

no/a.

AUTOR

AÑO



FACULTAD DE CIENCIAS DE LA SALUD

ESCUELA DE MEDICINA

" Identificación de genes core y vías clave asociadas a fibrosis uterina mediante análisis bioinformático integrado"

Trabajo de titulación presentado en conformidad con los requisitos establecidos para optar por el título de Médico Cirujano.

PROFESOR GUÍA:

"María José Vallejo"

AUTOR:

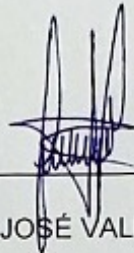
"Ana Gabriela Díaz Défaz"

QUITO, ECUADOR

2022

DECLARACIÓN DEL PROFESOR GUÍA

Declaro haber dirigido el trabajo, "Identificación de genes core y vías clave asociadas a fibrosis uterina mediante análisis bioinformático integrado" a través de reuniones periódicas con la estudiante Ana Gabriela Díaz Défaz, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación

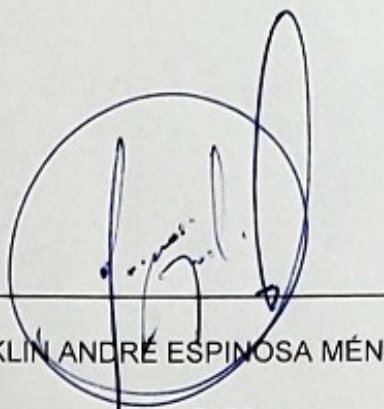


MARÍA JOSÉ VALLEJO LOPEZ

1715291140

DECLARACIÓN DEL PROFESOR CORRECTOR

"Declaro haber revisado este trabajo, "IDENTIFICACIÓN DE GENES CORE Y VÍAS CLAVE ASOCIADAS A FIBROSIS UTERINA MEDIANTE ANÁLISIS BIOINFORMÁTICO INTEGRADO", dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación."

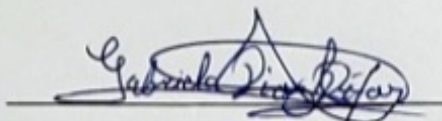
A handwritten signature in blue ink, consisting of stylized letters, is written over a horizontal line. The signature is enclosed within a large, hand-drawn circle.

FRANKLIN ANDRÉ ESPINOSA MÉNDEZ

1722016456

DECLARACIÓN DEL AUTORÍA DEL ESTUDIANTE

Declaro que este trabajo es original, de mi autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes

A handwritten signature in blue ink, which appears to read 'Ana Gabriela Díaz Défaz', is written over a horizontal line.

Ana Gabriela Díaz Défaz

1722654413

AGRADECIMIENTOS

A mi tutora, María José Vallejo, por la ayuda brindada a lo largo de este tiempo; ya que sin su apoyo y guía no hubiese sido posible realizar este trabajo de titulación.

A la Universidad de Las Américas por el conocimiento que me ha brindado durante estos 6 años de formación profesional y darme las herramientas para poder realizar esta investigación.

A mis profesores por los consejos brindados y enseñanzas que me han permitido crecer como ser humano y profesional.

A mis padres por el apoyo y soporte en toda mi vida académica. Por ser mi apoyo, fortaleza y la confianza que me han dado durante todo el camino.

DEDICATORIA

Dedicado a mis padres Rafael, Yolanda.

A mi hermano David por ser mi amigo incondicional.

A mis tías Margarita y Blanqui.

A mis abuelos Ulpiano, Mariana, Teresa y Julio.

A Monchito y Lulú.

Y a Cristhian.

ABREVIACIONES

GnRh: Hormona liberadora de gonadotropina

HMGA2: High Mobility Group AT-Hook 2, grupo AT-hook 2 de alta movilidad

COL4A5: Colágeno tipo 4 Alfa 5

COL4A6: Colágeno tipo 4 Alfa 6

MED12: Complejo mediador de la subunidad 12

ARN: Ácido Ribonucleico

ADN: Ácido Desoxirribonucleico

WNT: Sitio de integración relacionado con Wingless

B-Catenina: Beta Catenina

PLAG 1: Gen adenoma pleomórfico 1

ATP: Adenosín trifosfato

FIH1-a: Factor 1 alfa inducible por hipoxia

TGF-a: Factor de crecimiento transformante alfa

GLUT1: Transportador de glucosa cerebral de tipo 1

PDGF: Factor de crecimiento derivado de plaquetas

PI3K/AKT: fosfatidilinositol-3-kinasa/serina-treonina cinasa

TGF-B: Factor de crecimiento transformante beta

NCBI: Centro Nacional de Información Biotecnológico

Herramienta DAVID: Database for Annotation, Visualization and Integrated Discovery

KEEG: Kyoto Encyclopedia of Genes and Genomes

STRING: Search Tool for the Retrieval of Interacting Genes

.txt: Texto delimitado por tabulaciones

DEGS: Genes Expresados Diferencialmente

log-CPM: log counts per millón. Logaritmo de Conteo por millón

método TMM: Media ajustada de M-valores

GO TERMS: Gene Ontology Term. Términos de ontología génica

TGFBI: Factor transformador de crecimiento beta inducido

SERPINH1: Serpin Family H Member 1. Miembro 1 de la Familia Serpin H

COL: Familia de genes de Colágeno

HOX: Factor de transcripción homeótico

NFIB: Factor nuclear 1 tipo B

KLF4: Factor 4 Kruppel-like

FSH: Hormona foliculoestimulante

LH: Hormona luteinizante

H3AF2: Histone 3 amplified forward 2. Histona 3 amplificada hacia adelante 2

PCNA: Antígeno nuclear de células en proliferación

MCM2: Minichromosome Maintenance Complex Component 2. Componente 2 del complejo de mantenimiento de minicromosomas

RESUMEN

“Identificación de genes core y vías clave asociadas a fibrosis uterina mediante análisis bioinformático integrado”

Facultad de Ciencias Médicas, UDLA

Introducción: La Fibrosis o leiomiomas uterinos es una enfermedad del músculo liso que se caracteriza por la presencia de tumores benignos. Estos tumores son del miometrio y tienen baja tasa de malignidad. Los síntomas que pueden presentar las pacientes son metrorragia o menorragia, dismenorrea, dolor pélvico no cíclico, dispareunia, Incontinencia o retención urinarias, constipación, síntomas asociados a anemia e infertilidad (Phelippeau & Fernandez, 2016). Las bases genéticas de la fibrosis uterina se basan en las diferencias transcripcionales, estas diferencias se caracterizan por los impulsores genéticos que pueden ser reconocidos (Mehine, et al., 2016). Se conoce que las diferencias transcripcionales pueden ser: reordenamiento del grupo de alta movilidad AThook2 (HMGA2), mutaciones de la subunidad 12 del complejo mediador, inactivación bialélica del fumarato hidratasa y colágeno y Deleciones IV, alfa5 y colageno, tipo IV alfa 6 (Mehine, et al., 2016) (Laganà, et al., 2017).

Objetivo: Determinar los genes core y vías clave asociadas a fibrosis uterina mediante análisis bioinformático integrado

Materiales y métodos: Se recolecto dos datasets del repositorio GEO NCBI: GSE169255 y GSE199849. Los datos fueron procesados para determinacion de DEGS a traves de un codigo realizado en lenguaje de programacion R. Posteriormente los DEGS identificados como comunes fueron analizados en DAVID para definir terminos GO, en base a tres categorias: funcion biologica, componente celular y funcion molecular. Se utilizo KEEG para determinar vias clave asociadas a los DEGS. Finalmente se procesaron los datos en la plataforma STRING para la contruccion de redes de interaccion y CYTOSCAPE para la selección de modulo hub.

