.sdstate.edu/IRIS/>). 26
- Figura 5. **Flujo de búsqueda y selección de artículos científicos y datasets**. Pasos empleados en la selección de la literatura y los datasets empleados en los análisis. 31
- Figura 6. **Ejemplo del ajuste al formato de cada dataset**. 41

Figura 7. **Árbol de decisión.** Se presenta un esquema que identifica los atributos más relevantes (tiempo de ejecución, capacidad de integración de datos, control de parámetros), en la ejecución de un metaanálisis. El nodo de probabilidad, corresponde a una opción que mostrará múltiples resultados inciertos. El nodo de decisión, indica una elección compuesta por diferentes variables. La ramificación alterna indica diferentes opciones dentro de un nodo. La ramificación dependiente del dataset, indica que alguno de los procesos descritos es directamente influenciado por el objeto (dataset). La valoración truncada corresponde a un proceso que no se puede evaluar. La ramificación rechazada alude a una variable que se descarta. Y el nodo terminal, se refiere a un resultado definido. N.E. No explorado. 47

Figura 8. **Análisis de cajas y bigotes datasets, ImaGEO.** GSE34628 infección con virus del dengue, GSE142762, infección con citomegalovirus. 53

Figura 9. **Mapa de calor, ImaGEO.** Cada color en la escala de degradado representa un grado de cambio de expresión de los genes por muestra en cada dataset. Los valores <0 representan subexpresión y los valores >0 representan, sobreexpresión del gen en cada muestra de cada dataset. Mapa resultante de la comparación de GSE34628 y GSE142762 por medio de la plataforma ImaGEO. 58

Figura 10. **Diagrama de Venn.** Comparativo de datasets entre el estudio de EndMT en infección de virus de Kaposi (GSE22522) y los datasets de disfunción endotelial (GSE34628) y (GSE142762). 60

Figura 11. **Red de moléculas asociadas a la disfunción endotelial y EndMT.** Los nodos verdes definen las moléculas, los hexágonos azules representan las rutas de expresión relacionadas con las moléculas (las rutas no necesariamente por completo en el proceso). Los nodos lila, representan aquellas rutas y moléculas que llevan el mismo nombre. Los rombos rojos

representan las moléculas, genes o virus que inducen lesión. Por último, los círculos amarillos, representan los tejidos implicados. 70

Figura 12. Red circular de moléculas asociadas a la disfunción endotelial y EndMT.

Indica el círculo más grande (izquierda) como las interacciones entre moléculas (nodos verdes), rutas de expresión (nodos azules), moléculas y rutas con el mismo nombre (nodos lila), agentes o moléculas inductoras de lesión (nodos rojos) y tejidos (nodos amarillos) en la disfunción endotelial, así mismo, las interacciones en la EndMT (círculo derecho). 71

1. INTRODUCCIÓN

El fenómeno de la plasticidad celular se refiere a la dediferenciación de células ya diferenciadas, hasta la adquisición de un fenotipo más inmaduro dentro de su mismo linaje celular (Tata & Rajagopal, 2016); también se le denomina reprogramación, mediante un proceso inverso de diferenciación en el que las células diferenciadas regresan hasta sus primeros “programas embrionarios” (Hincapié-Monsalve & Gallego-Gómez, 2020). La diferenciación celular ocurre en los procesos de embriogénesis y organogénesis para la formación de tejidos especializados partiendo de un mismo linaje celular. Este proceso de reversión de las células especializadas a fenotipos más indiferenciados es de gran relevancia en las técnicas de la medicina regenerativa, que realiza procedimientos con células madres, por su mayor potencial de plasticidad.

Esta transición celular se ha evidenciado en células endoteliales, caracterizada por la inducción de anomalías en las uniones intercelulares, la disminución de expresión de moléculas como la cadherina vascular endotelial (VE-cadherina), sobreproducción de factores de crecimiento, mayor presencia y expresión de moléculas proinflamatorias, así como, aumento del estrés oxidativo y mayor permeabilidad en la capa celular, entre otras (Sena et al., 2013a), que conlleva a una transformación progresiva al fenotipo mesenquimal.

En la actualidad se ha reportado que esta plasticidad celular puede estar causada por la infección de virus, como el caso del virus del Herpes del sarcoma de Kaposi (P. Y. Chen et al., 2012; Cheng et al., 2011), el citomegalovirus (G. Chen et al., 2019) y recientemente, experimentos in vitro de células endoteliales de microvasculatura expuestas a medios condicionados por infección con el virus del dengue, mostraron tanto cambios fisiológicos como

morfológicos, así como, migración celular y expresión de microARN similares al fenómeno de transición endotelial a mesenquimal (Álvarez-Díaz et al., 2019).

De los estudios realizados en la disfunción endotelial inducida por la infección del virus del dengue (DENV) (Puerta-Guardo et al., 2019; Singh et al., 2018), se observó que los cambios fenotípicos de las células endoteliales son similares a los que ocurren en la transición endotelial a mesenquimal, por lo que surge la hipótesis de que, la disfunción endotelial en la infección viral, corresponde en realidad, a una transición endotelial a mesenquimal. Sin embargo, no hay conocimiento total de las moléculas y genes implicados en este proceso de disfunción endotelial inducida por este virus y, por ende, la relación real entre ambos procesos no ha podido ser definida con claridad.

De acuerdo a lo anterior, esta investigación reunió información teórica de genes relacionados con la disfunción endotelial en la infección de virus, así como los genes involucrados en la transición endotelial a mesenquimal, inducida por diversas lesiones y por infección viral. Así mismo, por medio de un protocolo estimado y evaluado para el metaanálisis, cotejamos datasets disponibles que contribuyan al enriquecimiento del listado de genes con cambios de expresión en estos fenómenos, a fin de que sean útiles como base teórica para posterior comprobación experimental.

2. MARCO DE REFERENCIA

2.1. Marco Institucional

Esta investigación se desarrolla en el marco del proyecto “*DENGUE: Disfunción Endotelial - Nuevo Giro que Urge en el Enfoque fisiopatológico para entender la enfermedad*” aprobado por Colciencias en la convocatoria “844-2019- PACTO PARA LA GENERACIÓN

DE NUEVO CONOCIMIENTO A TRAVÉS DE PROYECTOS DE INVESTIGACIÓN”, del programa nacional de CTeI de Ciencia, Tecnología e Innovación en Salud, bajo el código 66951.

El proyecto tiene una duración de 36 meses y la entidad ejecutora es la Universidad de Antioquía (Grupo de Medicina Molecular y de Traslación y Grupo Coloides). Cuenta con la co-ejecución de la Universidad del Quindío (Grupo de Inmunología Molecular), la Corporación Universitaria y Empresarial Alexander Von Humboldt (Grupo de Pedagogía y Responsabilidad Social Universitaria) y la Universidad de Pamplona (Grupo de Ecología y Biogeografía).

Los descriptores del proyecto son: La ciencia de datos, la biomecánica celular, IMATINIB, MICRORNAS, Choque, C-ABL, Disfunción Endotelial y Dengue.

El objetivo general busca: *Determinar si la infección por virus Dengue induce disfunción endotelial mediante transición endotelio-mesénquima*. En este sentido, nuestra propuesta de investigación se alinea con el objetivo específico número uno: *Determinar si factores solubles de la infección por virus Dengue inducen disfunción endotelial*, con el cual se espera generar conocimiento sobre *si la disfunción endotelial tiene marcadores moleculares y celulares compatibles con fenotipo mesenquimatoso*.

2.2. Transición Endotelial al Mesenquimal (EndMT)

La transición endotelial a mesenquimal, EndMT o EndoMT (Endothelial to mesenchymal transition), es un proceso mediante el cual, las células endoteliales pierden sus uniones célula a célula y cambian su fenotipo redondeado progresivamente, hasta un estado mesenquimal. En este punto, su morfología se transforma en una célula alargada y puntiaguda, y sucede la disminución

en la expresión de genes propios de células endoteliales, dando paso al aumento en la expresión de genes propios de células mesenquimales (figura 1) (Piera-Velazquez & Jimenez, 2019).

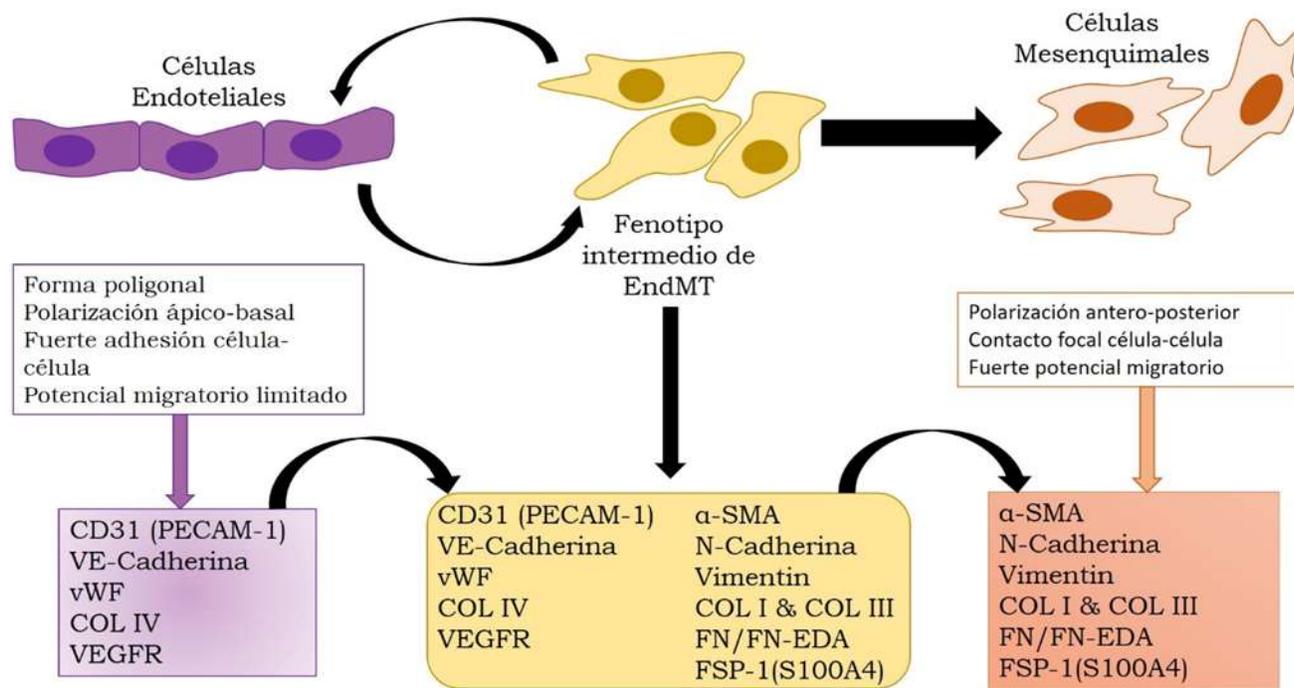


Figura 1. Transición endotelio-mesénquima y moléculas Involucradas Conocidas (Adaptado de (Piera-Velazquez & Jimenez, 2019)). Se parte de una morfología endotelial con características de monocapa estrechamente unida, que durante el proceso va disminuyendo la expresión de productos como CD31/ moléculas de adhesión entre células endoteliales y plaquetas (PECAM-1), cadherina endotelial vascular (VE-cadherin), COL4, receptor de factor de crecimiento epidérmico-vascular (VEGRF) y el factor von Willebrand. Esto da lugar a la nueva morfología en las células con aumento en la expresión de moléculas como, actina de músculo liso- α (SMA- α), COL1, COL3, fibronectina (FN), dominio extra de FN-A (EDA)/N-cadherina y proteína específica de fibroblastos-1 (FSP-1).

Cabe destacar que los cambios de estado a nivel morfológico y funcional en la EndMT, así como los cambios en los patrones de expresión, se han observado en la transición inducida por algunas lesiones del tejido (Rieder et al., 2011a). También, por infecciones virales como el caso del virus del herpes asociado al sarcoma de Kaposi (KSHV), en el que las células endoteliales pierden la expresión de marcadores fenotípicos endoteliales maduros y adquieren marcadores fenotípicos, y funcionales de células mesenquimales (Gasperini et al., 2012).

Durante el proceso de EndMT, se efectúan modificaciones en las que son responsables algunos genes, señales de inducción molecular, ARN no codificante, moléculas proinflamatorias, y factores de transcripción. Gran parte de las moléculas reguladoras no han sido completamente dilucidadas. Sin embargo, diversos estudios demuestran que los miembros de la familia de factores de crecimiento transformante como TGF β 1, TGF β 2, y TGF β 3, son los principales inductores de EndMT (Arciniegas et al., 1992; Goumans et al., 2008; Mihira et al., 2012; Wertz et al., 2011).

Además de las vías de señalización de TGF β , se ha demostrado la participación de otras vías moleculares que incluyen la señalización *Notch* (Gasperini et al., 2012), la vía *Wnt* canónica, así como otros mecanismos reguladores importantes inducidos por las citoquinas y otros mediadores inflamatorios, endotelina-1 (ET-1), el factor inducible de hipoxia-1 (HIF-1) y la respuesta al estrés oxidativo. Otras moléculas como factores de crecimiento derivados de plaquetas PDGFAA y PDGFBB también tienen un rol en la iniciación, progresión y regulación de EndMT (Piera-Velazquez & Jimenez, 2019)

Por otro lado, algunos mediadores inflamatorios son inductores eficientes de la EndMT, siendo IL-1 uno de los más potentes, (Rieder et al., 2011a)afirman que la combinación de TGF- β 1, IL-1 β y TNF- α , o los sobrenadantes de células mononucleares de lámina propia (LPMC) activados inducen cambios morfológicos y fenotípicos consistentes con EndMT, con un efecto dominante de IL-1. Así mismo, en la *Transición Epitelial a Mesenquimal* (EMT), TGF- β 1, el factor de necrosis tumoral – α (TNF- α) e IL-1 β , son los principales impulsores de EndMT, lo que sugiere similitudes mecanicistas entre los dos procesos.

2.3. Disfunción endotelial

Las células endoteliales vasculares son altamente organizadas, fisiológicamente especializadas y constituyen la barrera principal que controla el flujo de nutrientes tisulares, la dinámica de la microcirculación y la permeabilidad de las macromoléculas. Estas funciones de la barrera endotelial, pueden verse afectadas por patologías como la sepsis, la hipertensión, la aterosclerosis y la lesión por reperfusión isquémica (Mudau et al., 2012; Steyers & Miller, 2014)

La disfunción en la barrera endotelial está implicada en varios tipos de enfermedades, como cardiovasculares, coronarias y pulmonares (Endemann & Schiffrin, 2004). Esta disfunción se caracteriza por sobreproducción de factores de crecimiento, un aumento en las moléculas Endotelina-1 (ET-1), Angiotensina-II (Ang II), Tromboxano A2, especies reactivas de Oxígeno (ROS) ocasionando la vasoconstricción (Sena et al., 2013a). También se presentan pérdidas de unión célula ocasionadas por la disminución en moléculas de adhesión tales como, moléculas de adhesión de células vasculares (VCAM)-1, e-selectina, moléculas de adhesión intercelular (ICAM)-1 (Muñoz-Hernández et al., 2020), aumento en la presencia de moléculas de inflamación, aumento del estrés oxidativo, entre otras (Sena et al., 2013b).

Las infecciones virales pueden también estar implicadas en la disfunción endotelial. Por ejemplo, (Armstrong et al., 2013) reportan que el virus de la Influenza A, genera elevados niveles de citocinas y quimiocinas proinflamatorias, capaces de inducir directamente interrupciones en las uniones célula-célula lo que genera una fuga endotelial. Para el virus del Dengue (DENV), así como para otros virus hemorrágicos como Arenavirus y Hantavirus, la disfunción endotelial presentó un bajo efecto citopático in vitro (Spiropoulou & Srikiatkachorn, 2013). Sin embargo, la disfunción endotelial en Dengue podría estar relacionada con un proceso de EndMT que alteraría la permeabilidad endotelial, mediante la desestabilización de las uniones célula-célula. Álvarez-Díaz y colaboradores (2019), demostraron que, los medios condicionados

por la infección con DENV, como tratamiento sobre células endoteliales de microvasculatura humana, inducen cambios funcionales y morfológicos compatibles con el remodelamiento microvascular, migración celular y EndMT. Así mismo, Cheng y colaboradores (2011), reportaron que KSHV reprograma células endoteliales linfáticas primarias a células mesenquimales a través de EndMT, infiriendo que la heterogeneidad de las células presentes en la lesión de Sarcoma de Kaposi, puede tener origen en el endotelio infectado a través de este proceso.

2.4. Análisis de expresión diferencial

Con los descubrimientos y avances en investigaciones hechas sobre los ácidos nucleicos y sus características, se obtuvo el desarrollo de los métodos de secuenciación de primera generación, alrededor de la década de los sesenta. Entre los más destacados se encuentran, el método de modificación y escisión química, desarrollado por Allan Maxam y Walter Gilbert, y el método de secuenciación por síntesis ideado por Frederick Sanger (Ortigosa & Cañas, 2020).

Con la ventaja que demostró la secuenciación, a nivel de investigativo y médico, se promovió los avances hasta la secuenciación de próxima generación (NGS, Next Generation Sequencing), que cuenta con numerosas ventajas en relación con la secuenciación de primera generación. Se destaca que, la NSG cuenta con un rango dinámico más amplio en cuantificación de los niveles de expresión, requiere menor cantidad de muestra (ADN o ARN), no requiere el conocimiento previo de las características del genoma y permite un mayor grado de reproducibilidad (*¿Qué es la secuenciación de ADN de próxima generación? | Genómica funcional II, s/f*)

Las tecnologías que han permitido los avances en la NGS son: tecnología 454 (2004-2005) considerada como la pionera. En 2006 *Illumina*, basada en secuenciación por síntesis. SOLiD que se fundamenta en secuenciación por ligación (2007), y por último *Ion Torrent* (2010) que basa su método en la detección de protones (H⁺) (Aguilar-Bultet & Falquet, 2015), dando como resultado millones de *reads* (denominación para lecturas en secuenciación que consiste en segmento de secuencia específicos estudiados) en un solo proceso de secuenciación.

Los *reads*, han sido de gran utilidad en la identificación y caracterización de los transcritos expresados en una célula o tejido específico, en la determinación de la estructura de genes, patrones de empalme y la cuantificación de expresiones diferenciales en condiciones fisiopatológicas (Costa et al., 2010), o de procesos relacionados, como entre la transición epitelio a mesénquima y la transición endotelio a mesénquima (Srivastava et al., 2013).

Para la medición de niveles de expresión de ARN en sistemas biológicos, los métodos principales que se utilizan son el RNA-seq y los micro-arreglos de ADN (Heller, 2002; Oszolak & Milos, 2011).

Los micro-arreglos (también llamados chips de ADN, ADNc) se fundamenta en millones de sondas de oligonucleótidos individuales fijadas en una superficie sólida (de sílice, plástico o vidrio). Cada sonda se hibrida directamente con una secuencia de un gen de interés, el cual será detectado mediante el escaneo de señales fluorescentes para cuantificar que su intensidad sea análoga a su grado de expresión (Collí-Dulá et al., 2016; Lovén et al., 2012). El RNA-seq cuenta con distinguidas ventajas sobre los micro-arreglos como lo son la identificación de eventos de translocación, variaciones de nucleótidos y modificaciones de bases postranscripcionales (Costa et al., 2010), la cuantificación de isoformas y transcritos desconocidos.