Resultados: Se identificaron un total de 444 DEGS down-regulated y 595 DEGS up-regulated. De ellos el 97.2% de genes se encuentran asociados en términos ontológicos con componentes celulares, caracterizado mayormente por la participación en la formación de matriz extracelular, citoesqueleto celular y regulación en la transcripción. Las vías clave en los genes up-regulated están asociadas a la replicación de ADN. Las vías clave dadas por los genes down-regulated se encuentran en relación con vías metabólicas como degradación de ácidos grasos, glucolisis, gluconeogénesis y resistencia a la insulina. Se determinaron como genes hub la familia de genes COL, TGFB-I y SERPHIN1 en la categoría up-regulated. En los genes down-regulated se determinaron como genes hub la familia de genes HOX, NFIB y KKL4.

Conclusiones: Los genes up-regulated se encuentran asociados con el crecimiento y proliferación de células para la formación de tumores benignos a través del aumento de matriz extracelular y aumento de la replicación celular. La regulación a la baja de la familia de genes HOX explica porque los leiomiomas no progresan a malignidad.

Palabras clave: Análisis bioinformático integrado, Fibrosis uterina, Genes core, DEGS, Vías clave, Genes HOX, Genes COL, SERPHIN1

SUMMARY

“Identification of hub genes and pathways in benign uterine fibroids by integrated bioinformatic analysis”

Faculty of Medicine, UDLA

Introduction: Uterine fibroids (leiomyomas) are a smooth muscle disease. Characterized by the presence of benign tumors of the myometrium with low rate of malignancy. The symptoms that patients may present are metrorrhagia or menorrhagia, dysmenorrhea, non-cyclical pelvic pain, dyspareunia, incontinence or urinary retention, constipation, symptoms associated with anemia and infertility (Phelippeau & Fernandez, 2016). Genetic basis is based on transcriptional differences. These differences are characterized by genetic drivers that can be recognized. (Mehine M, 2016) It is known that transcriptional differences can be: rearrangement of the high mobility group AT-hook2 (HMGA2), mutations of subunit 12 of the mediator complex (MED12), biallelic inactivation of fumarate hydratase and collagen and Deletions IV, alpha5 and collagen, type IV alpha 6 (Mehine, et al., 2016) (Laganà, et al., 2017).

Objective: To determinate the core genes and key pathways associated with uterine fibrosis through integrated bioinformatic analysis

Materials and methods: Two datasets were collected from the GEO NCBI repository: GSE169255 and GSE199849. The data was processed to determine DEGS through a code made in R programming language. Subsequently, the DEGS identified as common were analyzed in DAVID to define GO terms, based on three categories: biological function, cellular component, and molecular function. KEGG was used to determine key pathways associated with DEGS. Finally, the data was processed in the STRING platform for the construction of interaction networks and CYTOSCAPE for the selection of the hub module.

Results: A total of 444 down-regulated DEGS and 595 up-regulated DEGS were identified. Of these, 97.2% of genes are associated in ontological terms with cellular components, characterized mainly by participation in the

formation of the extracellular matrix, cellular cytoskeleton, and regulation of transcription. Key pathways in up-regulated genes are associated with DNA replication. The key pathways provided by down-regulated genes are found in relation to metabolic pathways such as fatty acid degradation, glycolysis, gluconeogenesis, and insulin resistance. The COL gene family, TGFB-I and SERPIN1 in the up-regulated category were determined as hub genes. In the down-regulated genes, the HOX gene family, NFIB and KLF4 were determined as hub genes.

Conclusions: Up-regulated genes are associated with the growth and proliferation of cells for the formation of benign tumors through increased extracellular matrix and increased cell replication. Downregulation of the HOX gene family explains why leiomyomas do not progress to malignancy.

Key words: Integrated bioinformatic analysis, Uterine fibrosis, Hub Genes, DEGS, Key pathways, Genes HOX, Genes COL, SERPIN1

Tabla de contenido

DECLARACIÓN DEL PROFESOR GUÍA	¡Error! Marcador no definido.
DECLARACIÓN DEL PROFESOR CORRECTOR.....	¡Error! Marcador no definido.
DECLARACIÓN DEL AUTORÍA DEL ESTUDIANTE	¡Error! Marcador no definido.
AGRADECIMIENTOS.....	5
DEDICATORIA	6
ABREVIACIONES	7
RESUMEN	9
SUMMARY	11
INDICE DE IMAGENES	16
INDICE DE TABLAS	18
Capítulo I - INTRODUCCIÓN	1
1.1 Marco conceptual de la investigación	1
1.2 Estrategia de búsqueda de bibliografía	4
1.3 Pregunta de investigación	5
1.4 Objetivos del Estudio	5
Objetivo general:	5
Objetivos secundarios:.....	5
1.5 Justificación del Estudio	6
Capítulo II - MARCO TEÓRICO	8
2.1 ANATOMIA Y FUNCION DEL UTERO	8
2.2 DEFINICIÓN	8
2.3 EPIDEMIOLOGÍA.....	9
2.4 FACTORES DE RIESGO.....	9
2.5 SINTOMATOLOGÍA	10
2.6 CLASIFICACION.....	11

2.7 IMPACTO	11
2.8 ROL DE LAS HORMONAS	12
2.9 EMBARAZO	12
2.10 PERIMENOPAUSIA Y MENOPAUSIA	13
Capítulo III - METODOLOGÍA	15
3.1 Diseño General de Estudio	15
3.2 Contexto o escenario	16
3.3 Hipótesis	17
3.4 Sujetos y población	17
3.4.1 Criterios de inclusión.....	18
3.4.2 Criterios de exclusión.....	18
3.5 Recolección de datos.....	18
3.6 Fuentes de datos.....	19
3.7 Bases de datos	19
3.8 Tamaño de la muestra.....	20
3.9 Variables.....	20
3.10 Análisis de los datos	23
3.10.1 Identificación de DEGS.....	23
3.10.2 Ontología Génica y KEEG pathway.....	27
3.10.3 Construcción de red PPI y selección de modulo hub	27
3.11 Consideraciones éticas	27
Capítulo IV - RESULTADOS	29
4.1 Resultados generales	29
4.1.1 Recolección de datos:	29
4.1.2 Normalización de datos	29
4.1.2 Agrupación no supervisada de muestras	31
4.1.3 Análisis de expresión diferencial.....	33
4.2 Identificación de DEGS comunes	35

4.3 Ontología Génica y KEEGs pathway	36
4.4 Construcción de red PPI y selección de módulo hub	43
Capítulo V - DISCUSIÓN	46
5.1 Hallazgos.....	46
5.2 Interpretación.....	47
Capítulo VI - CONCLUSIONES Y LIMITACIONES	51
6.1 Conclusiones	51
6.2 Limitaciones.....	52
6.3 Implicaciones de salud pública	54
Bibliografía	56
ANEXOS.....	63
Anexo 1	64
Anexo 2	65
Anexo 3	65
Anexo 4	66
Anexo 5	67
Anexo 6	68
Anexo 7	69
Anexo 8	70
Anexo 9	71
Anexo 10	72
Anexo 11	73
Anexo 12	74