La tecnología RNA-seq., consiste en la secuenciación de pequeños fragmentos de ADN complementario, estas secuencias se ensamblan de *novo* o a partir de un genoma o transcriptoma de referencia para la anotación los patrones de expresión de genes (Costa-Silva et al., 2017). El nivel de expresión de cada unidad de ARN se cuantifica por el número de fragmentos (reads), que se asignan a la transcripción, relacionándose con su nivel de abundancia (Rapaport et al., 2013), como se ilustra en la figura 2.

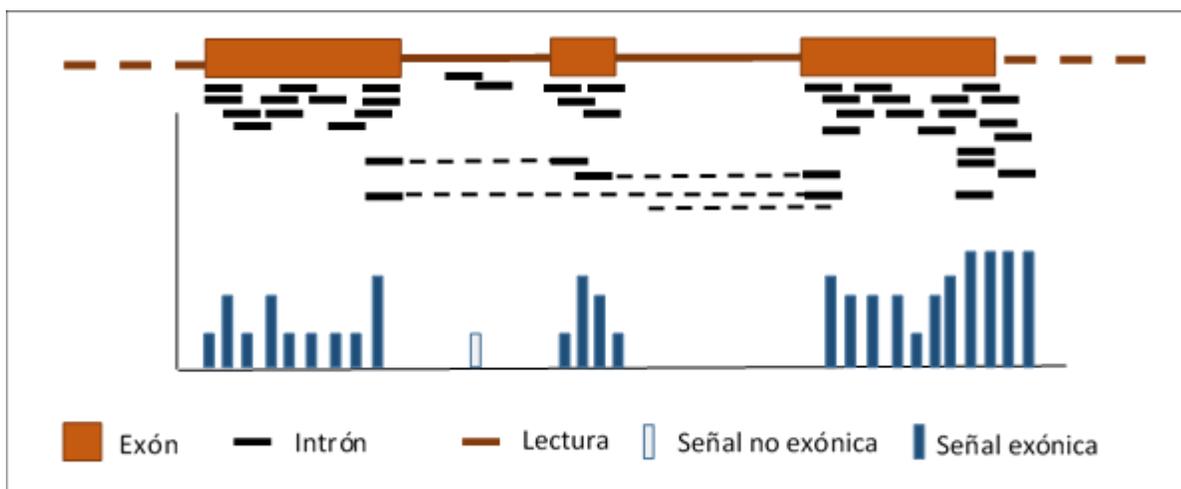


Figura 2. Ilustración de mapeo y cuantificación de genes expresados diferencialmente. Adaptado de (Costa et al., 2010).

El análisis de expresión diferencial de genes, permite cuantificar la diferencia de la expresión de los transcritos presentes bajo dos o más condiciones. La selección de los transcritos cuya expresión diferencial es significativa se hace principalmente con base a una combinación de parámetros como el umbral de cambio ‘Fold Change’, referido a la razón de cambio entre la expresión del mismo gen en dos condiciones evaluadas.

2.5. Revisión sistemática y Metaanálisis

La revisión sistemática, es entendida como un proceso objetivo y replicable, que recopila todos los estudios disponibles relacionados con una pregunta específica, donde se analizan y compilan los resultados (Ahn & Kang, 2018).

Las revisiones sistemáticas utilizan métodos explícitos y reproducibles, a fin de identificar, valorar críticamente y categorizar investigaciones hechas que se relacionen con la pregunta o hipótesis planteada. Con ella se busca obtener datos e información de dichas investigaciones con el propósito de poder obtener conclusiones válidas y antecedentes de robustez, acorde con la importancia de su aporte y su fuerza de recomendación (tabla 1).

Tabla 1. Clasificación del nivel de evidencia y fuerza de la recomendación, tomado de (M^a et al., 2008).

Fuerza de la recomendación	Nivel de evidencia	Fuente
A	1a.	Revisión sistemática de ensayos clínicos aleatorizados
	1b.	Ensayo clínico controlado aleatorizado
B	2a.	Evidencia de por lo menos un estudio controlado sin aleatorización
	2b.	Al menos otro tipo de estudio cuasiexperimental o estudios de cohorte
C	3	Evidencia de un estudio descriptivo no experimental, tales como estudios comparativos, estudios de correlación, casos y controles y revisiones clínicas

D	IV	Evidencia de comité de expertos, publicaciones, opiniones o experiencia clínica de autoridades en la materia, o de ambas.
----------	----	---

En un meta-análisis médico/terapéutico, la obtención de datos disponibles puede contribuir en la resolución de problemas, al integrar información que permite dilucidar factores, puntos concéntricos o patrones en distintas patologías (Castro Ávalos, 2020; Piras et al., 2018; Tejera-Vaquerizo et al., 2016) De esta forma, los meta-análisis de genes, contribuyen a identificar firmas y vías genicas correlacionadas entre ciertas infecciones y las respuestas inmunes que desencadenan, identificando patrones de coexpresión en grupos de células o tejidos (Badr & Häcker, 2019)

Existen diversos métodos para la ejecución de un meta-análisis, de los que podemos nombrar: la combinación de valores de significancia (p), puntajes Z, rangos o tamaños de efectos, análisis por conteo de votos, análisis agrupado; que encuentran diferencia en el marco estadístico para cada caso (Sweeney et al., 2017)Adicionalmente, existen programas y plataformas especializadas para meta-análisis (por ejemplo: GeneMeta (Lusa et al., 2006), MetaDE (Wang X. et al 2012), ExAtlas (Sharov et al., 2016), rmeta (Lumley T. 2015), metafor (Viechtbauer W.2010), IRIS-EDA (Monier et al., 2019), ImaGEO (Toro-Domínguez et al., 2019), Networks Analyst (Zhou et al., 2019), MetaOmics (João F. et al., 2017))

De los cuales resaltamos:

2.5.1. ImaGEO

Se presenta como una herramienta en línea para el Meta-análisis de expresión génica de acceso gratuito, fácil acceso y cuenta con una plataforma de interfaz sencilla y práctica a los usuarios. Una de las principales ventajas a destacar, se refiere al momento de ingreso de los datos, donde presenta la opción de insertar los identificadores (ID. GEO), sin la necesidad de descargarlos en el ordenador antes de ser analizados. Esta aplicación cuenta con diferentes métodos y modelos para la ejecución del metaanálisis.

Su flujo de trabajo (fig. 3) comprende, (i). ingreso y selección de parámetros para el análisis. (ii). Control de calidad de las muestras ingresadas en cada dataset. (iii). Compaginación de métodos, adaptados de las funciones del paquete MetaDE de R. (iv). Análisis funcional de genes sub-expresados o sobre-expresados en el meta-análisis. (v). Reporte de análisis.

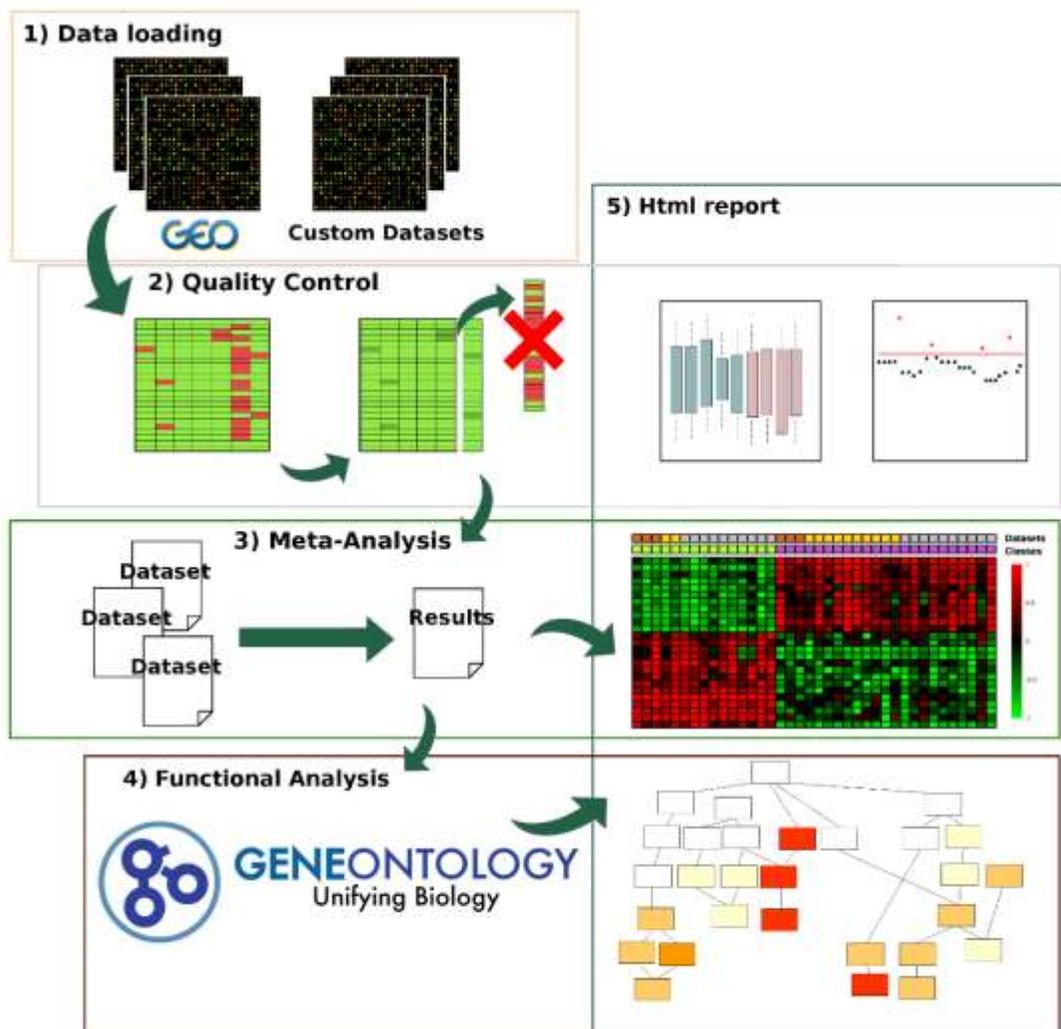


Figura 3. **Flujo de trabajo de ImGEO:** 1. Entrada de los datos, 2. Control de calidad de los datos ingresados, 3. Meta-análisis de los datos suministrados, 4. Análisis funcionales (opcional de acuerdo al objetivo del estudio), 5. Reporte de resultados. (Toro-Domínguez et al., 2019).

2.5.2. IRIS-EDA

Interactive RNA-Seq Interpretation System for Expression Data Analysis (IRIS-EDA), es una plataforma interactiva online de acceso gratuito. Permite el procesamiento y análisis de datos completos de RNA-seq. en un flujo de trabajo continuo. IRIS-EDA está integrado por

Differential Gene Expression (DGE), a través de uno de los tres paquetes de R ampliamente utilizados, DEseq2, edgeR y Limma.

Su flujo de trabajo (fig. 4) consiste en, (i) la entrada de dos conjuntos de datos que, de tipo, matriz de estimación de expresión génica (EEM), b) matriz de condición con niveles de factor de EEM o, c) matriz de longitud de gen. En este primer paso, se hace la normalización de los datasets, en caso de requerirse. (ii) Proceso de análisis por descubrimiento, basado en algoritmos de análisis y visualización de la información de EEM, para identificación de patrones o tendencias en los datos. (iii) Análisis de expresión diferencial (DEA), ejecutado según al paquete escogido (DEseq2, edgeR o limma). (iv) Generación de resultados del análisis por descubrimiento y del análisis de expresión diferencial (Fernandez Chinguel et al., 2019).

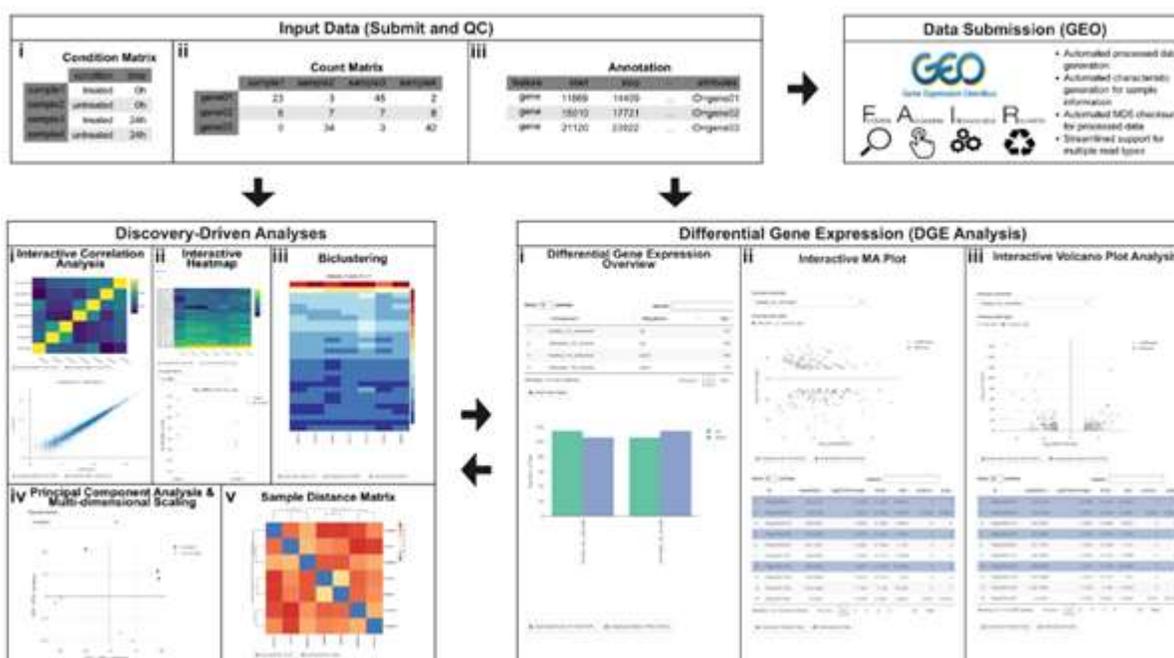


Figura 4. Flujo de trabajo de IRIS-EDA: sección de ingreso de datos propios desde servidor Input Data (Submit and QC), así como desde un repositorio por medio del ID. GEO en Data Submission (GEO). Una sección dedicada al análisis de los datos, Discovery-Driven Analyses. Y una sección dedicada al

análisis de expresión diferencial, Differential Gene Expression (DGE Analysis)
(<http://bmb1.sdstate.edu/IRIS/>).

3. OBJETIVOS

3.1. Objetivo general

Realizar un meta-análisis de texto e integrar datos públicos de expresión génica, para proponer rutas y genes en la disfunción endotelial inducida por virus, así como su posible relación con una transición endotelial a mesenquimal.

3.2. Objetivos específicos

Respecto de la realización de una revisión sistemática y un meta-análisis para integrar datos públicos de expresión génica y proponer rutas y genes en la disfunción endotelial inducida por virus, así como su posible relación con una transición endotelial a mesenquimal, éste trabajo pretende específicamente:

3.2.1. Definir el set de datos con los estudios relevantes en disfunción endotelial y transición endotelial a mesenquimal inducidas por virus, para el meta-análisis.

3.2.2. Determinar el protocolo más adecuado para el meta-análisis, a partir de la comparación de las relaciones de flujo de los softwares más utilizados.

3.2.3. Identificar los genes relevantes en la disfunción endotelial y la transición endotelial a mesenquimal inducida por virus mediante meta-análisis de cada proceso.

3.2.4. Especificar los genes y/o moléculas en las diferentes rutas dentro de los procesos y proponer un posible mecanismo en la EndMT

4. METODOLOGÍA

4.1. Definición de datasets y estudios relevantes

4.1.1. Rastreo y selección de artículos científicos

La búsqueda de documentos científicos se organizó en cuatro niveles que corresponden a: (i). La disfunción endotelial asociada al virus del dengue, (ii). La disfunción endotelial asociada a otros virus y (iii). La transición endotelial a mesenquimal asociada con virus u otros procesos y (iv). La transición endotelial a mesenquimal en la infección del dengue.

La búsqueda se hizo en cuatro diferentes repositorios científicos, NCBI PubMed, EMBL-EBI (European Bioinformatic Institute), Scopus y ScienceDirect. Se emplearon ecuaciones de búsqueda con términos clave para cada nivel (tabla 2). Posteriormente, y según la necesidad, se agregaron términos en cada plataforma, que contribuyeron con la exclusión de objetos de estudio no relacionados con el enfoque definido. Luego del proceso de exclusión por términos clave, se agregaron filtros propios de cada plataforma, como: tipo de documento, (article, journal article, full text, final stage, open access, research article, literature); año de publicación (2011-2021); especie (*Homo sapiens*, Human); idioma (español e inglés); hasta obtener un compendio de artículos a fin con esta investigación.

Tabla 2. Ecuaciones de búsqueda. Los términos en negrita comprenden la ecuación base de la búsqueda y los demás términos aluden a los excluyentes utilizados. Se empleó el comodín () para mayor exactitud de la búsqueda.*

Nivel	Repositorio	Ecuación de búsqueda con excluyentes
i	NCBI	(Endothel* AND Dysfunction*) AND Dengue (Viral OR Virus) NOT review NOT vaccine NOT (animal* OR mouse OR mice)
	Scopus	(endothel* AND dysfunction*) AND (dengue AND (virus OR viral)) AND NOT review AND NOT vaccine AND NOT

	endogenous AND NOT vector* AND NOT covid AND NOT (mice OR mouse OR animal*)
ScienceDirect	Endothelial dysfunction AND dengue virus NOT review NOT vaccine NOT endogenous NOT vector NOT animal NOT lymphocyte NOT transplantation
EMBL	(endothel* AND dysfunction*) AND dengue (viral OR virus) NOT review NOT vaccine NOT endogen* NOT vector NOT (animal* OR mouse OR mice) NOT blood
NCBI	(Endothel* AND Dysfunction*) AND (Viral OR Virus) NOT REVIEW NOT ENDOGENOUS NOT vectors AND infect* NOT vaccine NOT COVID-19 NOT (mice OR Mouse OR animal*) NOT transplantation
ii	(endothel* AND dysfunction*) AND (virus OR viral) AND NOT review AND NOT vaccine AND NOT endogenous AND NOT vector* AND NOT covid AND NOT (mice OR mouse OR animal*) AND NOT mitochondria* AND NOT blood AND NOT lymphocyte*
ScienceDirect	Endothelial dysfunction AND virus NOT review NOT vector NOT endogenous NOT transplantation NOT animal NOT mouse OR mice
EMBL	(endothel* AND dysfunction*) AND (viral OR virus) NOT review NOT endoge* NOT vector* AND infect* NOT vaccine NOT transplant* NOT cancer NOT (rat OR mice OR mouse) NOT animal* NOT mitochond* NOT lymphocyt* NOT blood NOT (clinical treatment) NOT covid
NCBI	(Endothel* AND to AND Mesenchy* Transition* OR EndMT OR EndoMT) AND (Viral OR Virus) NOT REVIEW NOT (mice OR Mouse OR animal*)
iii	(Endothel* to Mesenchy* Transition OR EndMT OR EndoMT) AND (virus OR viral) AND NOT review AND NOT vaccine AND NOT endogen* AND NOT vector* AND NOT (animal* OR mouse OR mice) AND NOT transplant* AND NOT

	lymphocyte* AND NOT blood AND NOT mitochondria* AND NOT covid
ScienceDirect	(Endothelial to Mesenchymal Transition) AND virus NOT vector NOT transplantation NOT endogenous NOT animal NOT mitochondria NOT epithelial NOT cancer
EMBL	(Endothel* to Mesenchy* Transition*) AND (virus OR viral) NOT review NOT vector NOT endogen* NOT vaccine NOT transplant* NOT (rat* OR mouse OR mice) NOT animal*
NCBI	(Endothel* to Mesenchy* Transition OR EndMT OR EndoMT) And Dengue
Scopus	((endothel* AND to AND mesenchy* AND transition OR endmt OR endomt) AND (dengue)) AND NOT proteomic*
iv	(Endothelial to mesenchymal transition OR EndMT OR EndoMT) AND (dengue) NOT review NOT therapeutic NOT (animal OR mouse OR mice)
EMBL	(Endothel* to Mesenchy* Transition* OR EndMT OR EndoMT) AND (dengue)

En cada nivel se tomó el total de artículos seleccionados finalmente, los cuáles se sometieron a un cribado manual en cada plataforma. De cada documento se tuvo en cuenta su respectivo resumen, en el que se analizó el enfoque de la investigación, el linaje celular objeto de estudio y su posible relación con alguno de los procesos de interés. La síntesis del procedimiento se muestra en la figura 5.