INDICE DE IMAGENES

<i>Figura 1 Términos de búsqueda para la investigación en PubMed NCBI como "Leiomyoma/physiopathology OR Leiomyoma/pathology" MESH Terms</i>	5
<i>Figura 2 Flujograma de trabajo de investigación</i>	15
<i>Figura 3 Modelo de matriz utilizado en el análisis. Donde 1 es presencia de tejido en la muestra y 0 es la ausencia del tejido en la muestra. En el eje de las y se observa el número de muestra a que corresponde</i>	25
<i>Figura 4 Aplicación de la función contr.matrix. El cual contrasta entre los dos tejidos seleccionados en este caso el grupo fibroid.fibroid y myometrium.nonfibroid.</i>	25
<i>Figura 5 Identificación de DEGS comunes. Aplicación de la fórmula decideTests que determina si un gen es significativo o no significativo mediante prueba de valor $p < 0.05$ integrado en el código.</i>	26
<i>Figura 6 Base de datos GSE169255 Normalizados. Se presenta el grupo de datos donde la primera columna corresponde al nombre de muestra tamaño de librería, datos normalizados cercanos a 1 y el tipo de tejido al que pertenece.</i>	30
<i>Figura 7 Base de datos GSE199849 Normalizados. Se presenta el grupo de datos donde la primera columna corresponde al nombre de muestra tamaño de librería, datos normalizados cercanos a 1 y el tipo de tejido al que pertenece</i>	30
<i>Figura 8 Diagrama de Cajas Base de datos GSE169255. A) presenta la distribución de log-CPM en los datos no normalizados B) se presenta la normalización de datos de cada muestra después de aplicar calcNormFactor por medio del método TMM</i>	31
<i>Figura 9 Diagrama de Cajas Base de datos GSE199849. A) presenta la distribución de log-CPM en los datos no normalizados B) se presenta la normalización de datos de cada muestra después de aplicar calcNormFactor por medio del método TMM</i>	31
<i>Figura 10 Mapa de calor de la base GSE169255. En rojo intenso se presentan los genes mayormente variables hacia up-regulated y en verde intenso los genes down-regulated</i>	32
<i>Figura 11 Mapa de calor de la base GSE199849. En rojo intenso se presentan los genes mayormente variables hacia up-regulated y en verde intenso los genes down-regulated</i>	33
<i>Figura 12 Modelo de la varianza de GSE169255. En la izquierda el modelo voom en el cual la media se presenta como una hipérbola. A la derecha el modelo final en el cual se aplica Ebayesiana y la tendencia de la media se vuelve una línea recta.</i>	34
<i>Figura 13 Modelo de la varianza de GSE199849. En la izquierda el modelo voom en el cual la media se presenta como una hipérbola. A la derecha el modelo final en el cual se aplica Ebayesiana y la tendencia de la media se vuelve una línea recta.</i>	34
<i>Figura 14 Diagrama de Venn de los genes down-regulated. En la intersección el número de genes similares entre los dos datasets.</i>	35

<i>Figura 15 Diagrama de Venn de los genes up-regulated. En la intersección el número de genes similares entre los dos datasets.</i>	36
<i>Figura 16 Clúster 1 genes up-regulated. En verde los genes asociados a función matriz extracelular, amarillo organizacion de matriz extracelular y naranja estructura de la matriz extracelular.</i>	37
<i>Figura 17 Clúster 2 genes up-regulated. Verde constitución estructural de matriz extracelular. Amarillo en trímico de colágeno y Naranja fuerza tensil de la matriz extracelular</i>	38
<i>Figura 18 Clúster 3 genes up-regulated. Se observa la influencia de los genes en distintos niveles de transcripción y crecimiento celular.</i>	39
<i>Figura 19 Clúster 4 genes up-regulated. Naranja unión de filamentos de actina, verde unión de actina y amarillo citoesqueleto de actina.</i>	40
<i>Figura 20 KEEG Pathway up-regulated. Intervienen vías asociadas replicación del ADN, división celular y generación de distintos cánceres.</i>	41
<i>Figura 21 Clúster 1 genes dow-regulated. se observa la categorización en diefernetes factores de transcripción y unión de ADN.</i>	42
<i>Figura 22 KEEG Pathway genes down-regulated. Las vías asociadas están en relación con el metabolismo como degradación de ácidos grasos, glucólisis y gluconeogénesis.</i>	43
<i>Figura 23 Red PPI up-regulated. En morado se observam los genes con mayor interacción en los módulos.</i>	44
<i>Figura 24 Red PPI down-regulated. En café se observan los genes con mayor interacción en los módulos. Se osberva los genes HOX como mayormente asociados.</i>	44
<i>Figura 25 Red PPI final módulo gênicos seleccionados. Se puede observar la interacción entre los genes determinados previamente.</i>	45

INDICE DE TABLAS

<i>Tabla 1 Complicaciones en embarazo asociadas con leiomiomas uterinos (Abdullah, Massey, & Liu, 2021)</i>	<i>13</i>
<i>Tabla 2 Variables de investigación.....</i>	<i>20</i>

Capítulo I - INTRODUCCIÓN

1.1 Marco conceptual de la investigación

La Fibrosis o leiomiomas uterinos es una enfermedad del músculo liso que se caracteriza por la presencia de tumores benignos, usualmente estos tumores son del miometrio y tienen baja tasa de malignidad. Se presentan en mujeres en edad reproductiva. Los síntomas que pueden presentar las pacientes son metrorragia o menorragia, dismenorrea, dolor pélvico no cíclico, dispareunia, Incontinencia o retención urinarias, constipación y signos asociados a anemia. Así, también puede causar infertilidad en la mujer que lo padece (Phelippeau & Fernandez, 2016).

Esta sintomatología afecta de manera directa a la calidad de vida de la paciente. Unas de las afecciones más grave que puede presentar la paciente es anemia grave debido a metrorragia. En Ecuador, la ENSANUT-ECU (2012) determinó que la prevalencia de anemia en las mujeres en edad fértil es de 15.0%. La razón de anemia en este grupo de pacientes se estudió por Julia Finelstein et al. En el estudio se determinaron que el 14.4% mostraron déficit de hierro. Como es sabido una de las causas para el déficit de hierro es metrorragia durante el período menstrual de la mujer (Finkelstein, y otros, 2019).

Su prevalencia es incierta debido a que la mayor cantidad de estudios se los ha realizado en mujeres asintomáticas y la misma puede variar desde 5.4% a 77%, con mayor incidencia en el grupo afro (Stewart, Cookson , Gandolfo, & Schulze-Rat, 217). Esto demuestra que mujeres que tengan fibrosis uterina asintomática no serán diagnosticadas. Al no contar con suficientes investigaciones en mujeres asintomáticas la fibrosis uterina sigue siendo una enfermedad cuyo estudio es complejo.

La fibrosis uterina tiene dos tipos de tratamiento: control de síntomas y recesión quirúrgica de los fibromas. No existe un método de control de síntomas a largo plazo por lo cual la mayor cantidad de mujeres sintomáticas terminar en tratamientos quirúrgicos ya sean conservadores o radicales. Entre los tratamientos para control de síntomas, principalmente para control de sangrado

previo a un tratamiento quirúrgico se encuentra el uso de progestágenos, levonorgestrel, ácido tranexámico y agonistas de GnRh. La fibrosis uterina es una de las indicaciones principales para cirugía ginecológica lo que afecta directamente a la calidad de vida y a la salud de la mujer que lo padece (Laganà, y otros, 2017) (Laughlin-Tommaso, 2016).

La fisiopatología de los leiomiomas uterinos no es bien conocida, y por ello, investigaciones recientes apuntan a la búsqueda de diferentes mutaciones génicas que puedan verse relacionadas con la misma (Chabbert-Buffet, Esber, & Bouchard, 2014). Se conoce que los leiomiomas son tumores que están altamente influenciados por niveles de hormonas sexuales: estrógenos y progesterona. Se cree que podría basarse por diferencias transcripcionales, estas diferencias se caracterizan por los impulsores genéticos que pueden ser reconocidos y las alteraciones dadas pueden darse a nivel genómico, transcriptómico o proteonómico (Alba Machado-Lopez, 2021) (Mehine, y otros, 2016). Se conoce que las diferencias transcripcionales pueden ser: reordenamiento del grupo de alta movilidad AT-hook2 (HMGA2), mutaciones de la subunidad 12 del complejo mediador, inactivación bialélica del fumarato hidratasa y deleciones IV colágeno alfa 5 y alfa 6 (Mehine, y otros, 2016) (Laganà, y otros, 2017). De ellos, las alteraciones más comunes presentadas en diferentes ensayos transcripcionales son MED 12 y HMGA2 (Mehine, Makinen, Heinonen, Aaltonen, & Vatheristo, 2014). Las mutaciones nombradas anteriormente han sido relacionadas como subtipos moleculares de los leiomiomas.

Las mutaciones antes mencionadas determinan principalmente alteraciones en la regulación de la tumorigénesis. MED12 (subunidad 12 del complejo mediador) es una subunidad de un complejo proteico que actúa en la regulación de la transcripción. Este actúa a nivel del complejo de iniciación ARN-ADN polimerasa II. Su principal acción se encuentra en la vía Wnt/B-catenina que se encarga de procesos como regeneración de tejidos, diferenciación de células madre y proliferación celular. Está asociada a la formación de tumores ya que, la alteración de esta vía por medio de MED12, permite la acumulación

de B-catenina en el núcleo celular que propone una activación permanente de la vía Wnt (Laganà, y otros, 2017) (Mantilla, Suarez, Duque, & Navas, 2015).

HMGA2 es una proteína de unión a cromatina no histónica que se enfocan en el dominio estructural cromático. Su principal acción es la modulación en la transcripción de varios genes, debido a que se unen en secuencias ricas en AT. Su reordenamiento y sobreexpresión se encuentra altamente ligada con una regulación positiva de PLAG1. PLAG1 es un protooncogén pleomórfico del adenoma en el gen 1 que promueve el tumor génesis (Laganà, y otros, 2017) (Zhang , Mo, & Wang, 2019).

El ciclo de Krebs es una vía metabólica que se encarga de la oxidación de hidratos de carbono, ácidos grasos y aminoácidos para producción de ATP. El fumarato hidratasa interviene en el ciclo de Krebs mediante la catalización del fumarato a malato. Las mutaciones de fumarato hidratasa provoca una pérdida de la capacidad celular para el uso del oxígeno. Por lo que se acumula fumarato en los tejidos diana que incrementa la cantidad de ATP. Altas concentraciones de fumarato sobre activan la vía FKH1-a que se encarga de activar TGF- α , GLUT1, PDGF y VEGF que actúan en procesos de proliferación celular y angiogénesis (Laganà, y otros, 2017) (Valencia & Arteaga, 2017).

Colágeno tipo IV contiene seis cadenas alfa, que van desde $\alpha 1$ a $\alpha 6$. En la importancia de esta investigación, la cadena tipo 5 y 6 están presente en mucosa oral de piel, músculo liso y riñón. Estos forman parte de la membrana basal. Deleciones en este tipo de colágeno son principalmente ligadas al cromosoma X. Y afecta en cambios estructurales del músculo liso (Wang, y otros, 2018) (Laganà, y otros, 2017).

Por otro lado, el entendimiento de qué vías clave son las que están asociadas a la producción y desarrollo de fibrosis uterina; pueden ser un indicador para dianas futuras dianas terapéuticas. Estudios sugieren que una de las vías que puede verse afectada es la vía Wnt/B-catenina. Dicha vía ha sido estudiada previamente por estar asociada a otras patologías como lo es el cáncer de mama y cánceres ginecológicos. La vía Wnt/B-catenina se encarga de

regular proliferación, migración y homeostasis celular (Mantilla, Suarez, Duque, & Navas, 2015).

Otra vía que se supone importante es PI3K/AKT. La misma está reconocida como una vía oncogénica, ya que su alteración provoca un crecimiento celular aumentado. La PI3K, también llamada fosfoinositida 3-quinasa; es una enzima “down-regulated” que afecta a los receptores de tirosina quinasa que posteriormente activaran la vía Mtorc 1 que provoca el aumento de síntesis de proteínas asociadas a la remodelación del citoesqueleto y la producción de membranas celulares (Miriam Martini, 2014).

Por último, la vía TGF-B se ha visto también alterada de manera significativa. El estudio realizado por Lee en 2001 determina que existe una sobre expresión de TGF-B en tumores de leiomiomas uterinos a diferencia del miometrio normal. El factor de crecimiento transformante beta es un conjunto de péptidos multifuncionales que se encargan de regular el crecimiento y la diferenciación celular. Su expresión se da en la cicatrización y reparación de tejido (Alba Machado-Lopez, 2021).

1.2 Estrategia de búsqueda de bibliografía

La búsqueda de bibliografía para la siguiente investigación se realizó con el fin de identificar los últimos avances en cuanto al estudio de las bases génicas sobre la fibrosis uterina y los tipos de investigación realizados. Se utilizó el portal PubMed proporcionado y desarrollado por NCBI. Los términos utilizados para la búsqueda de información fueron: Leiomyoma/genetics AND (Leiomyoma/physiopathology OR Leiomyoma/pathology)

The screenshot shows the PubMed search interface. At the top, the NIH National Library of Medicine logo is visible. The search bar contains the query: "leiomyoma/genetics"[MeSH Terms] AND ("leiomyoma/pathology"[MeSH Terms]). Below the search bar, there are options for "Advanced", "Create alert", "Create RSS", and "Clipboard (16)". The search results section shows "142 results" and "Sorted by: Best match". There are also buttons for "Save", "Email", and "Send to". A filter bar indicates "Filters applied: Abstract, in the last 5 years, Clear all".

Figura 1 Términos de búsqueda para la investigación en PubMed NCBI como "Leiomyoma/physiopathology OR Leiomyoma/pathology" MESH Terms

De los 142 resultados se tomaron solo aquellos que tengan interés para esta investigación y estén únicamente asociados con leiomiomas uterinos. En total se tomaron 20 artículos para la producción teórica de esta investigación.

1.3 Pregunta de investigación

¿Existen genes y vías claves asociados a fibrosis uterina que pueden ser identificados mediante un análisis bioinformático?

1.4 Objetivos del Estudio

Objetivo general:

Determinar los genes Core y vías clave asociadas a fibrosis uterina mediante análisis bioinformático integrado

Objetivos secundarios:

- Identificar perfiles de expresión génica que se encuentren asociados a fibrosis uterina.
- Realizar el cribado de genes expresados diferencialmente en fibrosis uterina.
- Analizar ontología genética y vía clave de los genes expresados en fibrosis uterina.

- Construir redes de interacción y selección de gene core en fibrosis uterina.
- Validar los genes core seleccionados mediante comparación con distintos datasets.

1.5 Justificación del Estudio

Los leiomiomas uterinos, más conocido como fibrosis uterina es una enfermedad ginecológica que afecta a mujeres en cualquier edad reproductiva. Su incidencia es incierta ya que los estudios epidemiológicos realizados se han visto sesgados por el uso único de pacientes con sintomatología. La progresión a leiomiosarcoma es extremadamente rara. Sin embargo, son los síntomas que se presentan lo que afecta directamente la calidad de vida de la paciente que la padece. Uno de los principales síntomas es metrorragia, que puede desencadenar en casos severos de anemia (Phelippeau & Fernandez, 2016).

Se conoce que la fibrosis uterina son neoplasias benignas monoclonales del miometrio (Wise & Laughlin-Tommaso, 2016). Su fisiopatología no se encuentra completamente estudiada. Hipótesis sugieren que es probable que la progresión de estas neoplasias este dada por un aumento en la respuesta al estrógeno y progesterona en las células miometriales (Mehine, Makinen, Heinonen, Aaltonen, & Vatheristo, 2014).