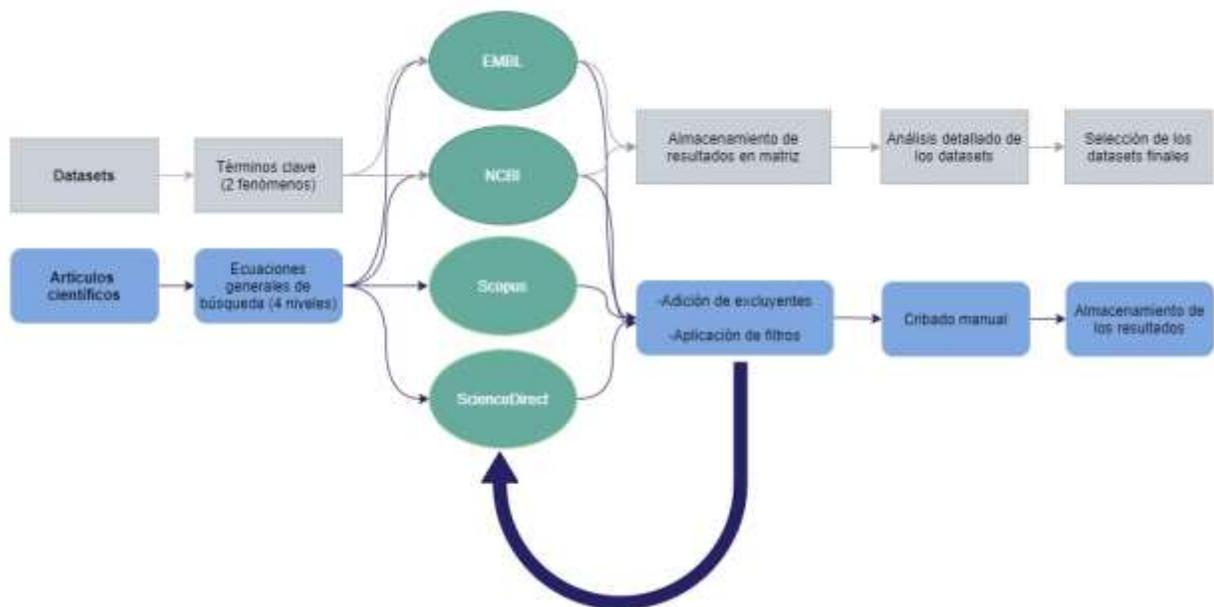


Figura 5. Flujo de búsqueda y selección de artículos científicos y datasets. Pasos empleados en la selección de la literatura y los datasets empleados en los análisis.

4.1.2. Selección de datasets para metaanálisis

La búsqueda de los datasets se hizo en las plataformas de NCBI-GEO (Gene Expression Omnibus) y EMBL-ENA (European Nucleotide Archive). Se emplearon ecuaciones de búsqueda con las palabras clave: Endothelial cell, Endothelial Dysfunction, Dengue Virus, Virus, Severe Dengue, RNAseq, MicroArray, Endothelial to Mesenchymal Transition (EndMT or EndoMT), Vascular Leakage.

El cribado de los datasets resultantes se hizo de forma manual en cada plataforma. Para este cribado se tuvo en cuenta, el tipo de experimento o estudio realizado, el tipo de tejido objeto de estudio, el tipo de tecnología de secuenciación utilizado y tratamiento aplicado. Finalmente, se seleccionados los datasets, se descargaron en el servidor para el posterior tratamiento de los mismos.

4.2. Protocolo para metaanálisis

4.2.1. Datasets de prueba

Los datasets para las pruebas de los softwares se tomaron de un metaanálisis hecho por (Moradifard et al., 2018), en un análisis de microARN (miARN) y perfiles de expresión génica (ARNm) en la enfermedad del Alzheimer. En dicho estudio, se utilizaron 6 dataset de ARNm y 1 de miARN, para un total de 264 muestras (113 de control y 151 casos). Para este análisis, se tomaron al azar tres de datasets de ARNm (GSE1297, GSE12685 y GSE16759), con un total de 54 muestras (21 de control y 32 de casos), la determinación de casos y controles se hizo conforme al estudio original.

4.2.2. Análisis de softwares

Para el análisis de cada software, se ingresaron los datasets y se conjugaron todas las posibles variables de los parámetros disponibles en cada plataforma, se evaluó el grado de complejidad en la interfaz, el método para ingreso de los datos, el tiempo de análisis de los datasets, conjugación de parámetros y la capacidad de manipulación de los mismos. Se hizo la comparación de los resultados de cada metaanálisis ejecutado. Posteriormente, se determinaron los atributos de interés y con ellos se construyó un árbol de decisión para la selección del protocolo más acertado para el metaanálisis.

4.3. Identificación de moléculas y genes

4.3.1. Revisión sistemática

Teniendo en cuenta los documentos finalmente seleccionados de los diferentes repositorios, se analizó cada uno de ellos rigurosamente bajo los intereses de nuestra investigación.

De cada documento incluido en la revisión sistemática, se hizo un resumen que destacó el linaje celular, tipo de tratamiento y tecnología usada por los autores. Además, se incorporó por

secciones (experimentos), los resultados obtenidos en cada caso y se resaltó la importancia, en este caso de las moléculas o rutas, el cambio expresado, su nivel y el tejido en el que se presentó.

El compendio de datos encontrados en los diferentes documentos analizados, se almacenó en una matriz externa por proceso (Libro1, “Documentos y moléculas”. Anexo)

4.3.2. Metaanálisis

4.3.2.1. Metaanálisis ImaGEO

La comparación de los datasets se hizo por afinidad de fenómeno. en la plataforma online ImaGEO (<https://imageo.genyo.es/>). Se ingresaron los datasets correspondientes a la disfunción endotelial (GSE34628 y GSE142762) en infección con virus del dengue y citomegalovirus, respectivamente. El método de comparación fue de valor-p, con el modelo de Stouffer (Darlington & Hayes, 2000). Se eligió un umbral ajustado de 0.05 para el valor de significancia y un 10% de valores atípicos permitidos. Para la selección de los casos y controles se asignaron conforme a la etiqueta original de cada estudio, comparando 10 muestras en total, 5 muestras de control (3 de GSE34628 y 2 de GSE142762) y 5 muestras de tratamiento (3 de GSE34628 y 2 de GSE142762).

4.3.2.2. Análisis de datasets por GEO2R

Para el aprovechamiento de toda la información obtenida en las búsquedas y distintos procesos de cribado, y, debido a la falta de datasets contrastante con los datos de EndMT en infección de virus de Kaposi (GSE22522), se hizo un análisis individual de la expresión diferencial de genes en los tres datasets desde la plataforma NCBI, en la extensión GEO, GEO2R.

Para el análisis individual, se ingresó el identificador o *Id* de dataset (GEO acceso), se definieron los grupos de comparación como “control” e “infected” para las muestras de control y casos, respectivamente. Posteriormente, se categorizaron las muestras en cada grupo según etiqueta. Para la definición de los parámetros, se usó igual en cada análisis de cada dataset, tomando,

- Ajuste al valor P: Benjamini & Hochberg (1995) (False discovery rate); seleccionado como predeterminado, ofrece un buen equilibrio entre el descubrimiento de genes estadísticamente significativos y la limitación de falsos positivos.
- Transformación de los datos por log: Yes; que indica a la plataforma una transformación Log_2 .
- Pesos de precisión de limma (vooma): Yes; la herramienta calcula ponderaciones apropiadas a nivel de observación, a partir de la relación entre media y varianza.
- Forzar normalización: Yes; indica que todas las muestras seleccionadas tengan una distribución de valores idéntica.
- Categoría para la muestra de resultados: Submitter supplied (suministrada por el operador).
- Visualización de gráficos: Nivel de significancia = 0.05;

Ejes = control vs infected

Se ejecutó el análisis y los resultados se descargaron como tabla en formato TSV.

4.3.2.3. Comparación de resultados de GEO2R

La comparación de los resultados de los análisis de expresión diferencial hecho en GEO2R, se hizo mediante Rstudio, comparando GSE22522_vs_GSE34628 y GSE22522_vs_GSE142762. Cada comparativo se almacenó como archivo.txt.

4.4. Análisis de vías de expresión de los genes diferencialmente expresadas

Los genes resultantes del análisis hecho en GEO2R se buscaron en las bases de datos KEGG PATHWAY Database: (<https://www.genome.jp/kegg/pathway.html>) y Reactome: (<https://reactome.org/PathwayBrowser/#/>) a fin de identificar las rutas a las que pertenecen y ver la agrupación entre ellas, así como, la influencia de estas rutas en los fenómenos de estudio.

4.4.1. Construcción de red de asociación de moléculas

Las moléculas encontradas en la revisión sistemática se asociaron con el tejido en el que se expresan, el cambio efectuado, la lesión que las induce y la ruta en la que se encuentran involucradas. Esta matriz se modificó como matriz de adyacencia, de base nodo-atributo, y mediante el paquete ‘igraph’ de Rstudio, se construyó la red de relación entre las moléculas y los demás atributos, la visualización y limpieza de la red se hizo en el programa Cytoscape3.9.0.

4.4.2. Identificación de moléculas relevantes en los fenómenos

La red de correlación contribuye con una visión general de los objetos involucrados en los fenómenos. Comparamos la presencia de las moléculas y la relación que contiene en los diferentes tejidos donde se produce la disfunción endotelial, para ver aquellas que tienen mayor influencia por tejido y por ruta de expresión.

También, se comparó la frecuencia de aparición de las moléculas en los fenómenos y las diferentes interacciones en las moléculas con mayor mención.

5. RESULTADOS

5.1. Cribado de artículos y datasets

La revisión sistemática consideró cuatro niveles para una mejor interpretación: (i) Disfunción endotelial y virus del dengue, (ii) Disfunción endotelial y virus, (iii) Transición endotelial a mesenquimal y virus, (iv) Transición endotelial y virus del dengue.

La búsqueda realizada en los cuatro repositorios científicos (PubMed, ScienceDirect, Scopus y EMBL), se sometió a una reestructuración de las ecuaciones de búsqueda para mayor especificidad de los resultados (tabla 1, anexos). En ella se dispuso de la adición de los términos de exclusión, que dependió de los resultados preliminares a la búsqueda por cada repositorio.

Posteriormente, el análisis manual de los estudios, dio como resultado la selección de 18 artículos para el primer nivel, 5 para el segundo, 7 para el tercero y cero (0) artículos para el cuarto nivel.

La tabla 3 muestra el resultado consolidado desde la ecuación sintética, que abarca la selección sin restricción y sus respectivos porcentajes.

Tabla 3. Síntesis porcentual del cribado en los repositorios científicos para la revisión sistemática. De acuerdo a las ecuaciones sintéticas o basales se indica, el número de estudios resultantes de la búsqueda sin restricción (SR), el número de estudios resultantes de la ecuación sintética con restricciones (CR), el estado porcentual de estudios seleccionados (%SRS) y de los estudios excluidos (%CREx). Los documentos finalmente seleccionados (FS) son producto de la aplicación de filtros a CR

desde cada repositorio, indicando el porcentaje de estudios excluidos (%FSEx) por los mismos y el porcentaje de seleccionados (%FS).

Ecuación sintética	Repositorio	SR	CR	% SRS	% CREx	FS	% FS	% FSEx
Dysfunction Endothelial + Dengue Virus	PubMed	72	28	38,89	61,11	11	39,29	60,71
	Scopus	104	39	37,50	62,50	24	61,54	38,46
	Science Direct	1.366	30	2,20	97,80	6	20,00	80,00
	EBI	95	31	32,63	67,37	20	64,52	35,48
		128				61	47,66	52,34
Dysfunction Endothelial + Virus	PubMed	1.604	308	19,20	80,80	85	27,60	72,40
	Scopus	62.868	360	0,57	99,43	107	29,72	70,28
	Science Direct	30.503	120	0,39	99,61	22	18,33	81,67
	EBI	1.697	143	8,43	91,57	73	51,05	48,95
		931				287	30,83	69,17
EndMT + Virus	PubMed	22	8	36,36	63,64	6	75,00	25,00
	Scopus	16.922	53	0,31	99,69	20	37,74	62,26
	Science Direct	7.440	25	0,34	99,66	5	20,00	80,00
	EBI	1.762	179	10,16	89,84	12	6,70	93,30
		265				43	16,23	83,77
EndMT + Dengue Virus	PubMed	0	0	0,00	0,00	0	0,00	0,00
	Scopus	1	0	0,00	100,00	0	0,00	0,00
	Science Direct	186	3	1,61	98,39	1	33,33	66,67
	EBI	7	7	100,00	0,00	1	14,29	85,71
		10				2	20,00	80,00

Alrededor de 124.649 registros totales fueron inspeccionados, considerando que de este gran número se repitieron algunos títulos. De ellos, 1.336 artículos fueron extractados con la restricción, para llegar a un número final seleccionado de 393 documentos.

El porcentaje de exclusión promedio para todas las ecuaciones y repositorios fue del 98.93%, el 1.03% extractados con las restricciones y finalmente solo el 0.31% seleccionado.

En cuanto a la selección de los datasets para el metaanálisis, se realizó una búsqueda en los repositorios de datos de NCBI GEO (Gene Expression Omnibus) y EMBL-ENA (European Nucleotide Archive), bajo búsquedas manuales, usamos como palabras clave, términos relacionados con los procesos de Disfunción Endotelial y Transición endotelial a mesenquimal, virus y virus del dengue específicamente.

Los criterios de inclusión de los resultados obtenidos de las búsquedas se hicieron desde la misma plataforma analizando, inicialmente, que fuesen datos de expresión diferencial génica,

tratamiento a células endoteliales o EndMT, la tecnología de secuenciación de tipo RNA seq. O Microarreglo, relación con infección viral.

El cribado permitió la preselección de 31 datasets (la información complementaria se encuentra en la tabla 2, anexos), de los cuales se excluyeron por características incompatibles con la necesidad de nuestro estudio (tabla 4), resaltando que el tratamiento experimental a células no endoteliales fue el mayor excluyente, con un 41.94% del total de datasets encontrados, seguido por estudios con tratamientos de tipo farmacológico, ensayos clínicos, reportes de caso, entre otros (tratamientos incompatibles), con un 25.81% de los datasets totales.

Tabla 4. Síntesis porcentual del cribado de datasets para el metaanálisis. De 31 datasets preseleccionados, se excluye el 90.32% y finalmente, solo el 9.68% es seleccionado.

Total	Criterios de exclusión de los datasets						Seleccionados
	Otro tipo de célula	Especie	Tipo de dato	Datos no disponibles	Sin respaldo de publicación	Tratamientos incompatibles	
31	13	1	3	1	2	8	3
%	41,94	3,23	9,68	3,23	6,45	25,81	9,68

Finalmente, solo 3 datasets (9.68%) del total, cumplió con la mayoría de los criterios mínimos requeridos para la inclusión en la investigación. Dos de los datasets seleccionados corresponden a disfunción endotelial en infección viral. **GSE34628**, un estudio realizado por (Dalrymple & Mackow, 2012), titulado “*Endothelial Cells Elicit Immune-Enhancing Responses to Dengue Virus Infection*”, **GSE142762**, estudio realizado por (Zhang et al., 2020), con título “*Human cytomegalovirus promoting endothelial cell proliferation by targeting regulator of G-protein signaling 5 hypermethylation and downregulation*” y un dataset relacionado con transición endotelial a mesenquimal en la infección con el virus de Sarcoma de Kaposi, **GSE22522**, realizado por (Cheng et al., 2011), titulado “*KSHV-Initiated Notch Activation Leads*

to Membrane-Type-1 Matrix Metalloproteinase-Dependent Lymphatic Endothelial-to-Mesenchymal Transition”.

Sin embargo, aclaramos que, los datasets relacionados con la disfunción endotelial (GSE34628 y GSE142762), presentan una incongruencia estadística en la presentación de las muestras, correspondiente a la falta de réplicas reportadas.

El estudio identificado con el serial GSE34628, involucra la infección en células endoteliales con virus del dengue, donde evalúan la expresión génica mediante la tecnología Matriz Affymetrix Human Genome U133 Plus 2.0 (MicroArray), en tres lapsos de tiempo diferente (12, 24 y 48 horas post infección). El dataset presenta un comparativo entre la expresión en células endoteliales de cordón umbilical humano (HUVEC) infectadas (3 muestras), con sus respectivos controles (3 muestras) de células no infectadas, en los mismos tiempos de análisis. El estudio no reporta réplicas en sus muestras.

El estudio identificado con el serial GSE142762, evalúa la función endotelial en la infección del citomegalovirus humano (HCMV), en presencia de hiperglucemia (HG) e hiperlipidemia (ox-LDL), en la hipertensión esencial. El estudio presenta 4 muestras en el estudio, que refieren a: 1. Células tratadas con HCMV + HG + ox-LDL, 2. células infectadas con HCMV, 3. células tratadas con HG + ox-LDL y 4. células sin tratamiento, el análisis se hizo mediante la tecnología Matriz de expresión de genes humanos Affymetrix (MicroArray). El estudio no presenta réplicas en sus muestras.

El estudio identificado con el serial GSE22522, corresponde a una investigación de células endoteliales linfáticas (LEC) infectadas por el virus del herpes del Sarcoma de Kaposi. El estudio hecho con tecnología Affymetrix Human Genome U133 Plus 2.0 Array. Este dataset comprende

6 muestras, 3 que corresponden a células infectadas y 3 muestras control. El estudio evalúa los cambios de expresión génica posterior a la infección.

5.2. Análisis de softwares

Para el análisis de los softwares usados en el metaanálisis, fueron seleccionados ImaGEO de genyo, NetworkAnalyst e IRIS-EDA (An Integrated RNA-Seq Interpretation System for gene Expression Data Analysis).

Los datos que sirvieron de prueba se tomaron de un estudio previo hecho por (Moradifard et al., 2018) en un metaanálisis de estudios relacionados a la enfermedad de Alzheimer. De este estudio, se tomaron tres datasets (GSE1297, GSE12685 Y GSE16759), de donde se compararon 54 muestras (21 de control y 32 de casos).

A continuación, se describen los resultados obtenidos con cada software

5.2.1. NetworkAnalyst 3.0

El acceso a la plataforma es online, en la dirección <https://www.networkanalyst.ca/>. Cuenta con una plataforma actualizada, numerosas funciones y herramientas. La subsección metaanálisis se encuentra en la opción, análisis múltiple de datos.

Para iniciar, se ingresa por la opción '*Multiple Gene Expression Tables*'. Los datos deben estar descargados en el ordenador, en formato de texto delimitado por tabulaciones (.txt) o en el formato comprimido (.zip).

En cuanto al formato del dataset, se debe hacer un ajuste previo en el texto, eliminando los metadatos que puedan existir de encabezado. La primera línea debe corresponder al ID de cada muestra, debe estar precedida por la etiqueta de categoría #NAME, cada muestra se separa por

una tabulación. La segunda línea corresponde a la etiqueta de cada muestra, que se precede con la etiqueta de categoría #CLASS, las etiquetas se deben separar por una tabulación como se muestra en el ejemplo en la (figura 6), y deben concordar con la muestra indicada.

```
#NAME "GSM318211"      "GSM318212"      "GSM318213"      "GSM318310"      "GSM318311"      "GSM318312"
"GSM318313"      "GSM318314"      "GSM318317"      "GSM318320"      "GSM318321"      "GSM318442"
"GSM318443"      "GSM318445"
#CLASS CONTROL      CONTROL      CONTROL      CONTROL      CONTROL      CONTROL
CONTROL      CONTROL      CASE      CASE      CASE      CASE      CASE      CASE
```

Figura 6. Ejemplo del ajuste al formato de cada dataset.

El tamaño de cada dataset no debe superar los 50 Megabytes. El tiempo de carga es directamente proporcional al tamaño del archivo.

El tratamiento de cada dataset en la plataforma se hizo por separado. El resumen del tratamiento y los parámetros seleccionados en la prueba, se muestran en la tabla 5.

Tabla 5. Parámetros usados para el tratamiento de los datos en NetworkAnalyst.

NetworkAnalyst				
		Datasets		
Parameters		GSE1297	GSE12685	GSE16759
ID conversion	Specify organism	<i>Homo sapiens</i>	<i>Homo sapiens</i>	<i>Homo sapiens</i>
	Gene Id or Probe platform	Affymetrix Human Genome U133A Array	Affymetrix Human Genome U133A Array	Affymetrix Human Genome U133 Plus 2.0 Array

Normalizati on	Data type	Microarray data (intensities)	Microarray data (intensities)	Microarray data (intensities)
	Normalizati on procedure	Variance Stabilizing Normalization (VSN)	Variance Stabilizing Normalization (VSN)	Variance Stabilizing Normalization (VSN)
	Perform auto-scaling	No	No	No
DE analysis	Contrast	Control=vs=Ca ses	Control=vs=Ca ses	Control=vs=Ca ses
	Set p-value (FDR) cutoff	0.05	0.05	0.05

Los resultados del análisis de los datasets pueden ser visualizados en la misma pestaña online, y, permite descargar solo aquellos que tengan relevancia con el interés del estudio.

El reporte ofrece:

- **Control de calidad**

- Un gráfico de análisis de componentes principales, que ofrece una visualización del comportamiento de los controles_vs_casos.
- Gráfico de densidad, que muestra la distribución relativa de diferentes recuentos en cada grupo.

- **Metaanálisis**, ofrece diferentes métodos y modelos, las variables se pueden manipular en la misma selección. El resultado de las pruebas con los diferentes métodos y modelos, se presentan en la tabla 6.
- **Genes significativos**, ofrece el listado de genes que se encontraron con cambios estadísticamente significativos, valores por dataset y valores combinados.
- **Descripción general del análisis**, cuenta con un gran número de herramientas como redes, que permite la asociación de los genes o moléculas encontrados a algunas rutas de expresión. También brinda mapas de calor, gráficos de volcán, diagrama de Venn o de acordes.
- **Personalización de redes**, que es el espacio donde se pueden configurar y aplicar los cambios más representativos como visualización de los resultados.
- **Visualización de redes**, una pestaña que configura parámetros estéticos de las redes o gráficos. Todas las redes y gráficos hechos se pueden descargar directamente
- **Descarga**, permite la descarga de datos, metadatos, script de R, datos conjugados, análisis de relación (en caso de haberlo requerido), gráficos del control de calidad.

El análisis de los 3 datasets estudiados no presentó genes diferencialmente expresados, bajo ningún modelo (tabla 6).

Tabla 6. Resultados del metaanálisis ejecutado en NetworkAnalyst. Descripción de los genes diferencialmente expresados y el tiempo de ejecución en cada modelo.

NetworkAnalyst			
Método	Modelo	Genes diferencialmente expresados	Tiempo de ejecución
Combining P values	Fisher	No encontrados	0'35"
Combining P values	Stouffer	No encontrados	0'27"

Combining Effect Sizes	Fixed Effect Model	No encontrados	0'07''
Combining Effect Sizes	Random Effect Model	No encontrados	0'04''
Vote counting	“2” sets the minimal number of votes	No encontrados	0'04''
Direct merging	-	No encontrados	0'01''

5.2.2. Análisis de ImaGEO

El acceso a la plataforma es online (<https://imageo.genyo.es/>). Su acceso es rápido, fácil al control del usuario. Cuenta con una interfaz amigable y de diseño dinámico. El portal de ayuda ofrece una dirección ejemplificada paso a paso desde el ingreso de los datos, hasta la interpretación de los resultados.

Para el metaanálisis, se ingresa el ID GEO de los datasets (Máximo 10). Para el caso de trabajar con datos externos, tiene como ventaja que no requiere el almacenamiento o tratamiento de los mismos como bases de datos. Sin embargo, también ofrece la opción de cargar datos propios.

La configuración de los parámetros para el metaanálisis se ejecuta desde la misma pestaña, cada método despliega las opciones de modelos. Manualmente se puede configurar el porcentaje de valores atípicos permitidos y el umbral de significancia con el que se van a medir los datos. Así mismo, se puede modificar el nombre de los grupos comparativos. Si se requiere es permitido seleccionar la opción de análisis funcional, donde se hace un análisis ontológico de los datos y los genes resultantes. El reporte se puede almacenar directamente en

el correo electrónico del usuario, antes de ejecutar el análisis se debe ingresar la dirección e-mail.

Para la asignación de las muestras de cada dataset a los grupos de casos y controles se hace por método de arrastre. Posteriormente se ejecuta el metaanálisis.

La plataforma permite la visualización del reporte de resultados en una pestaña emergente y se envía copia directamente al correo (en caso de haberse solicitado previamente).

El reporte ofrece:

1. Resumen: **Resumen del flujo de ejecución; información de los datasets** ingresados, que contiene el número de casos y control por dataset; **resumen del metaanálisis** con los parámetros usados.
2. Control de calidad entre estudios: **diagrama de cajas y bigotes** de cada muestra, de cada dataset. Gráfica del porcentaje de **'missing values'** en la que las muestras superiores a dicho valor, son ignoradas. Tabla de comparación de las muestras antes y después del **control de calidad**, mostrando el número de genes imputados.
3. Resultados: **resumen de genes diferencialmente expresados**, de cada dataset y de metaanálisis; **resultado del metaanálisis**, presentando el ID del gen, p-value del gen (pval), p-value ajustado del gen (fdr_pval), valor del cambio de expresión del conjunto de datasets, representado por FoldChange (FC_mean) y el valor de FoldChange de cada gen, en cada dataset; **mapa de calor** del cambio en la expresión de los genes.
4. Código de ejecución y archivos para repetir el análisis, en formato descargable.

El resumen del metaanálisis ejecutado con diferentes pruebas se presenta en la tabla 7

Tabla 7. **Resultados del metaanálisis realizado en ImaGEO.** Se indica el número de genes diferencialmente expresados en cada dataset, la conjugación promedio de los 3 datasets y el tiempo de ejecución tardado en cada modelo.

ImaGEO						
Método	Modelo	Genes diferencialmente expresados			Promedio metaanálisis	Tiempo de ejecución
		GSE129 7	GSE1268 5	GSE1675 9		
Effect Size	Fixed	6	2	0	0	1'00"
Effect Size	Random	6	2	0	0	0'59"
P-value	maxP	6	2	0	3	1'14"
P-value	minP	6	2	0	0	1'13"
P-value	Fisher	6	2	0	0	1'11"
P-value	Stouffer	6	2	0	0	1'13"

5.2.3. Análisis de IRIS-EDA (An Integrated RNA-Seq Interpretation System for gene Expression Data Analysis)

El comportamiento de esta plataforma online, no permitió su análisis debido a numerosas desconexiones con el servidor. IRIS-EDA presentó inestabilidad constante, lo que anuló el comando de carga de los datos.

Sin embargo, se destaca que su diseño web indica ser de fácil manipulación. Los datasets se deben cargar desde el ordenador en formato CSV con un tamaño no superior a 50KB y solo permite la comparación de dos datasets por sesión.

5.2.4. Árbol de decisión

La figura 7, comprende las opciones viables para hacer metaanálisis de expresión diferencial de genes, que es el primer nodo de decisión y tiene tres diferentes opciones. Los nodos de probabilidad, representados por un círculo gris oscuro, indican los softwares evaluados por su utilidad en el metaanálisis (IRIS-EDA, ImaGEO y NetworkAnalyst). Estos nodos tienen la característica de poseer generalmente múltiples resultados, aunque en la mayoría de los casos son de relación binaria (*Árbol de decisión*, s/f). Estos resultados expresados en relación única, para los casos de contraste con otros nodos de probabilidad, binaria o múltiple. En el caso de relación binaria, se identifican como nodos de decisión (cuadro azul) o directamente como nodo terminal (cubo, azul esmeralda).

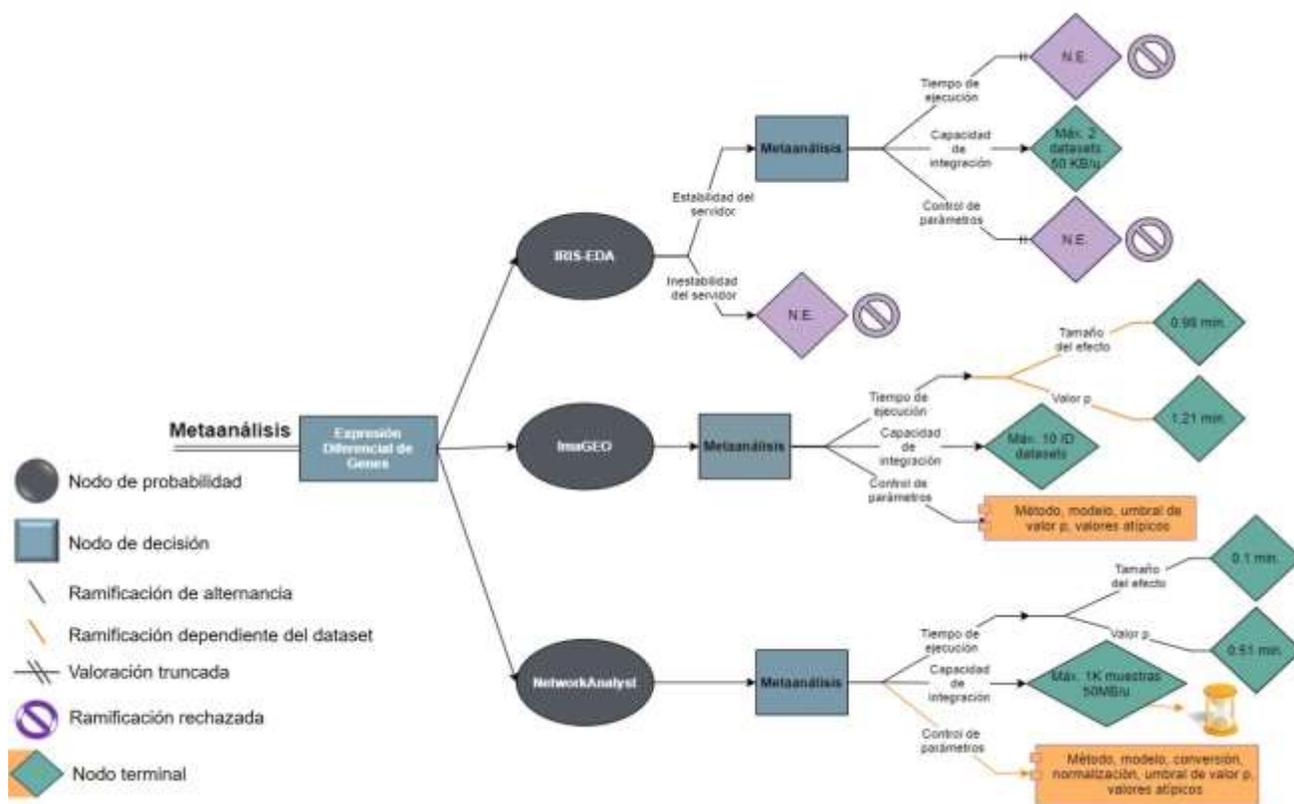


Figura 7. *Árbol de decisión*. Se presenta un esquema que identifica los atributos más relevantes (tiempo de ejecución, capacidad de integración de datos, control de parámetros), en la ejecución de un metaanálisis. El nodo de probabilidad, corresponde a una opción que mostrará múltiples resultados inciertos. El nodo de decisión, indica una elección compuesta por diferentes variables. La ramificación

alterna indica diferentes opciones dentro de un nodo. La ramificación dependiente del dataset, indica que alguno de los procesos descritos es directamente influenciado por el objeto (dataset). La valoración truncada corresponde a un proceso que no se puede evaluar. La ramificación rechazada alude a una variable que se descarta. Y el nodo terminal, se refiere a un resultado definido. N.E. No explorado.

Los nodos de decisión comprenden el metaanálisis como un sistema de atributos con funciones específicas, en este caso, la integración de datos de expresión génica. Los atributos que se encuentran en cada nodo de decisión, fueron los mismos para cada caso, a fin de facilitar la comparación y evaluación de cada nodo de probabilidad (software). Los atributos evaluados fueron, tiempo de ejecución en metaanálisis, capacidad de integración de datos y control de los parámetros para el metaanálisis.

IRIS-EDA, presentó una relación binaria. Por una parte, un nodo terminal caracterizado por la inestabilidad en el servidor, que imposibilitó la terminación de la evaluación y, por otra parte, un nodo de decisión, donde únicamente se pudo evaluar la interfaz de inicio, indicando que, la capacidad de integración de datos no soporta más de dos datasets por sesión y un tamaño superior a 50 Kilobytes de cada uno. En cuanto a los atributos de tiempo de ejecución y control de parámetros, son ramificaciones rechazadas (signo de bloqueo), por valoración truncada (\neq) a raíz de la inestabilidad de la plataforma.

ImaGEO, presentó una relación unitaria, donde su nodo de decisión (metaanálisis), comprendió la ramificación 'tiempo de ejecución' como dependiente del dataset, indicando que, cada metaanálisis ejecutado es dependiente del tamaño de los datos analizados, sin embargo, el tiempo promedio de los dos métodos no excedió los 66 segundos, siendo el método de comparación de los valores-p, el que empleó mayor tiempo (73 segundos).

La ramificación que corresponde a la capacidad de integración, indica como nodo terminal que ImaGEO soporta un máximo de 10 datasets, los cuales se admiten con solo el ID de GEO (Gene Expression Omnibus) o cargados desde el servidor (el tiempo de carga se hace dependiente del tamaño de los datasets).

El control de parámetros, como nodo terminal también, indica que la plataforma permite la manipulación del método, modelo, asignación del umbral de significancia (Threshold) y el porcentaje de valores atípicos permitidos (Allowed missing values %).

Por último, NetworkAnalyst presentó una relación unitaria, donde su nodo de decisión (metaanálisis), presentó una ramificación de alternancia en el tiempo de ejecución, concluyendo en dos nodos terminales. Para el método de tamaño del efecto se calculó un promedio en tiempo de 0.1 minutos (~6 seg) y para el método de comparación de valores-p, un promedio de 0.51 minutos (~31 seg), los datos exactos se muestran en la tabla 6.

La capacidad de integración de datos, indica que el software soporta una capacidad máxima de 1000 muestras (independiente del número total de datasets) y cada dataset cargado no puede exceder los 50 Megabytes, estos datos pueden cargarse desde el servidor en formato .txt o su versión comprimida en formato .zip. Esta ramificación genera una variable correspondiente al tiempo de carga que es directamente proporcional al tamaño de cada dataset.

La última ramificación que corresponde al control de parámetros, se presentó como dependiente de los datasets, debido a que el parámetro ‘conversión’ soporta cierta población limitada de tecnologías de secuenciación, usadas en la obtención de los datos. Otros parámetros que se encuentran disponibles en la plataforma, y que no son dependientes del dataset son,

método, modelo, normalización, umbral de significancia, porcentaje de valores atípicos permitidos.

5.3. Identificación de moléculas y genes

5.3.1. Moléculas de la revisión sistemática

De los artículos de disfunción endotelial en infección de virus del dengue, 18 en total, se analizaron los tratamientos de cada documento y se reportó un total de 122 moléculas con importancia en la infección del virus del dengue con influencia directa sobre el endotelio o plaquetas, que *in vivo* tienen relación en patologías inducidas por el virus del dengue. También se pudo observar que la proteína no estructural-1 (NS1), expresada por células infectadas, indujo cambios de expresión en 21 moléculas, mientras que la infección directa con el virus del dengue de diferentes serotipos (I y II), indujo el cambio de expresión en 16 moléculas.

En la tabla 8, se encuentra la síntesis de los atributos que presentaron las moléculas, teniendo en cuenta que, para el proceso de disfunción endotelial en la infección del dengue (A), algunas moléculas presentan diferentes relaciones con tejidos, como la quimiocina CXCL11, presente en riñón, pulmón, tejido cardíaco y umbilical, la molécula de adhesión intercelular ICAM-1, presente en tejido dérmico, pulmón y riñón, la interleucina IL-8 presente en el tejido dérmico, pulmonar, renal y cardíaco, y el factor de necrosis tumoral TNF- α , expresado en tejido umbilical, pulmonar, dérmico, renal, cardíaco y dérmico. Por tanto, a estas diferentes relaciones entre molécula y sus atributos se nombran ‘entradas’, presentes en la base de datos con sus asociaciones directas por fila.