Últimos estudios realizados han identificado posibles mutaciones que podrían llevar al aumento de dicha respuesta hormonal. Estas investigaciones se han realizado mediante análisis transcriptómicos mediante el uso de arrays de fibromas. Si bien estos estudios brindan un avance en el entendimiento de genes posiblemente involucrados, estos son realizados en grupos reducidos de pacientes in situ y de una sola localidad. Las alteraciones que se creen podrían estar asociadas son: mutaciones de MED 12, sobreexpresión de HMGA2, mutaciones de fumarato hidratasa y deleciones de COL4A5 y COL4A6. Por lo que, si bien dan un acercamiento al entendimiento de las bases genéticas de la patología, la recomendación final de manera característica de estos estudios es

el realizar a posterior un análisis bioinformático integrado que adopte todas las muestras de estos estudios y determine de manera más global y con claridad, si los genes determinados son efectivamente genes que están asociados a vías clave para el desarrollo de la patología.

De esta manera, el siguiente trabajo de investigación plantea recopilar todas las muestras que se encuentren en la base de datos del GEO NCBI respecto a muestras de leiomiomas y someterlas a un análisis bioinformático integrado. En este contexto, el siguiente trabajo de investigación tiene como objetivo llegar a determinar genes core y vías clave para el desarrollo de fibrosis uterina. Con el fin de explorar el significado biológico de los datos y las vías biológicas que están en asociación con la enfermedad. Y reforzar la determinación de los genes asociados a la fibrosis uterina. Los resultados que se den dentro de la investigación podrán ser validadas de forma externa al comparar con los resultados de otras investigaciones.

Capítulo II - MARCO TEÓRICO

2.1 ANATOMIA Y FUNCION DEL UTERO

El útero es uno de los órganos que conforman el aparato reproductor femenino. Es un órgano hueco compuesto mayormente muscular. De manera macro está compuesto por cuello, istmo y cuerpo (Bosquet, 2020) (Pöder, 2018).

El cuello del útero es un canal que conecta la vagina con el cuerpo del útero. Consta de dos porciones, una porción vaginal y una porción fija que es supravaginal. El cuello se conecta directamente con el orificio interno del istmo. El istmo es un canal que conecta cuello uterino y cuerpo del útero (Bosquet, 2020) (Pöder, 2018).

El cuerpo del útero está compuesto por tres capas. El peritoneo que es su capa externa y recubre todo el cuerpo del útero. El miometrio que a su vez se subdivide en tres zonas. El perimetrio es una capa delgada apegada al peritoneo. Se ubica encima de los vasos arcuatos. Su función es estabilizar la pared uterina. La zona media es la capa más gruesa y uniforme. Es altamente vascularizada. Se encarga de la perfusión por medio de una red vascular. La zona interna es la capa más fina y compacta. Se ubica inmediatamente debajo del endometrio. Es hipo vascular e hipoecoica, no suele visualizarse en exámenes de imagen. Esta zona separa el endometrio de la zona media del miometrio durante el ciclo menstrual. La última capa del cuerpo uterino es el endometrio. El endometrio es la capa más interna y típica de mucosa. Esta se subdivide en dos capas funcionales: el estrato basal y estrato funcional (Bosquet, 2020) (Pöder, 2018).

Su función se ve ligada a la gestación de una nueva vida. Acoge al feto para su desarrollo durante la etapa de gestación. Da nutrición y protección hasta que el feto complete con el tiempo necesario para su natalicio (Bosquet, 2020).

2.2 DEFINICIÓN

Los leiomiomas son tumores benignos del miometrio uterino que se presentan en mujeres de edad fértil. Son tumores monoclonales que nacen de una célula con mutación génica y se ven altamente influenciados por hormonas.

Se considera que para la formación de leiomiomas se ven envueltos factores genéticos, como mutaciones biológicas y ambientales. Su aparición modifica la anatomía propia del útero y altera las sus funciones (Giuliani, As-Sanie, & Marsh, 2020) (Erica & Ayman, 2018) (Abdullah, Massey, & Liu, 2021).

2.3 EPIDEMIOLOGÍA

En cuanto a su epidemiología, es variable. Ya que es influenciada por tipo de estudio, método diagnóstico y aspectos demográficos y étnicos, variando entre una prevalencia de 4.5% a 68.6%. La presentación varía según el grupo étnico, el caso de mujeres afro es detectado hasta en un 80%, y en mujeres caucásicas hasta 70%. Estos datos fueron recolectados por Baird & et. All que lograron observar que los leiomiomas son principalmente detectados en ecografía. En otros grupos étnicos como mujeres hispanas o asiáticas los datos son limitados y no pueden ser evaluados (Erica & Ayman, 2018) (Giuliani, As-Sanie, & Marsh, 2020).

La prevalencia aumenta con la edad, se considera que existe un pico de presentación a los 50 años. Posterior a este pico la prevalencia disminuye, esto puede ser determinado por la disminución de las hormonas en perimenopausia o por tratamientos invasivos realizados a estas mujeres (Erica & Ayman, 2018) (Giuliani, As-Sanie, & Marsh, 2020) (Giuliani, As-Sanie, & Marsh, 2020).

2.4 FACTORES DE RIESGO

Existen varios factores de riesgo que pueden influir en la formación de fibromas uterinos. El más importante es la raza negra, que en este caso presenta la enfermedad a edades más tempranas, mayor cantidad y mayor tamaño. Un factor de riesgo en el que se puede crear una influencia positiva es la obesidad. En la obesidad se da un aumento del tejido adiposo que beneficia a la mayor producción de estrógenos por conversión periférica. Otros factores de riesgo conocidos son nuliparidad, hipertensión, menopausia tardía, menarquia temprana e historia familiar de fibromas (Giuliani, As-Sanie, & Marsh, 2020) (Erica & Ayman, 2018) (Mara & Mohamed Ali, 2020).

2.5 SINTOMATOLOGÍA

Alrededor de la mitad de las pacientes presenta síntomas. Esta se encuentra determinada por la cantidad, tamaño y distribución de los tumores. Cabe recalcar que en mujeres Afro los síntomas se presentan a edades tempranas y cuadros más graves. Entre la clínica que se puede presentar están:

- Dolor abdominal o lumbar
- Presión pélvica
- Sensación de llenura
- Sangrado uterino anormal
- Dismenorrea
- Síntomas urinarios o intestinales
- Disfunción sexual
- Infertilidad
- Riesgo obstétrico
 - Parto prematuro
 - Necesidad de cesárea
 - Sangrados durante el embarazo
 - Restricción de crecimiento

Entre otros síntomas que pueden presentar mujeres con leiomiomas están los síntomas psicológicos-emocionales, esto por el impacto en la calidad de vida de la paciente que presenta esta enfermedad. Usualmente las pacientes pueden presentar depresión, estrés emocional y ansiedad (Giuliani, As-Sanie, & Marsh, 2020) (Pöder, 2018) (Lewis, Malik, Britten, San Pablo, & Catherino, 2018).

2.6 CLASIFICACION

Los síntomas son altamente dependientes de su localización, cantidad y tamaño, por ello es importante determinar su clasificación. Ya que de ello también dependerá el tratamiento para la paciente. Los leiomiomas pueden ser clasificados en dos categorías: tamaño y localización. La clasificación mayormente utilizada es por localización. Por el tamaño se pueden clasificar:

- **Pequeño:** diámetro es menor a 2 cm.
- **Mediano:** El diámetro varía entre 2 a 6 cm.
- **Grande:** diámetro mayor a 6 cm o el volumen del útero es similar a un embarazo de 14 semanas.