Tabla 8. Propiedades de las moléculas integradas desde la literatura. i) disfunción endotelial en la infección del virus del dengue se reportaron 165 moléculas con diferente comportamiento, ii) disfunción endotelial en infección viral, se reportaron 24 moléculas y iii) fenómeno de Transición endotelial a

mesenquimal, reporta 122 moléculas; la frecuencia de los cambios por molécula (Change type) donde Up, se refiere a las moléculas sobre expresadas, Down, las moléculas subexpresadas, Present, se refiere a las moléculas que tiene influencia en el fenómeno sin cambios de expresión, Inhibited, alude a las moléculas inhibidas, P+, aquellas moléculas que se fosforilaron. La frecuencia por los tejidos más relevantes menciona, U.C. (cordón umbilical), Derm. (dermis), Lung (pulmón), Card. (cardiaco), Brain (cerebral), Int. (intestinal), LEC (células endoteliales linfáticas), Other (otros tejidos involucrados); las moléculas expresadas a nivel mRNA (ARN mensajero), protein (proteína) o both (ARN mensajero y proteína), N.I. (No identificado); los tipos de molécula reportados, Qc (quimiocinas), Enz (enzimas), NF (factor de necrosis), Gp (Glucoproteína), GF (factor de crecimiento).

	Change type		Tissue		Level		Molecule Type	
i	Up	113	U.C.	51	mRNA	4	Qc	32
	Down	16	Derm.	28			Enz	12
	Present	31	Lung	25	Protein	159	NF	9
	Inhibited	2	Card.	10			Gp	9
	P+	3	Brain	0	N.I.	2	GF	8
			Other	51			Other	95

ii	Up	21	U.C.	7	mRNA	0	Qc	2
	Down	3	Int.	0			Enz	2
	Present	0	LEC	0	Protein	16	NF	0
	Inhibited	0	Card.	10			Gp	0
	P+	0	Other	7	Both	8	GF	3
							Other	17

iii	Up	68	U.C.	30	mRNA	7	Qc	0
	Down	22	Int.	32			Enz	4
	Present	30	LEC	20	Protein	108	NF	2
	Inhibited	1	Card.	13			Gp	21
	P+	1	Other	27	Both	7	GF	13
							Other	82

A raíz de estas interacciones, se identificaron, para el proceso de disfunción endotelial en infección del dengue (i), 166 entradas de 106 moléculas únicas, para el proceso de disfunción endotelial en infección viral (ii), se registraron 24 moléculas y entradas, respectivamente y el para los hallazgos de los documentos de Transición Endotelial a Mesenquimal (EndMT) (iii), se registraron 122 entradas de 63 moléculas únicas.

A partir de las asociaciones por tejido se identificó que en el proceso (i) hay predominancia de las células endoteliales del cordón umbilical con un 30.91%, seguido por el tejido dérmico con un 16.96% del total de entradas por tejido y otros tejidos ya mencionados que comprenden 52.13%. En el proceso (ii), se encontró que el tejido predominante fue el tejido cardiaco con un 41.67% sobre el total. Y para el proceso (iii), predominó el tejido intestinal con un 26.23%, seguido de cerca por las células de cordón umbilical con el 24.59%, para indicar que el resto de tejidos presentes suman el 49.18%.

En los tres niveles estudiados, se pudo notar que los cambios de expresión de las moléculas ocurren en su mayoría a nivel de proteína, una minoría a nivel de ARN mensajero y para algunos casos el cambio de expresión se dio tanto a nivel de ARNm como de proteína.

Asociamos las moléculas por tipo funcional, así logramos determinar qué grupos son los más relevantes en los fenómenos. En el caso de la disfunción endotelial, los niveles (i) y (ii), las quimiocinas (Qc), enzimas (Enz) y factores de crecimiento (GF), fueron las familias de moléculas con mayor presencia, mientras que para EndMT (iii), las glicoproteínas (Gp) y factores de crecimiento (GF), fueron las que presentaron mayor incidencia.

5.3.2. Genes resultantes del metaanálisis

El análisis de expresión diferencial comparó los dos datasets relacionados con la infección en virus del dengue y citomegalovirus (GSE34628 y GSE142762), a través de la plataforma ImGEO por el modelo de Stouffer. Este análisis no mostró valores atípicos por encima del rango asignado en los parámetros (10%). Además, el análisis de cajas y bigotes indicó una distribución normal de los datos en ambos datasets (Figura 8).

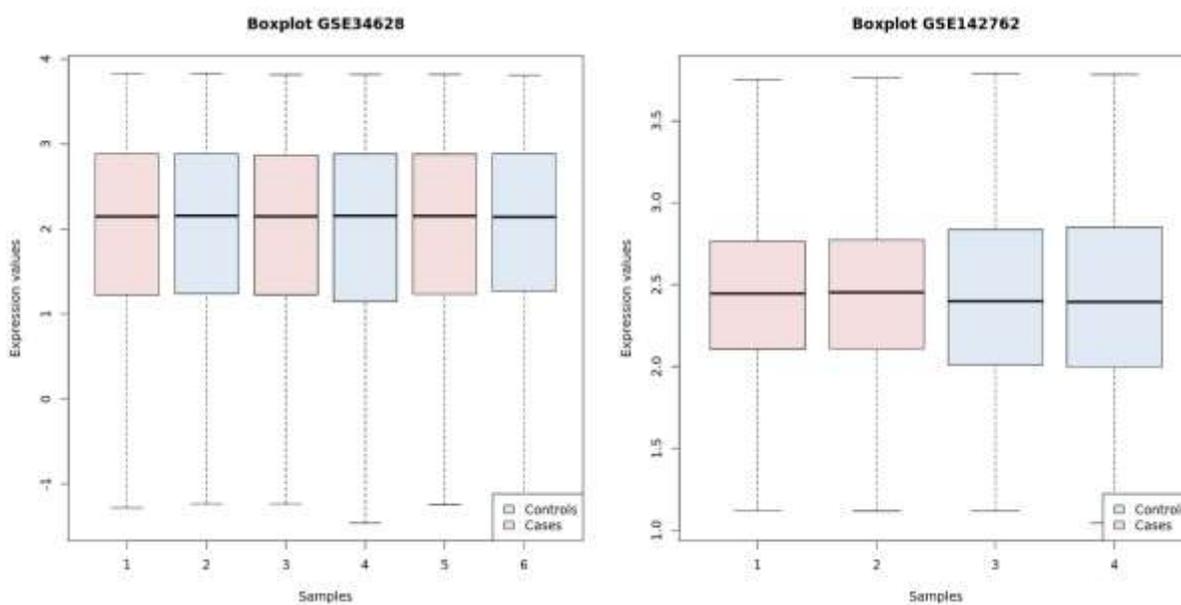


Figura 8. Análisis de cajas y bigotes datasets, ImGEO. GSE34628 infección con virus del dengue, GSE142762, infección con citomegalovirus.

Se pudo observar que los valores de las medianas tienen una relación uniforme en las muestras del dataset GSE34628, tanto para los casos, como para los controles, en un valor cercano a 2.1. Mientras que, para el dataset GSE142762, se distinguió una diferencia entre la mediana de los casos (~2.42), y los controles (~2.37). Aunque las muestras de casos se observaron con medianas homogéneas entre sí, al igual que las medianas correspondientes a las muestras de controles, la mediana de las muestras de los controles fue inferior a la de los casos.

También es claro afirmar que las muestras del dataset de infección con virus del dengue (GSE34628), presenta una mayor dispersión de sus datos, en un rango [~ -1.5 , ~ 3.8]. Por otra parte, la distribución de los datos del estudio en infección con citomegalovirus, indicó menor dispersión, con un rango [1 , ~ 3.7]. No se presentaron valor atípicos, ni extremos en ninguno de los datasets.

El metaanálisis ejecutado en imaGEO identificó 42 genes diferencialmente expresados. La tabla 9, describe los genes encontrados junto con sus valores descriptos del valor-p, el valor-p ajustado por logaritmo y el fold change promedio.

Tabla 9. Genes diferencialmente expresados en disfunción endotelial. Se indican los genes con diferencia estadísticamente significativa, resultante de la comparación de GSE34628 y GSE142762 por medio de la plataforma ImaGEO.

Gen	Nombre	p-valor	p-valor ajustado	Fold Change, promedio
CACNB2	calcium voltage-gated channel auxiliary subunit beta 2	1.6e-06	0.019	-1.2
GIPC3	GIPC PDZ domain containing family member 3	2.8e-06	0.019	-0.27

TMC8	transmembrane channel like 8	4.4e-06	0.019	-0.49
TRIM15	tripartite motif containing 15	4.7e-06	0.019	0.59
RIMKLA	ribosomal modification protein rimK like family member A	8.1e-06	0.024	0.48
IQCF1	IQ motif containing F1	8.8e-06	0.024	1.1
ZNF385D	zinc finger protein 385D	0.000011	0.026	-0.83
CCDC122	coiled-coil domain containing 122	0.000014	0.028	-0.77
MAF	MAF bZIP transcription factor	0.000016	0.029	-1.2
NKAIN1	sodium/potassium transporting ATPase interacting 1	2,00E-05	0.029	0.59
HTN1	histatin 1	2,00E-05	0.029	-0.87
PADI4	peptidyl arginine deiminase 4	0.000022	0.029	-0.75
TKTL1	transketolase like 1	0.000027	0.029	0.97
RCOR2	REST corepressor 2	0.000028	0.029	0.0049
POU2F3	POU class 2 homeobox 3	0.000031	0.029	-0.65
KIAA1751	NA	0.000033	0.029	1.7
MGEA5	meningioma expressed antigen 5 (hyaluronidase)	0.000034	0.029	-0.35
F5	coagulation factor V	0.000034	0.029	2.5
PRSS33	serine protease 33	0.000035	0.029	1
SLC25A48	solute carrier family 25, member 48	0.000036	0.029	0.96

ST18	ST18, C2H2C-type zinc finger	0.000038	0.029	0.62
REG1A	regenerating family member 1 alpha	0.000041	0.03	-0.47
TCTE1	t-complex-associated-testis- expressed 1	0.000043	0.031	0.45
PXT1	peroxisomal testis enriched protein 1	0.000048	0.033	0.018
BIN2	bridging integrator 2	0.000052	0.034	-0.29
KCNJ11	potassium voltage-gated channel subfamily J member 11	0.000064	0.039	1.7
AGRP	agouti related neuropeptide	0.000066	0.039	-0.47
TRPC5	transient receptor potential cation channel subfamily C member 5	7,00E-05	0.039	0.82
ROR2	receptor tyrosine kinase like orphan receptor 2	7,00E-05	0.039	-0.19
LMO3	LIM domain only 3	0.000075	0.041	0.9
MPZ	myelin protein zero	0.000088	0.046	0.31
GABRR3	gamma-aminobutyric acid type A receptor rho3 subunit (gene/pseudogene)	0.000092	0.046	-0.79
PCDH7	protocadherin 7	0.000094	0.046	-0.036
SRSF1	serine and arginine rich splicing factor 1	0.0001	0.048	-0.7
OR51M1	olfactory receptor family 51 subfamily M member 1	0.00011	0.048	-0.0064
NXNL2	nucleoredoxin like 2	0.00011	0.048	0.048

CPEB4	cytoplasmic polyadenylation element binding protein 4	0.00011	0.048	-0.38
TMEM155	transmembrane protein 155	0.00012	0.048	-0.011
DOC2A	double C2 domain alpha sodium/potassium	0.00012	0.048	1.2
NKAIN3	transporting ATPase interacting 3	0.00012	0.048	0.68
ADCY9	adenylate cyclase 9	0.00012	0.048	0.29
SCNN1G	sodium channel epithelial 1 gamma subunit	0.00012	0.048	1.3

Los valores positivos expresados en la columna de *fold change promedio* indica que son genes regulados positivamente, es decir, sobreexpresados. Para el caso del factor de coagulación v (F5), fue el gen que presentó mayor cambio en comparación con los demás, con un valor de 2.5, seguido por los genes KIAA1751 y KCNJ11 con valor de 1.7 cada uno. Otros genes que indicaron sobreexpresión significativa fueron, SCNN1G, DOC2A y IQCF1 con valores de 1.3, 1.2 y 1.1 respectivamente.

En cuanto a los valores negativos encontrados el *fold change promedio*, indican aquellos genes que se encontraron regulados negativamente (subexpresados), siendo el gen calcium voltage-gated channel auxiliary subunit beta 2 (CACNB2), el más significativo con un valor de -1.2. De los 42 genes identificados, 20 de ellos se presentaron regulados a la baja, mientras que los 22 restantes estaban regulados positivamente.

El mapa de calor mostrado en la figura 9, describe el comportamiento de cada gen diferencialmente expresado. La figura cuenta con una escala de color que describe el valor desde -1 que es el grado mínimo representado en la firma de calor para indicar subexpresión del gen.

En cambio, el valor 1, describe el valor máximo representado para indicar sobreexpresión del gen en cada muestra de cada dataset.

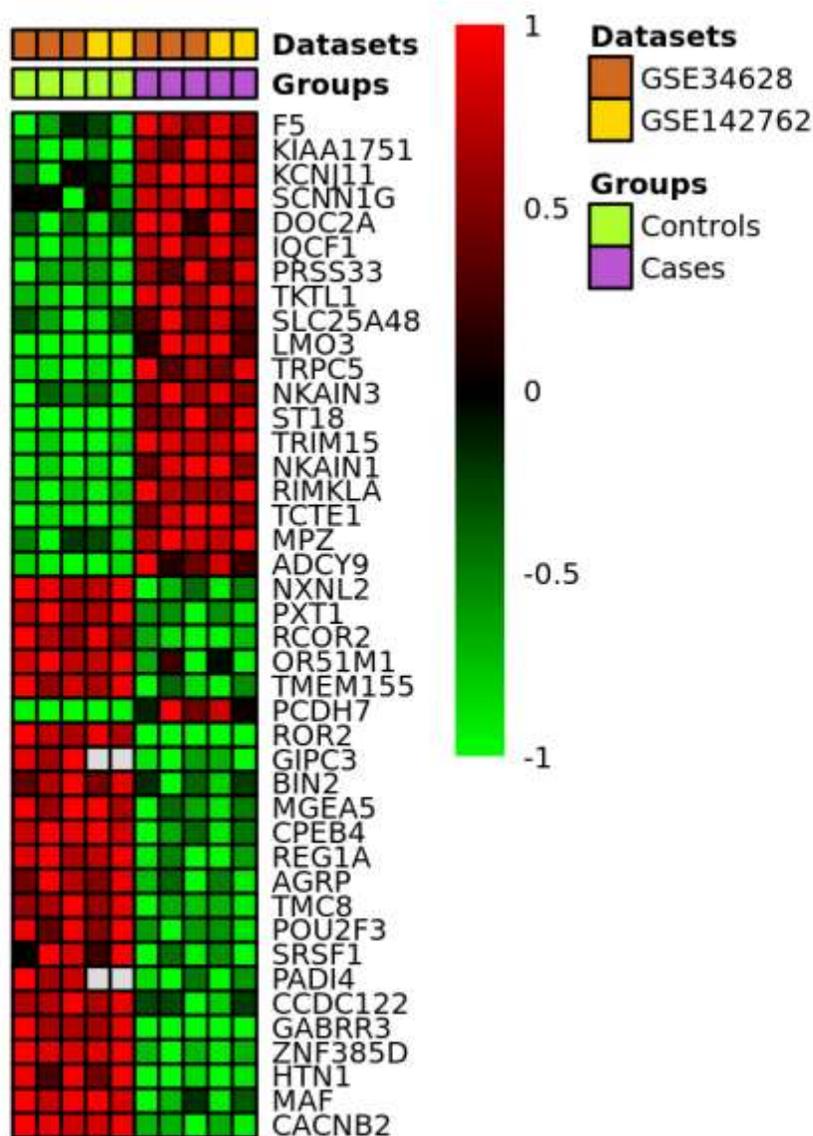


Figura 9. Mapa de calor, ImaGEO. Cada color en la escala de degradado representa un grado de cambio de expresión de los genes por muestra en cada dataset. Los valores <0 representan subexpresión y los valores >0 representan, sobreexpresión del gen en cada muestra de cada dataset. Mapa resultante de la comparación de GSE34628 y GSE142762 por medio de la plataforma ImaGEO.

El mapa de calor, al igual que el registro estadístico indicó que 20 de los genes presentaron sobreexpresión en las muestras de casos, en comparación con los controles, aunque los valores

fueron heterogéneos en cada muestra. Para los 22 genes restantes, se evidenció de igual forma, una heterogeneidad marcada en los valores por cada muestra. Estos 22 genes mostraron subexpresión en las muestras de casos, en comparación con los controles.

Debido a que solo contamos con un único dataset relacionado con el proceso de EndMT, se hizo el análisis individual en la extensión de NCBI, GEO2R. Cada dataset cumplió los mismos parámetros de análisis: el ajuste del valor p , *False discovery rate* postulado por Benjamini y Hochberg (1995), transformación de logaritmo por *Log2*, pesos de precisión de *limma*, se forzó normalización y se asignó valor de significancia en 0.05.

Analizados los datos se hizo la comparación directa entre el estudio de EndMT en infección de virus de Kaposi (GSE22522) y cada dataset de disfunción endotelial de la siguiente forma, GSE22522-GSE34628 y GSE22522-GSE142762. Los resultados de la comparación son producto de un ajuste en el valor de significancia de 0.5, debido a que en el valor 0.05, no se identificaron resultados para ningún caso.

El comparativo mostró que para el estudio de disfunción endotelial en la infección con el virus del dengue GSE142762, hay 22.070 genes con expresión diferencial en un valor superior a 0.5, mientras que, el estudio relacionado con la disfunción endotelial en la infección por citomegalovirus GSE34628, contiene únicamente 6 genes, iguales o superiores a este valor. El dataset único de EndMT GSE22522 cuenta con 28.027 genes diferencialmente expresados. Para el caso de GSE22522-GSE142762 se encontraron 149 genes comunes y, GSE22522-GSE34628 se encontraron 5 genes comunes (figura 10).



Figura 10. **Diagrama de Venn.** Comparativo de datasets entre el estudio de EndMT en infección de virus de Kaposi (GSE22522) y los datasets de disfunción endotelial (GSE34628) y (GSE142762).

Los genes encontrados en cada dataset cuentan con identificación por ID GeneBank Acc. Para la decodificación de estos ID. Se emplearon bancos de genes como Uniprot y Reactome, así como desde la misma plataforma de GeneBank y de las plataformas de secuenciación usadas en cada estudio. Sin embargo, sólo se pudo decodificar una fracción de ellos, correspondiente a 91 genes en total, 4 de la comparación GSE22522-GSE34628 y 87 de GSE22522-GSE142762.

Los genes que se relacionaron en el comparativo GSE22522-GSE34628, fueron: Band4.1-like protein 2 (EPB41L2), Mitogen-activated protein kinase 13 (MAP3K13), Phosphatidylserine synthase 2 (PTDSS2PSS2) y Junctophilin-2 (JPH2).