Por su localización, pueden presentarse de manera intramural, submucosa y subserosa.

- **Intramural:** Es el tipo de localización más frecuente. No distorsiona la cavidad uterina y genera síntomas leves.
- **Submucosa:** Distorsiona la cavidad uterina. Pueden ser pediculados o no pediculados
- **Subserosa:** Se presentan cuando el mioma sobresale de la superficie serosa del útero.

Ciertas presentaciones de los leiomiomas, especialmente en aquellos de gran tasa de crecimiento y tamaño no es posible clasificarlos según su localización, por lo que su tratamiento es mayormente complejo y se prefiere histerectomía (Erica & Ayman, 2018) (Pöder, 2018) (Abdullah, Massey, & Liu, 2021) (Jacques Donnez, Pablo Arriagada, Olivier Donnez & Marie-Madeleine, 2018).

2.7 IMPACTO

Por su alta prevalencia, los fibromas uterinos tienen un impacto directo en los costos de salud de manera directa e indirecta. El costo se ha estimado en pocos países. En Estados Unidos se estima que los costos anuales llegan a 34.4

billones de dólares; mientras que en Alemania y Francia bordean los 348 millones y 120 millones respectivamente. En Ecuador no se cuentan con datos de los costos anuales que se dirigen para Ecuador (Giuliani, As-Sanie, & Marsh, 2020) (Erica & Ayman, 2018).

En el caso de las pacientes, los leiomiomas uterinos presentan un impacto en su vida diaria. En primer lugar, en la calidad de vida por la sintomatología de la paciente. Psicológicamente la paciente puede verse ampliamente afectada llegando a presentar estrés emocional que desencadena en depresión y ansiedad (Giuliani, As-Sanie, & Marsh, 2020) (Mara & Mohamed Ali, 2020).

Una encuesta realizada en el 2018 en Estados Unidos demostró que las mujeres hispanas presentaban una mayor gravedad en los síntomas, con sangrados más abundantes y un impacto significativo en su actividad laboral. Por ello, este grupo prefirió el tratamiento quirúrgico para la resolución de los síntomas (Erica & Ayman, 2018).

2.8 ROL DE LAS HORMONAS

Los leiomiomas son tumores que se encuentran altamente influenciados por hormonas ováricas. Diferentes estudios han demostrado que los fibromas presentan mayor cantidad de receptores de estrógeno y progesterona a comparación del miometrio normal (Mara & Mohamed Ali, 2020) (Abdullah, Massey, & Liu, 2021).

2.9 EMBARAZO

La prevalencia conocida en el embarazo de los fibromas uterinos se encuentra en un rango de 0,1% hasta 11% se cree que puede estar infraestimada. Si bien su prevalencia es baja, la literatura informa que la presencia de esta enfermedad puede dar complicaciones obstétricas hasta en un 40%. Los efectos pueden asociarse al tamaño, localización, tipo y cantidad. Las complicaciones en el embarazo pueden ser definidas en tres categorías: durante el embarazo, intraparto o posparto y complicaciones fetales (Abdullah, Massey, & Liu, 2021). En la tabla 3.1 se describen las posibles complicaciones que se pueden presentar.

Durante embarazo	Intraparto y posparto	Complicaciones fetales
Amenaza de aborto	Retención de placenta	Disminución de crecimiento:
Aborto espontáneo	Cesárea	Reducción de miembros
Ruptura prematura de membranas	Sepsis congénita	Deformidades de cabeza no asociada a síndromes genéticos
Parto pretérmino	Hemorragia posparto	Torticollis congénito
Restricción de crecimiento intrauterino		
Placenta previa		
Radiculopatía		
Falla renal aguda		
Preeclampsia		
Incarceración uterina		
Presentación anómala		

Tabla 1 Complicaciones en embarazo asociadas con leiomiomas uterinos (Abdullah, Massey, & Liu, 2021)

2.10 PERIMENOPAUSIA Y MENOPAUSIA

Durante la perimenopausia hay una alteración de los niveles hormonales que alteran la forma en la que se desarrollan los fibromas. En esta etapa de la vida de la mujer existe una disminución gradual del estrógeno y la progesterona.

Usualmente las mujeres mayormente sintomáticas se presentan durante este momento (Mara & Mohamed Ali, 2020).

La menopausia está definida como el cese permanente de la menstruación durante 12 meses sin presentar una causa patológica. Debido a la disminución de estrógenos y progesterona, se observa que durante la menopausia los leiomiomas uterinos disminuyen de tamaño. Sin embargo, esto no suele pasar en todos los casos. En el caso de las mujeres afrodescendientes no disminuye el crecimiento y es probable que se necesite de una intervención quirúrgica para la resolución de los fibromas uterinos en esta etapa de la vida (Mara & Mohamed Ali, 2020).

Capítulo III - METODOLOGÍA

3.1 Diseño General de Estudio

La siguiente investigación se basa en un estudio bioinformático analítico. En el cual se plantea el uso de diversas herramientas informáticas para el procesamiento de datasets asociados a leiomiomas uterinos, recuperados de la base de datos global NCBI. Los pasos que se realizan en el análisis bioinformático pueden ser observados en la figura 2.



Figura 2 Flujograma de trabajo de investigación

Para llevar a cabo los pasos mencionados, se requiere el uso de los programas bioinformáticos numerados a continuación. Los mismos que se encuentran disponibles en internet para su uso de manera libre y gratuita. Los programas que se usan en el análisis bioinformático brindan la mayor cantidad de información sobre los genes identificados y su asociación con la presentación de leiomiomas uterinos.

- Lenguaje R: identificación de DEGS
- Herramienta DAVID (Database for Annotation, Visualization and Integrated Discovery): análisis ontológico

- Software y base de datos KEEG: análisis de vías clave
- Base de datos STRING (Search Tool for the Retrieval of Interacting Genes): creación de red de interacción proteína-proteína
- Software Cytoscape: análisis de genes core

3.2 Contexto o escenario

Los datos usados son recabados de una base de datos pública. Estos datos son el resultado de estudios previamente realizados en grupos reducidos de pacientes que presentaron lesiones de leiomiomas uterinos.

Desde la creación del proyecto del genoma humano en 1990, nació un enfoque diferente en la medicina conocida como medicina personalizada. Con el avance y desarrollo de nuevas tecnologías esta se transformó en medicina de precisión. La medicina de precisión busca estrategias preventivas, diagnósticas y terapéuticas que se adapten a la variabilidad de cada paciente, entendiendo que cada ser humano es un individuo único. El desarrollo de medicina personalizada cambia el paradigma clásico de las ciencias médicas en el cual se establece un modelo único basado en frecuencia de casos a un modelo adoptado según cada paciente. Con el tiempo el desarrollo de la biología, genómica y bioinformática se han convertido en pilares fundamentales para el descubrimiento de nuevos biomarcadores o dianas terapéuticas para enfermedades crónicas o alta mortalidad, siendo el principal medio de investigación y ejemplo el área de oncología; en el cual el conocimiento de la biología tumoral ha permitido la identificación de biomarcadores específicos para determinar sujetos en riesgo y su eventual diagnóstico temprano disminuyendo así las tasas de mortalidad. Los avances realizados en la medicina de precisión en el área oncológica han llevado al uso de la bioinformática en otras áreas médicas (Morocho, Espinoza, & Villarruel, 2019) (Sacristan, 2018).

Con el aumento de recursos dirigidos a la investigación génica para medicina personalizada, se han generado grandes bases de datos que pueden ser utilizadas para diferentes análisis bioinformáticos y que logran aportar nueva información o contrastar información necesaria. El análisis bioinformático puede

realizarse en cualquier tipo de enfermedad que tenga una base de datos de secuenciación génica (Sacristan, 2018) (Alonso, Laserna, Varo, Molina, & Orera, 2018).