Por otra parte, algunos de los genes identificados Enel comparativo GSE22522-GSE142762 son: Cadherin 22 (CDH22), protein MAK 16 homolog (MAK16), Guanylate-binding protein 1 (GBP1), Metastasis-associated proteín (MTA3) y Thrombospondin-2 (THBS2).

5.3.3. Rutas de expresión

Con el objetivo de conocer qué rutas moleculares se encuentran involucradas en los procesos de disfunción endotelial y EndMT, se hizo la búsqueda por los genes que se encontraron diferencialmente expresadas. Los resultado del metaanálisis en disfunción endotelial indicó que, de los 42 genes, 11 se encuentran asociados en, al menos 18 rutas moleculares diferentes. Entre las más importantes se destacan MAPK, NF-kappa, Ras, Rap1, Wnt, IL-17, cAMP y la ruta de señalización de quimiocinas.

Los genes encontrados en el análisis comparativo de datasets por R, indicó que, ocho (8) de los 149, se encuentran presentes en al menos, ocho (8) rutas diferentes, de las que destacamos, MAPK, PI3K-Akt y la ruta de p53.

Al asociar los resultados del metaanálisis hecho en ImaGEO y el análisis comparativo de R, se evidencia la presencia de una sola ruta en ambos análisis. La ruta MAPK integra el gen *MAP3K13LZK*, del análisis comparativo y el gen *CACNB2* del metaanálisis de disfunción endotelial.

6. Discusión

CONSIDERACIONES SOBRE LA REVISIÓN SISTEMÁTICA

En el rastreo de información para seleccionar los artículos más importantes en cada uno de los fenómenos, disfunción endotelial y EndMT, tomamos las ecuaciones base con la ayuda del comodín: asterisco (*). Este comodín otorga mayor precisión en cuanto al rastreo por término clave (Comodines y operadores de búsqueda - Documentación de IBM, s/f). Por ejemplo, en el caso de sintetizar conjuntos de palabras como, endothel, endothelial, endothelin, endotheliitis, se indica una sección base y se agrega el comodín de la siguiente forma: “endothel*”. De esta

manera, el algoritmo de búsqueda reconoce las diferentes opciones en conjugación con los demás términos clave, para dar mayor precisión a la búsqueda.

Obtenida la ecuación base, como, por ejemplo: (Endothel* Dysfuntion*) AND (virus OR viral), y al observar el número tan amplio de resultados, se sometió la búsqueda a un cribado de exclusión, que implica la eliminación de resultados con el criterio a excluir. Teniendo en cuenta algunos patrones detectados en los resultados, como virus tomados como vector, tratamientos farmacológicos y otros, se definieron los operadores de exclusión con los términos no deseados. Entre los términos excluyentes más usados se hallaron: review, vaccine, animals, Covid*, (mouse Or mice), endogenous, vector.

Estos excluyentes se usaron por motivo que en la mayoría de los casos distorsionan el enfoque de nuestra investigación, para el caso de “vector” indicaba la infección viral como un transporte dirigido de algún tipo de tratamiento celular, “vaccine” interfiere en que los datos presentados indican estudios de síntomas y reacciones. Para el caso de “covid*” fue un término que se tuvo que excluir, que, aunque presentan estudios relacionados a disfunción endotelial y patologías relacionadas, los estudios son poco precisos, contienen poca replicabilidad y aluden a síntomas y tratamientos experimentales.

Adicionalmente se destaca que la limpieza de los resultados se hizo mediante filtros propios de las plataformas, como fue el caso de, año de publicación 2011-2021, tipo de documentos (research or journal) de estado final, especie Homo sapiens y lenguaje español o inglés. Estos filtros permiten tener un alineamiento más objetivo entre los resultados, así como un grado de confianza de la calidad bibliográfica encontrada.

Cada ecuación de búsqueda se adaptó de acuerdo a los resultados de cada plataforma, así también los filtros son dependientes de la interfaz de ella. En el caso de la plataforma ScienceDirect, su búsqueda avanzada no permite el uso de comodines y para las ecuaciones de búsqueda tiene un límite permitido de 8 términos. De igual forma, las plataformas ofrecen distintos tipos de filtros para las búsquedas, lo que representa que, para hacer estudios de bibliometría, la comparación de sus resultados se puede ver sesgada por estos parámetros.

La selección de datos involucró, además de una ecuación de búsqueda base como, Endothel* dysfunction* AND (virus OR viral), términos clave (RNAseq OR MicroArray) y se aplicó el filtro inclusivo Associated data. Sin embargo, la búsqueda y descarga se hizo específicamente desde la extensión GEO (Gene Expression Omnibus) de NCBI y desde la extensión ENA (European Nucleotide Archive) de EMBL.

Los datos que coincidieron en mayor parte con los criterios de búsqueda se descargaron en el servidor para la inspección manual. De un total de 31 datasets que contienen 13.817 muestras, se analizó su composición, y se descartó el 90.32% de estudios y por tanto se incluyó en el estudio únicamente 3 datasets, que representan el 9.68% de ellos.

Los datos con mayor índice de exclusión corresponden a estudios en otro linaje celular (41.94%), seguido de aquellos que presentaron incompatibilidad en el tratamiento con los intereses de esta investigación (25.81%) y otros parámetros como especie, disponibilidad de los datos o respaldo bibliográfico de los mismos, que suman el 22.54% del total.

CONSIDERACIONES DEL DESEMPEÑO DE LAS HERRAMIENTAS INFORMÁTICAS

A partir de los resultados obtenidos, y dada la incompatibilidad de la plataforma Affymetrix Human Gene Expression Array, usada en la secuenciación del dataset GSE142762

con las aceptadas en el proceso de conversión de NetworkAnalyst, que no fue posible hacer la evaluación de los softwares con el conjunto de datasets obtenidos para el estudio. Por este motivo, se tomaron 3 datasets usados en el metaanálisis en la enfermedad de Alzheimer por Moradifard et al., (2018), que sirvieron de objeto de estudio para los softwares y así poder obtener la comparabilidad de ellos. En casos como el anterior, se deben tener en cuenta las plataformas de secuenciación permitidas por el software, lo que, a su vez, pueden indicar la elección de las herramientas más eficientes.

La evaluación detallada de los softwares nos permitió interpretar que, la plataforma IRIS-EDA al presentar constantes desconexiones con el servidor, ser una plataforma desactualizada y con baja capacidad de recepción de datasets (2), es una herramienta poco viable. Situación por la que su uso se descartó para futuros estudios. En síntesis, no se recomienda su uso.

Por otra parte, en el caso de la plataforma ImaGEO, fue apreciable su desempeño y atributos favorables para un metaanálisis online. El manejo de su interfaz es de fácil acceso, su manipulación es práctica.

ImaGEO, cuenta además con métodos de análisis en comparación de tamaño del efecto, con soporte en modelos de efecto fijo (FEM). Estos modelos interpretan las entradas como estudios que provienen de un tamaño del efecto real más el error de medición. Al mismo tiempo, identifican los genes con efecto más fuerte en estudio.

Desde el modelo de efectos aleatorios (REM), en el que cada estudio contiene un efecto aleatorizado que puede incorporar heterogeneidades de estudios cruzados por la presencia de plataformas o lotes diferentes. REM permite la identificación de genes con el efecto promedio más fuerte en una población hipotética de estudios (Toro-Domínguez et al., 2019). Para poder

definir cuál de los dos modelos puede llegar a ser más eficiente en el análisis de los datos, estos se deben someter a un análisis de heterogeneidad mediante la ‘prueba Q’ propuesta por DerSimonian y Laird (Fernandez et al., 2019).

Otro método incorporado en ImaGEO, es la combinación del valor-p, que cuenta con cuatro (4) estadígrafos diferentes. El primero, el modelo de Fisher que está definido como la suma de $-\log(\text{valor-p})$ entre los datos. El segundo, modelo de Stouffer, que cuenta con un grado de similitud con el modelo de fisher, pero difiere en que este modelo basa su análisis en puntuaciones Z. Se destaca que la transformación Z, mide la diferencia entre una estadística observada, los niveles de expresión génica y su parámetro de población hipotético en unidades de la desviación estándar. El tercero es el estadígrafo maxP, que identifica y asigna a todos los estudios el valor-p máximo. Y, por último, el modelo minP, identifica y asigna el valor-p mínimo a todos los estudios. Este tipo de combinaciones cuenta con la ventaja para la estandarización de asociaciones de datos génicos a una escala común, lo que facilita la comparación de datos muy heterogéneos (Toro-Domínguez et al., 2019).

Además, ImaGEO, empleó poco tiempo en la ejecución de los análisis, así mismo destacó en la generación de las visualizaciones de los resultados, que se dan de forma ordenada, clara y precisa con la opción de descargar, en caso de ser requerido.

NetworkAnalyst, por su parte, resultó ser la plataforma más completa en el estudio. Es una herramienta virtual de acceso gratuito y licencia abierta para replicación de los análisis. Cuenta con un servidor que soporta una cantidad considerable de muestras y por ende de datasets. A pesar de contar con un número inferior de métodos y modelos para el metaanálisis, en comparación por ejemplo con ImaGEO, tiene los más usados.

Para el método de comparación de valores-p, contiene el método de Fisher y Stouffer. Para el método de combinación de tamaños del efecto, cuenta con los modelos de efecto fijo y efecto aleatorio, pero cuenta además con el análisis de conteo de votos, que consiste en identificar los genes diferencialmente expresados en cada dataset y posteriormente calcular el número total de veces que se presentó en todos los datasets, este método puede establecer un número mínimo de votos, para mayor representatividad de los resultados. Y el método de fusión directa, que asocia todas las muestras como un solo dataset, indicando aquellos genes que presentan expresión diferencial, los resultados se muestran en forma de lista fusionada y solo se puede aplicar cuando los datos tienen una amplia similitud entre sí (Zhou et al., 2019).

NetworkAnalyst también permite el control del tratamiento de los datos, lo que permite la visualización del comportamiento de las muestras de cada dataset antes de la combinación. El tiempo de ejecución del metaanálisis es considerablemente corto. En cuanto al tratamiento de los resultados, esta plataforma ofrece una amplia gama de visualizaciones y herramientas interactivas para análisis más profundos.

Los anteriores elementos de análisis fueron la base para la postulación de nuestro protocolo que proporciona un procedimiento eficiente en el metaanálisis de expresión génica.

Entre los aspectos importantes para tener en cuenta está inicialmente que, los datos deben cumplir con los siguientes criterios para el estudio: secuenciación por tecnología RNAseq, o Microarreglo, definición de muestras en dos categorías (casos y controles), debidamente etiquetadas y depositadas en documento .txt delimitado por tabulación, con un tamaño inferior a 50 Megabytes, en caso de exceder este límite, se puede comprimir en archivo .zip.

En la plataforma online de NetworkAnalyst <https://www.networkanalyst.ca/>, se ingresa a la opción Multiple Gene Expression Tables, se cargan los datos y se especifican los parámetros de la siguiente forma:

En la opción ID converter, se indica el organismo objeto de estudio en el dataset y se ingresa la plataforma de secuenciación usada. Indicar, *proceed*.

Para la normalización, en caso de que sea requerida se seleccionará, si es RNAseq. Se recomienda usar la transformación Log2. Y en caso de tener muestras de MicroArreglos, se sugiere la transformación Variance Stabilizing Normalization (VSN). Indicar, *proceed*.

Además, se debe indicar un contraste comparativo entre caso y control, así mismo se debe asignar el umbral de significancia deseado, se recomienda 0.05. Indicar, *proceed*.

Finalmente, para la visualización y otros análisis se debe considerar que estos dependen del interés propio del estudio en particular.

CONSIDERACIONES DE LAS MOLÉCULAS ASOCIADAS A LA DISFUNCIÓN Y EndMT.

En el proceso de disfunción endotelial o también llamada activación del endotelio, las células endoteliales aumentan los niveles de expresión de moléculas de adhesión a la superficie endotelial. Por ello se hacen evidentes moléculas como, molécula de adhesión intercelular ICAM-1, molécula de adhesión celular vascular VCAM-1 y e-selectin (Soe et al., 2017; Yacoub et al., 2016). Dado que en muchos casos, los cambios de expresión molecular, pueden estar inducidos por lesiones en el endotelio, se destaca la sobreexpresión en moléculas proinflamatorias como las interleucinas y los interferones (Barbachano-Guerrero et al., 2020; Dalrymple & Mackow, 2012; Patkar et al., 2013).

A parte de las lesiones físicas observadas, se ha reportado que infecciones virales o bacterianas, también son inductoras de remodelamiento en el endotelio vascular. La razón radica en que algunos virus que tienen como blanco de infección células endoteliales, ocasionan la activación del endotelio, lo que conlleva a fugas vasculares y hemorragias (Soe et al., 2017; van de Weg et al., 2015). En este aspecto, el virus del dengue, es uno de los agentes más estudiados, debido a su relación con la patología de fiebre hemorrágica y síndrome de shock.

Para el caso de infecciones virales, algunas de ellas ocasionan graves patologías como la hipertensión arterial, que pueden provenir inicialmente de una activación del endotelio. Como se mencionó, tal activación está fuertemente relacionada con los cambios en los patrones de expresión de moléculas estructurales o funcionales de la célula endotelial. Algunos de los virus que identificamos con importancia en la disfunción endotelial son citomegalovirus humanos (Yaiw et al., 2014; Zhang et al., 2020), que pertenece a la familia Herpesviridae y se relaciona estrechamente con la aterosclerosis (Valle et al., 2003), una condición común en infecciones virales con afección al endotelio. Otros virus: El virus de inmunodeficiencia humana (VIH), reconocido como un inductor de la disfunción endotelial (Frustaci et al., 2014; Gupta et al., 2019), así como el hantavirus (Krautkrämer et al., 2016) y virus de la hepatitis (Kim et al., 2008).

Si bien, el virus del dengue ha sido ampliamente estudiado, más que nada dentro de un contexto sintomatológico de fiebre hemorrágica, fuga vascular y principalmente, activación del endotelio. Cabe destacar que nuestra investigación se enfocó en determinar el alcance de la infección a nivel molecular, de rutas de expresión y órganos. A partir de esta inspección, se evidencian estudios hechos *in vitro* o reportes de caso, que indican que el tejido sufre activación en al menos 7 órganos diferentes, endotelio de cordón umbilical, de pulmón, sublingual, dérmico, cardiaco, cerebral y renal.

Desde otro enfoque, los cambios de expresión observados, muestran un grupo de moléculas que sufrieron cambio de tipo inhibitorio dentro del contexto infeccioso, como el caso del ácido siálico y sindecan-1 (Puerta-Guardo et al., 2019). Otro tipo de cambio evidenciado fue la fosforilación a nivel proteico de Akt, RhoA y VE-cadherin (Immenschuh et al., 2013; Singh et al., 2018).

Algunas moléculas no presentaron cambios en el patrón de expresión, sin embargo, su presencia dentro del fenómeno infeccioso fue de importancia en la activación de algunas moléculas o incluso de relevancia en la activación o inhibición de rutas de expresión génica, como el caso de ISG56 bloquea la inducción del interferón IFN- β en la infección del dengue. (Dalrymple & Mackow, 2012)

Se pudo observar también, que la infección viral tiende en la mayoría de los casos a regular positivamente la expresión de las moléculas de tipo proinflamatorio como en el caso de las interleucinas (IL- 6, 7, 8, 10) (Álvarez-Díaz et al., 2019; Soe et al., 2017), quimiocinas (CCL- 2, 5, 20), (CXCL- 1, 3, 4, 5, 8, 9, 10, 11) (Dalrymple & Mackow, 2012; Soe et al., 2017), en otros casos de tipo enzimático como Akt, catepsin L, CD73, OAS (2, 3) (Patkar et al., 2013).

Dichos aumentos en la expresión molecular, pueden estar relacionados con procesos como la sobreexpresión de proteínas de adhesión a plaquetas o células inmunitarias en la membrana celular (Armstrong et al., 2013; Quirino-Teixeira et al., 2020), o también puede estar relacionado a procesos de angiogénesis (Álvarez-Díaz et al., 2019).

La figura 11, comprende una relación donde se puede observar la construcción de diferentes comunidades de moléculas, que se asocian ya sea a un tejido por infección o a rutas de expresión génica. Destacamos que, en las interacciones por infección, la molécula no estructural 1 (NS1)

corresponde al nodo con mayor agrupamiento, incluso por encima de los nodos que representan los serotipos del virus del dengue.

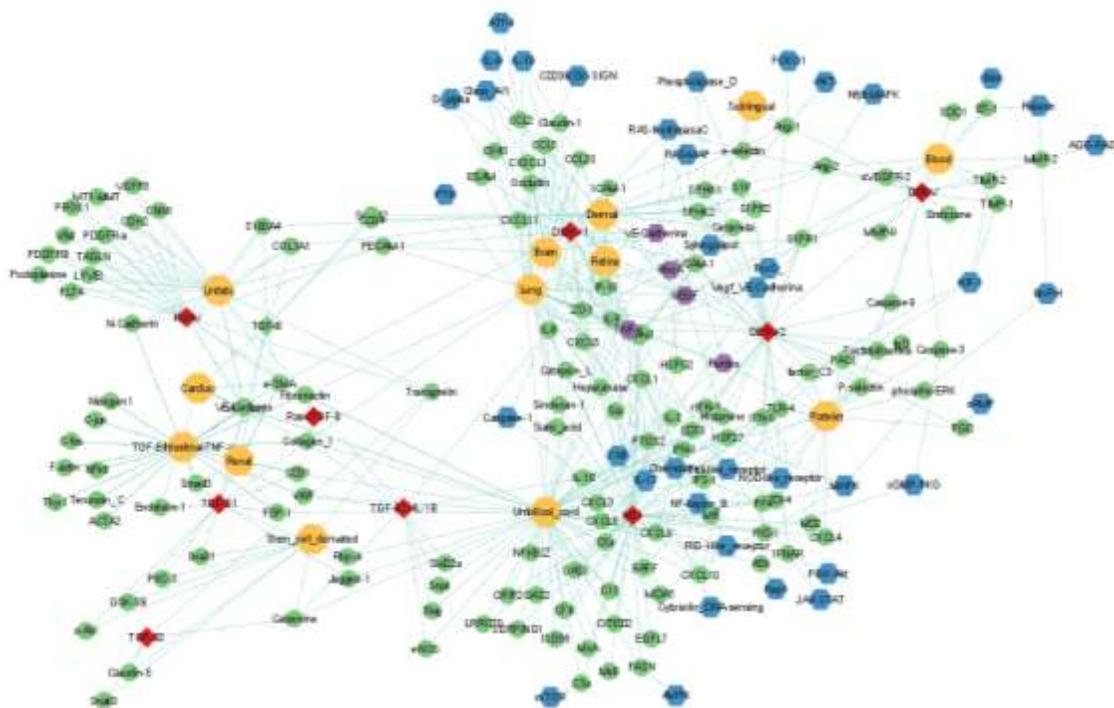


Figura 11. Red de moléculas asociadas a la disfunción endotelial y EndMT. Los nodos verdes definen las moléculas, los hexágonos azules representan las rutas de expresión relacionadas con las moléculas (las rutas no necesariamente por completo en el proceso). Los nodos lila, representan aquellas rutas y moléculas que llevan el mismo nombre. Los rombos rojos representan las moléculas, genes o virus que inducen lesión. Por último, los círculos amarillos, representan los tejidos implicados.