Una de las enfermedades más prevalentes en mujeres en edad fértil en todo el mundo son los leiomiomas uterinos. Tiene una prevalencia de 77%. El diagnóstico tardío de esta enfermedad predispone a la realización de histerectomías cuando se llega a tener una afectación en la calidad de vida de la paciente, debido a metrorragia y por consiguiente anemia. Y su diagnóstico temprano, usualmente un hallazgo fortuito, solo proporciona medicina que difiere momentáneamente el crecimiento de los fibromas. En este contexto, el estudio bioinformático que plantea esta investigación busca determinar genes clave que en siguientes investigaciones puedan evaluarse como biomarcadores o dianas terapéuticas para un mejor diagnóstico y tratamiento de dicha enfermedad (Alonso, Laserna, Varo, Molina, & Orera, 2018).

3.3 Hipótesis

Existen genes core que tienen conexiones directas con el desarrollo de fibrosis uterina.

3.4 Sujetos y población

La población que se utiliza en este estudio es finita y accesible. Está constituida por datasets recopilados del repositorio GEO del Centro Nacional de Información Biotecnológico (NCBI) (<http://www.ncbi.nlm.nih.gov/geo/>). Dichos datasets son correspondientes a muestras de pacientes que presentan leiomiomas uterinos. Los datasets seleccionados son de acceso público. Se evalúan todos los datasets recopilados y se trabaja con aquellos que cumplen los criterios de inclusión y exclusión. A su vez, como criterio adicional para una normalización de datos se escogen los estudios que su tipo de experimento fuese perfiles de expresión génica de secuenciación de alto rendimiento y por lo tanto sean estudios recientes.

3.4.1 Criterios de inclusión

- Perfiles de expresión génica que se encuentren disponibles en NCBI
- Perfiles de expresión génica que cuenten con muestras de leiomioma y miometrio uterinos sano
- Perfiles de expresión génica que pertenezcan a muestras de pacientes humanas
- Perfiles cuyo tipo de experimentación sea perfiles de expresión génica de secuenciación de alto rendimiento.

3.4.2 Criterios de exclusión

- Perfiles de expresión génica que no cuenten con muestras de leiomioma uterino.
- Perfiles de expresión génica que teniendo muestras de leiomioma y miometrio uterinos, además tengan muestras de otra patología uterina.
- Perfiles de expresión génica que contengan muestras animales de leiomioma uterino.
- Perfiles de expresión génica que no tengan muestras de miometrio uterino sano de la misma paciente.

3.5 Recolección de datos

Los datos se extraen desde la base de datos de acceso libre GEO datasets del Centro Nacional Biotecnológico (NCBI), recopilado de: <http://www.ncbi.nlm.nih.gov/geo/>.

Para recolectar los datos, se ingresa en la página del ncbi (<https://www.ncbi.nlm.nih.gov/>) y se selecciona la opción GEO Dataset. EN el buscador se coloca "leiomyoma "[MeSH Terms] AND "uterus"[MeSH Terms]). Se encuentra un total de 1348 resultados, a los cuales se les aplica dos filtros: "Entry Type: Series" y "Study type: Expression profiling by array". El primer filtro se lo

toma ya que estos registros son los proporcionados por los investigadores principales, pueden contener tablas con los datos tomados, conclusión o análisis. Su forma de archivo es tipo tar, que son los archivos originales de cada investigación en formato .txt. El segundo filtro reduce la cantidad de búsquedas a los estudios que sean de nuestro interés en el tipo de estudio. En el caso de esta investigación, se selecciona perfiles de expresión génica mediante secuenciación de alto rendimiento. Una vez aplicados los dos filtros, los resultados se reducen a 27. Aplicados los criterios de inclusión y exclusión, se eliminan 23 resultandos. Se toman para el estudio los 2 resultados que pasaron los filtros y los criterios de selección.

Los datos seleccionados son aquellos que obedecen los criterios de inclusión y exclusión. Los datasets seleccionados son: GSE169255 y GSE199849. De estos, las series GSE199849 se basa en la plataforma GPL18573 correspondiente a Illumina NextSeq 500 (Homo Sapiens). GSE169255 se fundamenta en la plataforma GPL24676 Illumina NovaSeq 6000 Homo sapiens.

3.6 Fuentes de datos

La base de datos utilizada es Centro Nacional Biotecnológico (NCBI), que es una base de acceso libre y gratuito que permite el acceso a información genética y bioinformática para el avance de nuevas ciencias que sean aplicables a la medicina. De esta se utilizó su subdivisión GEO Datasets que es una base de datos en la cual investigadores pueden subir datasets de sus experimentos para ser estudiados de manera libre. La información entregada en estos datasets se descarga como archivos .rar para ser extraídos y en su contenido tienen archivos .txt en el que se presenta toda la información de cada muestra y su conteo.

3.7 Bases de datos

Para la creación de base de datos se descargaron todos los datasets. Cada data set cuenta con un archivo .txt (Texto delimitado por tabulaciones) en el que se presenta nombre de gen, nombre de muestra y el conteo que se presenta por cada muestra, en forma de una matriz.

Dichos archivos .txt se convirtieron en archivos.xlsx para su manejo y depuración en Excel. Por la cantidad de datos obtenidos por cada dataset se decidió realizar cuatro bases de datos, una por cada dataset para evitar la pérdida de datos. Para la base de datos final se mantuvieron en el libro de Excel, llamado “BASE FINAL”, únicamente las tres bases para el análisis principal. La cuarta base de datos se presenta como un archivo separado ya que es utilizada en la fase final de validación de genes.

A su vez, se crea otra base de datos llamada “Sample info” en la cual. Al igual que la base de datos previa se realiza un datasheet por cada base de datos para evitar la confusión de datos al momento de procesar. En cada datasheet se presente el nombre de la muestra, el tipo de célula que contiene y el estado de dicha célula, es decir, si es célula del miometrio o del fibroma y si es tomada de una zona fibroide o tomada de un miometrio sano.

3.8 Tamaño de la muestra

Debido al modelo de estudio se requiere la mayor cantidad de información. Una vez expuestos los criterios de inclusión y exclusión se toma como muestra todos los datasets que cumplan con los criterios. No se requiere el cálculo de la muestra, ya que se utiliza el universo muestral.

3.9 Variables

Tabla 2 Variables de investigación

Variabl e	Definición conceptual	Definición operacional	Catego ría	Tipo	Indicador	Técnica o instrume ntos
Perfiles de expresi ón génica	Medida de la actividad de expresión génica de miles genes simultáneame	Módulos de secuencias genéticas que se encuentre asociadas a fibrosis	Cualitat ivo	Nomina l	1Presencia de secuencias génicas asociadas a fibrosis uterina 2 Ausencia de secuencias	T: análisis bioinform ático

	nte para crear una imagen global de la función celular	uterina				génicas asociadas a fibrosis uterina	
Genes expresados diferencialmente (DEGS)	Niveles de expresión de genes son significativamente diferentes. Fuerza impulsadora molecular y/o biomarcadores moleculares de diferentes fenotipos	Genes que tengan un mayor nivel de expresión génica respecto a fibrosis uterina	Cualitativa	Nominal	1 estadísticamente significativo: $p < 0.5$ 2. No es estadísticamente significativo: $p > 0.5$	Es	T: análisis bioinformático I: análisis estadístico
Proteína	Clase importante de moléculas que se encuentran en las células vivas. Se	Genes codificantes de proteínas que tienen relación con procesos biológicos y/o funcionalida	Cualitativa	Nominal	1 Función molecular 2 Procesos biológicos 3 Componentes celulares		T: análisis bioinformático

compone d molecular
de una o de fibrosis
mas uterina
cadenas
largas de
aminoácido
s cuya
secuencia
correspond
e a la
secuencia
de
ADN del
gen que la
codifica