En cuanto a la regulación negativa en la expresión génica, corresponde en muchos casos a moléculas de adhesión estrecha como ZO-1 (Raekiansyah et al., 2014), ICAM-1 (Soe et al., 2017) lo que indica un remodelamiento del endotelio y fuga vascular.

Desde otra perspectiva, lo encontrado en los estudios de Transición Endotelial a Mesenquimal, indica que los cambios de expresión están más relacionados con la transformación celular. La infección viral promueve los cambios en los patrones de expresión, desde una

disminución progresiva de los marcadores moleculares del endotelio como CD31, VE-cadherin, factor von Willebrand (vWF) (Maleszewska et al., 2013; Rieder et al., 2011b; Tian et al., 2019), hasta un aumento en los marcadores moleculares mesenquimales como α -SMA, COL1A1, FSP-1 (Cheng et al., 2011; Goumans et al., 2012; Rieder et al., 2011b).

Estos hallazgos también convergen en señalar que dentro de la disfunción endotelial existe la expresión de TNF- α , reportado como un factor regulado positivamente en la disfunción endotelial por (Álvarez-Díaz et al., 2019; Barbachano-Guerrero et al., 2020; Patkar et al., 2013; Soe et al., 2017) y en la EndMT es indicado de participar como inductor por (Rieder et al., 2011b).

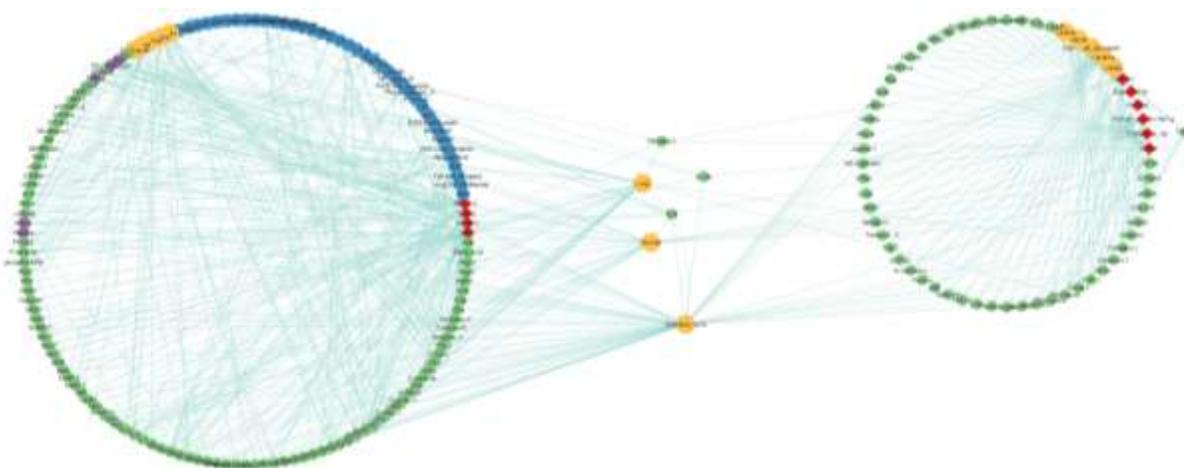


Figura 12. Red circular de moléculas asociadas a la disfunción endotelial y EndMT. Indica el círculo más grande (izquierda) como las interacciones entre moléculas (nodos verdes), rutas de expresión (nodos azules), moléculas y rutas con el mismo nombre (nodos lila), agentes o moléculas inductoras de lesión (nodos rojos) y tejidos (nodos amarillos) en la disfunción endotelial, así mismo, las interacciones en la EndMT (círculo derecho).

La figura 12 indica la relación que existe entre los dos procesos, los nodos puente son moléculas que tienen relación estrecha con otros nodos en ambos fenómenos. Se destacan 3 nodos de moléculas, PECAM-1, IL-1 β e IL-6. La PECAM-1 o también conocida como CD31

(marcador molecular endotelial), cumple una función de receptor inhibidor que, mediante la fosforilación secuencial, limita las respuestas de la activación celular (del endotelio). En cuanto a la interleucina IL-1 β , conocida como pirógeno leucocitario, es mediador importante en la respuesta inflamatoria y se encuentra asociada también con otros procesos importantes como la proliferación, diferenciación y apoptosis celular. En cuanto a la interleucina IL-6, también con importancia inflamatoria, se encuentra estrechamente relacionada como mediadora de la fiebre y reacciones agudas en infecciones de varios tipos. También

De igual forma se evidenció la importancia a nivel de tejido que, el endotelio pulmonar, dérmico y de cordón umbilical, tienen relación estrecha en los dos fenómenos, por lo tanto, pueden ser objeto de estudio en la relación real de dichos procesos.

7. CONCLUSIONES

Del cribado usado en más de 124 mil estudios encontrados, se determinó que hay abundancia en lo que se relaciona con la disfunción endotelial y virus, sin embargo, gran parte de ellos se enfocan en casos clínicos, tratamientos farmacológicos, muestras y estudios sanguíneos, respuestas inmunes y estudios sintomatológicos. Por otra parte, la Transición Endotelial a Mesenquimal, sigue siendo un tema novedoso, indicando que es un tema de gran impacto para la comunidad científica en este momento. De igual forma, en lo que refiere a los datos estadísticos disponibles, se presentan grandes vacíos, tanto en la parte de diseño experimental, como de un enfoque relacionado al fenómeno a nivel molecular.

El metaanálisis es una metodología de gran potencial en el análisis de expresión diferencial de genes. Recomendamos como plataforma confiable a NetworkAnalyst, por contener una interfaz más completa hasta el momento. No obstante, se ofrece como alternativa eficiente la plataforma ImaGEO.

La disfunción endotelial inducida por la infección viral, comprende la activación de numerosas rutas de expresión génicas y el efecto en moléculas de distinta categoría. Se identificó un total de 164 moléculas diferencialmente expresadas en este fenómeno, 122 identificadas en la revisión sistemática y 42 en el metaanálisis ejecutado en ImaGEO. De los estudio de transición endotelial a mesenquimal, se identificaron relevantemente un total de 154 moléculas involucradas, de las cuales, 91 comprenden una relación con la disfunción endotelial determinado por el análisis comparativo de los datasets.

La disfunción endotelial y la EndMT, entendidos desde una posible relación molecular en la infección viral, comprenden moléculas ya establecidas e identificadas. Nosotros logramos, además, aportar una red de correlación que muestra gráficamente las interacciones de cada fenómeno entre moléculas, rutas, agentes inductores y tejidos. Así mismo, presentamos algunos nodos puente (las moléculas PECAM-1, IL-6 e IL-1 β), con importancia en la unión de los procesos. Así mismo, algunos tejidos identificados como blanco opcional en futuros estudios relacionados. Todo esto se presenta como base teórica, con miras a la próxima comprobación experimental.

8. REFERENCIAS BIBLIOGRÁFICAS

¿Qué es la secuenciación de ADN de próxima generación? | Genómica funcional II. (s/f).

Aguilar-Bultet, L. & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas

bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*,

37(2), 125–132.

Ahn, E. & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean*

Journal of Anesthesiology, 71(2), 103. <https://doi.org/10.4097/KJAE.2018.71.2.103>

Álvarez-Díaz, D. A., Gutiérrez-Díaz, A. A., Orozco-García, E., Puerta-González, A., Bermúdez-Santana, C. I. & Gallego-Gómez, J. C. (2019). Dengue virus potentially promotes migratory responses on endothelial cells by enhancing pro-migratory soluble factors and miRNAs. En *Virus Research* (Vol. 259). <https://doi.org/10.1016/j.virusres.2018.10.018>

Árbol de decisión. (s/f).

Arciniegas, E., Sutton, A. B., Allen, T. D. & Schor, A. M. (1992). Transforming growth factor beta 1 promotes the differentiation of endothelial cells into smooth muscle-like cells in vitro. *Journal of Cell Science*, 103(2), 521–529.

Armstrong, S. M., Darwish, I. & Lee, W. L. (2013). Endothelial activation and dysfunction in the pathogenesis of influenza A virus infection. *Virulence*, 4(6).
<https://doi.org/10.4161/viru.25779>

Badr, M. T. & Häcker, G. (2019). Gene expression profiling meta-analysis reveals novel gene signatures and pathways shared between tuberculosis and rheumatoid arthritis. *PLoS ONE*, 14(3), 1–16. <https://doi.org/10.1371/journal.pone.0213470>

Barbachano-Guerrero, A., Endy, T. P. & King, C. A. (2020). Dengue virus non-structural protein 1 activates the p38 MAPK pathway to decrease barrier integrity in primary human endothelial cells. *Journal of General Virology*, 101(5), 484–496.
<https://doi.org/10.1099/jgv.0.001401>

Castro Ávalos, M. A. (2020). Efecto de la Vacuna del Papiloma Humano en la prevención del cáncer cervicouterino: Metaanálisis. *Archives of Nursing Research*, 3(1), 89.

<https://doi.org/10.24253/anr.3.89>

- Chen, G., Yang, X., Wang, B., Cheng, Z. & Zhao, R. (2019). Human cytomegalovirus promotes the activation of TGF- β 1 in human umbilical vein endothelial cells by MMP-2 after endothelial mesenchymal transition. *Advances in Clinical and Experimental Medicine*, 28(11), 1441–1445. <https://doi.org/10.17219/acem/109199>
- Chen, P. Y., Qin, L., Barnes, C., Charisse, K., Yi, T., Zhang, X., Ali, R., Medina, P. P., Yu, J., Slack, F. J., Anderson, D. G., Kotelianski, V., Wang, F., Tellides, G. & Simons, M. (2012). FGF Regulates TGF- β Signaling and Endothelial-to-Mesenchymal Transition via Control of let-7 miRNA Expression. *Cell Reports*, 2(6), 1684–1696. <https://doi.org/10.1016/j.celrep.2012.10.021>
- Cheng, F., Pekkonen, P., Laurinavicius, S., Sugiyama, N., Henderson, S., Günther, T., Rantanen, V., Kaivanto, E., Aavikko, M., Sarek, G., Hautaniemi, S., Biberfeld, P., Aaltonen, L., Grundhoff, A., Boshoff, C., Alitalo, K., Lehti, K. & Ojala, P. M. (2011). KSHV-initiated notch activation leads to membrane-type-1 matrix metalloproteinase-dependent lymphatic endothelial-to-mesenchymal transition. *Cell Host and Microbe*, 10(6), 577–590. <https://doi.org/10.1016/j.chom.2011.10.011>
- Collí-Dulá, R. C., Martyniuk, C. J., Streets, S., Denslow, N. D. & Lehr, R. (2016). Molecular impacts of perfluorinated chemicals (PFASs) in the liver and testis of male largemouth bass (*Micropterus salmoides*) in Minnesota Lakes. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, 19, 129–139. <https://doi.org/10.1016/j.cbd.2016.02.001>
- Costa-Silva, J., Domingues, D. & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*, 12(12), 1–18.

<https://doi.org/10.1371/journal.pone.0190152>

Costa, V., Angelini, C., De Feis, I. & Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine and Biotechnology*, 2010.

<https://doi.org/10.1155/2010/853916>

Dalrymple, N. A. & Mackow, E. R. (2012). Endothelial Cells Elicit Immune-Enhancing Responses to Dengue Virus Infection. *Journal of Virology*, 86(12), 6408–6415.

<https://doi.org/10.1128/jvi.00213-12>

Endemann, D. H. & Schiffrin, E. L. (2004). Endothelial dysfunction. *Journal of the American Society of Nephrology*, 15(8), 1983–1992.

<https://doi.org/10.1097/01.ASN.0000132474.50966.DA>

Fernandez Chinguel, J. E., Zafra Tanaka, J. H., Goicochea Lugo, S., Peralta, C. I., Taype Rondan, A., Huo, Z., Tang, S., Park, Y., Tseng, G. C., Tejera-Vaquerizo, A., Descalzo-Gallego, M. A., Otero-Rivas, M. M., Posada-García, C., Rodríguez-Pazos, L., Pastushenko, I., Marcos-Gragera, R., García-Doval, I., Siddaway, A. P., Wood, A. M., ... García-Doval, I. (2019). Meta-analyses in mental health research . A practical guide. *Bioinformatics*, 33(3), 19–22. <https://doi.org/10.1093/nar/gkz240>

Frustaci, A., Petrosillo, N., Vizza, D., Francone, M., Badagliacca, R., Verardo, R., Fedele, F., Ippolito, G. & Chimenti, C. (2014). Myocardial and microvascular inflammation/infection in patients with HIV-associated pulmonary artery hypertension. *Aids*, 28(17), 2541–2549.

<https://doi.org/10.1097/QAD.0000000000000426>

Gasparini, P., Espigol-Frigole, G., McCormick, P. J., Salvucci, O., Maric, D., Uldrick, T. S., Polizzotto, M. N., Yarchoan, R. & Tosato, G. (2012). Kaposi sarcoma herpesvirus promotes

- endothelial-to-mesenchymal transition through notch-dependent signaling. *Cancer Research*, 72(5), 1157–1169. <https://doi.org/10.1158/0008-5472.CAN-11-3067>
- Goumans, M. J., van Zonneveld, A. J. & ten Dijke, P. (2008). Transforming Growth Factor β -Induced Endothelial-to-Mesenchymal Transition: A Switch to Cardiac Fibrosis? *Trends in Cardiovascular Medicine*, 18(8), 293–298. <https://doi.org/10.1016/j.tcm.2009.01.001>
- Goumans, M. J., van Zonneveld, A. J., Ten Dijke, P., Gasperini, P., Espigol-Frigole, G., McCormick, P. J., Salvucci, O., Maric, D., Uldrick, T. S., Polizzotto, M. N., Yarchoan, R., Tosato, G., Rieder, F., Kessler, S. P., West, G. A., Bhilocha, S., De La Motte, C., Sadler, T. M., Gopalan, B., ... Manasek, F. J. (2012). Molecular mechanisms of endothelial to mesenchymal cell transition (EndoMT) in experimentally induced fibrotic diseases. *American Journal of Pathology*, 103(2), 177–186. <https://doi.org/10.1016/j.ceb.2016.07.005>
- Gupta, S. K., Liu, Z., Sims, E. C., Repass, M. J., Haneline, L. S. & Yoder, M. C. (2019). Endothelial colony-forming cell function is reduced during HIV infection. *Journal of Infectious Diseases*, 219(7), 1076–1083. <https://doi.org/10.1093/infdis/jiy550>
- Heller, M. J. (2002). DNA Microarray Technology: Devices, Systems, and Applications. *Annual Review of Biomedical Engineering*, 4(1), 129–153. <https://doi.org/10.1146/annurev.bioeng.4.020702.153438>
- Hincapié-Monsalve, V. & Gallego-Gómez, J. . (2020). Transición Epitelio-Mesénquima Inducida por Virus. *Revista Acta Biológica Colombiana. Manuscrito N° 7935. Aceptado .*
- Immenschuh, S., Rahayu, P., Bayat, B., Saragih, H., Rachman, A. & Santoso, S. (2013). Antibodies against dengue virus nonstructural protein-1 induce heme oxygenase-1 via a redox-dependent pathway in human endothelial cells. *Free Radical Biology and Medicine*,

54, 85–92. <https://doi.org/10.1016/j.freeradbiomed.2012.10.551>

- Kim, W. U., Yoo, S. A. & Kwok, S. K. (2008). Proinflammatory role of vascular endothelial growth factor in the pathogenesis of rheumatoid arthritis: Prospects for therapeutic intervention. *Mediators of Inflammation*, 2008. <https://doi.org/10.1155/2008/129873>
- Krautkrämer, E., Nussbag, C., Baumann, A., Schäfer, J., Hofmann, J., Schnitzler, P., Klempa, B., Witkowski, P. T., Krüger, D. H. & Zeier, M. (2016). Clinical characterization of two severe cases of hemorrhagic fever with renal syndrome (HFRS) caused by hantaviruses Puumala and Dobrava-Belgrade genotype Sochi. *BMC Infectious Diseases*, 16(1), 1–8. <https://doi.org/10.1186/s12879-016-2012-2>
- Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D. L., Lee, T. I. & Young, R. A. (2012). Revisiting global gene expression analysis. *Cell*, 151(3), 476–482. <https://doi.org/10.1016/j.cell.2012.10.012>
- Lusa, L., Cappelletti, V., Gariboldi, M., Ferrario, C., De Cecco, L., Reid, J. F., Toffanin, S., Gallus, G., McShane, L. M., Daidone, M. G. & Pierotti, M. A. (2006). Questioning the utility of pooling samples in microarray experiments with cell lines. *International Journal of Biological Markers*, 21(2), 67–73. <https://doi.org/10.5301/JBM.2008.2616>
- M^a, R., González, A. & Uk, A. (2008). *UNIVERSIDAD DE ALCALÁ FACULTAD DE MEDICINA Tesis Doctoral CORE View metadata, citation and similar papers at core.*
- Maleszewska, M., Moonen, J. R. A. J., Huijkman, N., van de Sluis, B., Krenning, G. & Harmsen, M. C. (2013). IL-1 β and TGF β 2 synergistically induce endothelial to mesenchymal transition in an NF κ B-dependent manner. *Immunobiology*, 218(4), 443–454. <https://doi.org/10.1016/j.imbio.2012.05.026>

- Mihira, H., Suzuki, H. I., Akatsu, Y., Yoshimatsu, Y., Igarashi, T., Miyazono, K. & Watabe, T. (2012). TGF- β -induced mesenchymal transition of MS-1 endothelial cells requires Smad-dependent cooperative activation of Rho signals and MRTF-A. *Journal of Biochemistry*, *151*(2), 145–156. <https://doi.org/10.1093/jb/mvr121>
- Monier, B., McDermaid, A., Wang, C., Zhao, J., Miller, A., Fennell, A. & Ma, Q. (2019). IRIS-EDA: An integrated RNA-seq interpretation system for gene expression data analysis. *PLoS Computational Biology*, *15*(2), 1–15. <https://doi.org/10.1371/journal.pcbi.1006792>
- Moradifard, S., Hoseinbeyki, M., Ganji, S. M. & Minucheher, Z. (2018). Analysis of microRNA and Gene Expression Profiles in Alzheimer's Disease: A Meta-Analysis Approach. *Scientific Reports*, *8*(1), 1–17. <https://doi.org/10.1038/s41598-018-20959-0>
- Mudau, M., Genis, A., Lochner, A. & Strijdom, H. (2012). Endothelial dysfunction: The early predictor of atherosclerosis. *Cardiovascular Journal of Africa*, *23*(4), 222–231. <https://doi.org/10.5830/CVJA-2011-068>
- Muñoz-Hernández, R., Ampuero, J., Millán, R., Gil-Gómez, A., Rojas, Á., Macher, H. C., Gallego-Durán, R., Gato, S., Montero-Vallejo, R., Rico, M. C., Maya-Miles, D., Sánchez-Torrijos, Y., Soria, I. C., Stiefel, P. & Romero-Gómez, M. (2020). Hepatitis C Virus Clearance by Direct-Acting Antivirals Agents Improves Endothelial Dysfunction and Subclinical Atherosclerosis: HEPCAR Study. *Clinical and translational gastroenterology*, *11*(8), e00203. <https://doi.org/10.14309/ctg.000000000000203>
- Ortigosa, J. & Cañas, R. (2020). Métodos de secuenciación de ácidos nucleicos : Primera generación Método de escisión química (Ma- xam y Gilbert). *Encuentros en la Biología.iología.*, *XIII*(173), 19–25.

http://www.encuentros.uma.es/assets/journals/13/173singles/173.7_secuenciacion.pdf

- Ozsolak, F. & Milos, P. M. (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics*, 12(2), 87–98. <https://doi.org/10.1038/nrg2934>
- Patkar, C., Giaya, K. & Libraty, D. H. (2013). Dengue virus type 2 modulates endothelial barrier function through CD73. *American Journal of Tropical Medicine and Hygiene*, 88(1), 89–94. <https://doi.org/10.4269/ajtmh.2012.12-0474>
- Piera-Velazquez, S. & Jimenez, S. A. (2019). Endothelial to mesenchymal transition: Role in physiology and in the pathogenesis of human diseases. *Physiological Reviews*, 99(2), 1281–1324. <https://doi.org/10.1152/physrev.00021.2018>
- Piras, I. S., Manchia, M., Huentelman, M. J., Pinna, F., Zai, C. C., Kennedy, J. L. & Carpiniello, B. (2018). Peripheral Biomarkers in Schizophrenia: A Meta-Analysis of Microarray Gene Expression Datasets. *International Journal of Neuropsychopharmacology*, 22(3), 186–193. <https://doi.org/10.1093/ijnp/pyy103>
- Puerta-Guardo, H., Glasner, D. R., Espinosa, D. A., Biering, S. B., Patana, M., Ratnasiri, K., Wang, C., Beatty, P. R. & Harris, E. (2019). Flavivirus NS1 Triggers Tissue-Specific Vascular Endothelial Dysfunction Reflecting Disease Tropism. *Cell Reports*, 26(6), 1598–1613.e8. <https://doi.org/10.1016/j.celrep.2019.01.036>
- Quirino-Teixeira, A. C., Rozini, S. V., Barbosa-Lima, G., Coelho, D. R., Carneiro, P. H., Mohana-Borges, R., Bozza, P. T. & Hottz, E. D. (2020). Inflammatory signaling in dengue-infected platelets requires translation and secretion of nonstructural protein 1. *Blood Advances*, 4(9), 2018–2031. <https://doi.org/10.1182/bloodadvances.2019001169>

- Raekiansyah, M., Espada-Murao, L. A., Okamoto, K., Kubo, T. & Morita, K. (2014). Dengue virus neither directly mediates hyperpermeability nor enhances tumor necrosis factor- α -induced permeability in vitro. *Japanese Journal of Infectious Diseases*, 67(2), 86–94. <https://doi.org/10.7883/yoken.67.86>
- Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D. & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, 14(9). <https://doi.org/10.1186/gb-2013-14-9-r95>
- Rieder, F., Kessler, S. P., West, G. A., Bhilocha, S., De La Motte, C., Sadler, T. M., Gopalan, B., Stylianou, E. & Fiocchi, C. (2011a). Inflammation-induced endothelial-to-mesenchymal transition: A novel mechanism of intestinal fibrosis. *American Journal of Pathology*, 179(5), 2660–2673. <https://doi.org/10.1016/j.ajpath.2011.07.042>
- Rieder, F., Kessler, S. P., West, G. A., Bhilocha, S., De La Motte, C., Sadler, T. M., Gopalan, B., Stylianou, E. & Fiocchi, C. (2011b). Inflammation-induced endothelial-to-mesenchymal transition: A novel mechanism of intestinal fibrosis. *American Journal of Pathology*, 179(5), 2660–2673. <https://doi.org/10.1016/j.ajpath.2011.07.042>
- Sena, C. M., Pereira, A. M. & Seica, R. (2013a). Endothelial dysfunction - A major mediator of diabetic vascular disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1832(12), 2216–2231. <https://doi.org/10.1016/j.bbadis.2013.08.006>
- Sena, C. M., Pereira, A. M. & Seica, R. (2013b). Endothelial dysfunction - A major mediator of diabetic vascular disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1832(12), 2216–2231. <https://doi.org/10.1016/j.bbadis.2013.08.006>

- Sharov, A. A., Schlessinger, D. & Ko, M. S. H. (2016). ExAtlas: An interactive online tool for meta-analysis of gene expression data. <http://dx.doi.org/10.1142/S0219720015500195>, *13*(6). <https://doi.org/10.1142/S0219720015500195>
- Singh, S., Anupriya, M. G., Modak, A. & Sreekumar, E. (2018). Dengue virus or NS1 protein induces trans-endothelial cell permeability associated with VE-Cadherin and RhoA phosphorylation in HMEC-1 cells preventable by angiopoietin-1. *Journal of General Virology*, *99*(12), 1658–1670. <https://doi.org/10.1099/jgv.0.001163>
- Soe, H. J., Khan, A. M., Manikam, R., Raju, C. S., Vanhoutte, P. & Sekaran, S. D. (2017). High dengue virus load differentially modulates human microvascular endothelial barrier function during early infection. *Journal of General Virology*, *98*(12), 2993–3007. <https://doi.org/10.1099/jgv.0.000981>
- Srivastava, S. P., Koya, D. & Kanasaki, K. (2013). MicroRNAs in kidney fibrosis and diabetic nephropathy: Roles on EMT and EndMT. *BioMed Research International*, *2013*. <https://doi.org/10.1155/2013/125469>
- Steyers, C. M. & Miller, F. J. (2014). Endothelial dysfunction in chronic inflammatory diseases. *International Journal of Molecular Sciences*, *15*(7), 11324–11349. <https://doi.org/10.3390/ijms150711324>
- Sweeney, T. E., Haynes, W. A., Vallania, F., Ioannidis, J. P. & Khatri, P. (2017). Methods to increase reproducibility in differential gene expression via meta-analysis. *Nucleic Acids Research*, *45*(1), 1–14. <https://doi.org/10.1093/nar/gkw797>
- Tata, P. R. & Rajagopal, J. (2016). Cellular plasticity: 1712 to the present day. *Current Opinion in Cell Biology*, *43*, 46–54. <https://doi.org/10.1016/j.ceb.2016.07.005>

- Tejera-Vaquerizo, A., Descalzo-Gallego, M. A., Otero-Rivas, M. M., Posada-García, C., Rodríguez-Pazos, L., Pastushenko, I., Marcos-Gragera, R. & García-Doval, I. (2016). Cancer Incidence and Mortality in Spain: A Systematic Review and Meta-Analysis. *Actas Dermo-Sifiliograficas*, *107*(4), 318–328. <https://doi.org/10.1016/j.ad.2015.12.008>
- Tian, D., Zeng, X., Wang, W., Wang, Z., Zhang, Y. & Wang, Y. (2019). Protective effect of rapamycin on endothelial-to-mesenchymal transition in HUVECs through the Notch signaling pathway. *Vascular Pharmacology*, *113*, 20–26. <https://doi.org/10.1016/j.vph.2018.10.004>
- Toro-Domínguez, D., Martorell-Marugán, J., López-Domínguez, R., García-Moreno, A., González-Rumayor, V., Alarcón-Riquelme, M. E. & Carmona-Sáez, P. (2019). ImaGEO: Integrative gene expression meta-analysis from GEO database. *Bioinformatics*, *35*(5), 880–882. <https://doi.org/10.1093/BIOINFORMATICS/BTY721>
- Valle, U., Teresa, M. & Valle, U. (2003). *Colombia Médica*.
- van de Weg, C. A. M., van den Ham, H. J., Bijl, M. A., Anfasa, F., Zaaraoui-Boutahar, F., Dewi, B. E., Nainggolan, L., van IJcken, W. F. J., Osterhaus, A. D. M. E., Martina, B. E. E., van Gorp, E. C. M. & Andeweg, A. C. (2015). Time since Onset of Disease and Individual Clinical Markers Associate with Transcriptional Changes in Uncomplicated Dengue. *PLoS Neglected Tropical Diseases*, *9*(3), 1–20. <https://doi.org/10.1371/journal.pntd.0003522>
- Wertz, M. S., Kyriss, T., Paranjape, S. & Glantz, S. A. (2011). The toxic effects of cigarette additives. Philip Morris' project mix reconsidered: An analysis of documents released through litigation. *PLoS Medicine*, *8*(12). <https://doi.org/10.1371/journal.pmed.1001145>
- Yacoub, S., Lam, P. K., Vu, L. H. M., Le, T. L., Ha, N. T., Toan, T. T., Thu Van, N., Quyen, N.

T. H., Duyen, H. T. Le, Kinh, N. Van, Fox, A., Mongkolspaya, J., Wolbers, M., Simmons, C. P., Srean, G. R., Wertheim, H. & Wills, B. (2016). Association of microvascular function and endothelial biomarkers with clinical outcome in dengue: An observational study. *Journal of Infectious Diseases*, 214(5), 697–706.

<https://doi.org/10.1093/infdis/jiw220>

Yaiw, K. C., Mohammad, A. A., Taher, C., Wilhelmi, V., Davoudi, B., Strååt, K., Assinger, A., Ovchinnikova, O., Shlyakhto, E., Rahbar, A., Koutonguk, O., Religa, P., Butler, L., Khan, Z., Streblow, D., Pernow, J. & Söderberg-Nauclér, C. (2014). Human cytomegalovirus induces upregulation of arginase II: Possible implications for vasculopathies. *Basic Research in Cardiology*, 109(2). <https://doi.org/10.1007/s00395-014-0401-5>

Zhang, X., Tang, N., Xi, D., Feng, Q., Liu, Y., Wang, L., Tang, Y., Zhong, H. & He, F. (2020). Human cytomegalovirus promoting endothelial cell proliferation by targeting regulator of G-protein signaling 5 hypermethylation and downregulation. *Scientific Reports*, 10(1).

<https://doi.org/10.1038/S41598-020-58680-6>

Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N. & Xia, J. (2019). NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research*, 47(W1), W234–W241.

<https://doi.org/10.1093/nar/gkz240>

ANEXOS

Tabla 10. Anexo #1. Tabla de ecuaciones de búsqueda utilizada en los repositorios científicos.

N°	Repositorio	Palabras clave	S.R.	C.R.	Filtros de la plataforma usados	Filtrados
1	PubMed	(Endothel* AND Dysfunction*) AND Dengue (Viral OR Virus) NOT review NOT vaccine NOT (animal* OR mouse OR mice)	72	28	Full text, Associated data, Journal article, 2011-2021, Human, Language.	11
	Scopus	(endothel* AND dysfunction*) AND (dengue AND (virus OR viral)) AND NOT review AND NOT vaccine AND NOT endogenous AND NOT vector* AND NOT covid AND NOT (mice OR mouse OR animal*)	104	39	2011-2021, Article, Language	24
	ScienceDirect	Endothelial dysfunction AND dengue virus NOT review NOT vaccine NOT endogenous NOT vector NOT animal NOT lymphocyte NOT transplantation	1.366	30	2012-2020, research articles.	6
	EBI-search	(endothel* AND dysfunction*) AND dengue (viral OR virus) NOT review NOT vaccine NOT endogen* NOT vector NOT (animal* OR mouse OR mice) NOT blood	95	31	Literature, 2011-2021	20
2	PubMed	(Endothel* AND Dysfunction*) AND (Viral OR Virus) NOT REVIEW NOT ENDOGENOUS NOT vectors AND infect* NOT vaccine NOT COVID-19 NOT (mice OR Mouse OR animal*) NOT transplantation	1.604	308	Full text, Associated data, Journal article, 2011-2021, Human, Language.	85
	Scopus	(endothel* AND dysfunction*) AND (virus OR viral) AND NOT review AND NOT vaccine AND NOT endogenous AND NOT vector* AND NOT covid AND NOT (mice OR mouse OR animal*) AND NOT mitochondria* AND NOT blood AND NOT lymphocyte*	62.868	360	2011-2021, Article, Final stage, Language.	107

ScienceDirect	Endothelial dysfunction AND virus NOT review NOT vector NOT endogenous NOT transplantation NOT animal NOT mouse OR mice	30.503	120	2012-2021, research articles.	22
EBI-Search	(endothel* AND dysfunction*) AND (viral OR virus) NOT review NOT endoge* NOT vector* AND infect* NOT vaccine NOT transplant* NOT cancer NOT (rat OR mice OR mouse) NOT animal* NOT mitochond* NOT lymphocyt* NOT blood NOT (clinical treatment) NOT covid	1.697	143	Literature, 2011-2021	73
3 PubMed	(Endothel* AND to AND Mesenchy* Transition* OR EndMT OR EndoMT) AND (Viral OR Virus) NOT REVIEW NOT (mice OR Mouse OR animal*)	22	8	Full text, Journal article, Human.	6
Scopus	(Endothel* to Mesenchy* Transition OR EndMT OR EndoMT) AND (virus OR viral) AND NOT review AND NOT vaccine AND NOT endogen* AND NOT vector* AND NOT (animal* OR mouse OR mice) AND NOT transplant* AND NOT lymphocyte* AND NOT blood AND NOT mitochondria* AND NOT covid	16.922	53	Open Access, 2012-2021, Article, Final stage, Journal, Human, Language.	20
ScienceDirect	(Endothelial to Mesenchymal Transition) AND virus NOT vector NOT transplantation NOT endogenous NOT animal NOT mitochondria NOT epithelial NOT cancer	7.440	25	2005-2021, Research articles.	5
EBI-search	(Endothel* to Mesenchy* Transition*) AND (virus OR viral) NOT review NOT vector NOT endogen* NOT vaccine NOT transplant* NOT (rat* OR mouse OR mice) NOT animal*	1.762	179	Literature, 2011-2021	12
4 PubMed	(Endothel* to Mesenchy* Transition OR EndMT OR EndoMT) And Dengue	0	0	-	0

Scopus	((endothel* AND to AND mesenchy* AND transition OR endmt OR endomt) AND (dengue)) AND NOT proteomic*	1	0	-	0
ScienceDirect	(Endothelial to mesenchymal transition OR EndMT OR EndoMT) AND (dengue) NOT review NOT therapeutic NOT (animal OR mouse OR mice)	186	3	Research article	1
EBI-search	(Endothel* to Mesenchy* Transition* OR EndMT OR EndoMT) AND (dengue)	7	7	Literature	1

Tabla 11. Anexo #2. Tabla de cribado y selección de datasets para el metaanálisis.

Referencia (año de publicación)	Serie	Tipo de célula	Tipo de dato	N° de muestras	Plataforma	Artículo
Erich R Mackow et al. (2012)	GSE34628	Células endoteliales primarias	Micro-arreglo	6	ENA	PMID: 22496214
Van de Weg CA et al. (2015)	E-MTAB-3162	Sangre total	Micro-arreglo	76	ENA	PMID: 25768297
Long HT et al. (2014)	GSE13052	Sangre total	Micro-arreglo	30	ENA	PMID: 19138155
Barbachano-Guerrero A (2020)	GSE139603	Células endoteliales cordón umbilical.	RNA seq.	6	NCBI	PMID: 32141809
Devignot S. et al (2010)	GSE17924	Sangre total	Micro-arreglo	48	ENA	PMID: 20652028
Oliveira LF. Et al. (2019)	GSE84536	Tejido hepático embebido en parafina fijada con formalina humana	-	15	NCBI	-
Tian Y. et al. (2020)	GSE132367	T CD4	RNA seq.	57	NCBI	PMID: 31875555
Robinson M. et al. (2019)	GSE124046	Sangre total	RNA seq.	103	NCBI	PMID: 30699342

Díaz-Badillo A. et al. (2015)	GSE2694 <u>3</u>	N.A.	Micro-arreglo	12	NCBI	PMID: 24776933
Zanini F. et al (2019)	GSE1166 <u>72</u>	Células mononucleares y Sangre total	viscRNA A-Seq	13127	NCBI	PMID: 31820734
Naim AN. Et al (2014)	GSE2898 <u>5</u>	N.A.	RNA seq.	96	NCBI	:
Weiskopf D. et al. (2019)	GSE8750 <u>5</u>	T CD4 y CD8	RNA seq.	108	NCBI	:
Zhu Y. et al. (2019)	GSE1190 <u>36</u>	Intestinales de <i>Aedes aegypti</i>	RNA seq.	4	NCBI	PMID: 31527795
Campor RK. Et al. (2019)	GSE1331 <u>11</u>	Células A549 (epiteliales alveolares)	RNA seq.	12	NCBI	:
Teramoto T. et al. (2019)	GSE1255 <u>80</u>	Las células albopictus (C6 / 36) o wMelPop-C6 / 36	RNA seq.	11	NCBI	PMID: 31092581
Cain MP et al. (2020)	GSE1446 <u>78</u>	Células endoteliales y alveolares tipo 2.	RNA seq.	2	NCBI	PMID: 32461220
Martin M, Shyy JY (2019)	GSE1217 <u>82</u>	Células endoteliales del cordón umbilical	RNA seq.	4	NCBI	N.A.
Frausto RF et al. (2019)	GSE1216 <u>80</u>	Células endoteliales corneales	RNA seq.	12	NCBI	PMID: 31194824
Glaser SF et al. (2020)	GSE1431 <u>51</u>	Células endoteliales	RNA seq.	39	NCBI	PMID: 32034099
Zhang T et al. (2019)	GSE1328 <u>33</u>	células endoteliales	ATAC seq	4	NCBI	PMID: 31341023
Doronso G et al. (2019)	GSE8889 <u>4</u>	Células endoteliales Tfeb	RNA seq.	2	NCBI	PMID: 30591554
Hébert M, Mignea	GSE1417 <u>31</u>	Células endoteliales	RNA seq.	4	NCBI	PMID: 32724121

ult F (2019)						
Zhang X. et al (2020)	GSE142762	Células endoteliales cordón umbilical.	Micro- arreglo	4	ENA	PMID: 32041970
Xianwu Lin, et al. (2021)	N.A.	Células endoteliales de microvasculat ura cerebral	RNA seq.	6	ENA	PMC (ID: PPR388296)
Messaoudi S, et al (2015)	GSE71838	Células endoteliales dérmicas	Micro- arreglo	No disponibl es en el momento	ENA	-
Trembinski DJ, et al. (2020)	GSE145697	Cardiomiocito s	RNA seq.	10	ENA	PMC32341350
Cao F., et al (2008)	-	Cardiomiocito s	Micro- arreglo	1 (Taxa)	ENA	PMC2565131
Eleanor S Roberts, et al (2003)	GSE2377	Células endoteliales	Micro- arreglo	1 (Taxa)	ENA	PMC12759259
Pluvinet R, et al (2008)	GSE10601	Células endoteliales, células T.	Micro- arreglo	1	ENA	PMID: 18669876
Fang C, et al (2011)	GSE22522	Células endoteliales	Micro- arreglo	6	NCBI	PMID: 32034099
Simone F, et al (2020)	GSE143150	Células endoteliales	Micro- arreglo	12	NCBI	PMID: 32034099