Gen Hub	Gen que tiene muchas interacciones con otros genes, generalmente tiene un papel esencial en la regulación de genes y proceso	Genes con alta correlación en un módulo genético	Cualitativo	Nominativo	1 Genes que presenten redes de conexión $P < 0.05$ 2 Genes que presenten redes de conexión $p > 0.05$	T: análisis estadístico, univariado y multivariado
----------------	---	---	--------------------	-------------------	--	---

biológicos			
-------------------	--	--	--

3.10 Análisis de los datos

3.10.1 Identificación de DEGS

Se utiliza R studio, que se basa en el lenguaje de programación R, para determinación de DEGS. El programa está dividido en tres archivos: Código GSE, GSE16 y GSE19. Código GSE contiene la programación para descarga e instalación de las bibliotecas necesarias para el procesamiento de los siguientes 2 códigos. Cada archivo contiene el modelo de programa realizado para el procesamiento de cada muestra, GSE16 procesa la base de datos GSE169255 y GSE19 procesa GSE199549.

Para el procesamiento de los datos en primer lugar es necesario filtrar los genes que presentan un bajo nivel de expresión y, por lo tanto, no son de importancia en este estudio. Para ello es necesario calcular el conteo por millón de cada gen, lo que permite generar un nuevo conteo de datos basado en el tamaño de cada base de datos. Adicionalmente se calcula el log-CPM, el cual es el logaritmo de conteos por millón, el cual se usa para un aumento en la precisión del conteo, de esta manera los genes que tengan el mismo conteo en bruto tendrán el mismo conteo en el cálculo de conteo por millón, lo que es un punto fundamental en la normalización de datos que se describe más adelante. Una vez aplicado log-CPM se filtra los genes con menor expresión, para ello se utiliza la función `filterByExpr` proporcionado por la librería `edgeR`. De manera automática la función mantiene genes que tengan un conteo igual o mayor a la mediana de la base de datos en todos los tipos de células que se presentan dentro de ella (Chen, Lun, & Smyth, 2022).

3.10.1.1 Normalización de datos

Ya que se han filtrado los genes con baja expresión génica, se desconoce el nuevo tamaño de las bases de datos y si la cantidad de datos entre sus muestras es similar. Es en este momento que se plantea realizar la última normalización de la base de datos para entrar al estudio de la identificación de los DEGS para asegurar que el rango y la distribución de valores es normal. En este caso se usa la función `calcNormFactors` que viene integrada en el paquete `Limma`. El cual calcula los factores de normalización mediante el método TMM. El método TMM (en español Media recortada de los valores M) fue propuesto por Robison y Oshlack en el año 2010. El método se basa en una medida estadística de tendencia central, en la cual se recalcula la media después del descarte de genes en el extremo bajo de expresión.

3.10.1.2 Agrupación no supervisada de muestras

Cuando los datos han sido normalizados se puede realizar agrupaciones jerárquicas no controladas. La generación de la agrupación jerárquica tiene como objetivo buscar patrones dentro de las muestras estudiadas. El caso en esta investigación se ha decidido crear un heatmap, la cual es la representación gráfica del cálculo de las distancias euclidianas de los log-CPM de cada muestra. La distancia euclídea es la longitud del segmento que une dos puntos. Si en un espacio dimensional, se toma a un punto x de distancia como el número de observaciones que tiene una variable, toda variable que tenga ese número de observaciones tendrá la misma distancia. La interpretación de la distancia euclídea después de las normalizaciones datos es proporcional a $(1-r)$ donde r es el coeficiente de correlación de Pearson entre los dos genes a validar. Todos los cálculos los hace el programa de manera intrínseca (Chen, Lun, & Smyth, 2022) (Ayala, 2022) (Rodrigo, 2017) (Anders, 2010).

Se realiza la creación de heatmap por cada base de datos con el uso de la función `aheatmap`, en el que previamente se han seleccionado los 100 genes más variables los cuales pueden ser up-regulated or down-regulated.

3.10.1.3 Análisis de expresión diferencial

En el análisis de la expresión diferencial se inicia con la creación de una matriz con la función `model.matrix`. Esta matriz está compuesta por el número de muestras en la primera columna y los tipos de tejidos en la primera fila como encabezados, que existen dentro la base de datos. Este tipo de matriz está lista para aplicar modelos lineales a los datos. En esta se determina 1 como la presencia del tejido en la muestra y 0 como la ausencia del tejido en la muestra cómo se observa en la figura 3

	fibroid.fibroid	myometrium.fibroid	myometrium.nonfibroid
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	1
5	0	0	1
6	0	0	1
7	0	1	0
8	0	1	0
9	0	1	0
10	0	1	0

Figura 3 Modelo de matriz utilizado en el análisis. Donde 1 es presencia de tejido en la muestra y 0 es la ausencia del tejido en la muestra. En el eje de las y se observa el número de muestra a que corresponde

De esta matriz se realiza el contraste en las muestras de interés, que son el tejido fibroso y el miometrio no fibroso. Se procede a aplicar la función `contr.matrix` con estos dos grupos. Figura 4

```
> contr.matrix16 <- makeContrasts(
+   FvsNF = fibroid.fibroid - myometrium.nonfibroid,
+   levels = colnames(design16))
> contr.matrix16
```

Levels	Contrasts
	FvsNF
fibroid.fibroid	1
myometrium.fibroid	0
myometrium.nonfibroid	-1

Figura 4 Aplicación de la función `contr.matrix`. El cual contrasta entre los dos tejidos seleccionados en este caso el grupo `fibroid.fibroid` y `myometrium.nonfibroid`.

Una vez hecho el procesamiento se utiliza la herramienta `voom` para graficar la tendencia de la variancia de los datos respecto a la media. Con esto se puede calcular el peso apropiado de cada observación de manera mas precisa. `Voom` toma los conteos y los convierte en valores `logCPM` con una

sumatoria de 0.5 a todas las muestras para evitar tener logaritmos de cero. La matriz se normaliza a sí misma. Posteriormente se usa la función `limfit` la cual ajusta modelos lineales para cada fila. Del ajuste del modelo lineal se puede aplicar `contrasts.fit` el cual estima y obtiene coeficientes y errores standard de cada matriz contrastada. Por último, se utiliza un modelo empírico de Bayes que realiza una prueba t-estadística moderada en la cual contrasta cada fila con cero. Este tipo de pruebas es análogo a la relación de pruebas t y pruebas F que se realiza en un ANOVA convencional. Sin embargo, el modelo de Bayes permite moderar los grados de libertad entre las filas mismas (McCarthy, 2009) (Páez, 2011).

3.10.1.4 Identificación de DEGS comunes

Se ejecuta la función `decideTest` la cual mediante pruebas t-estadísticas determina si un gen es up-regulated, down-regulated o no significativo. `decideTest` cuenta de manera integrada con el valor de $p < 0.05$ y la matriz resultante se presenta con resultados de -1,0,1 por cada gen. En el cual 0 se refiere a no estadísticamente significativo; 1 a genes up-regulated y -1 genes down-regulated. Para conocer de que genes se tratan se descarga la información mediante `write.csv`.

```
decide17 <- decideTests(efit17)
summary(decide17)

write.csv(efit17, file = "decide17.csv", row.names = FALSE)
```

Figura 5 Identificación de DEGS comunes. Aplicación de la fórmula `decideTests` que determina si un gen es significativo o no significativo mediante prueba de valor $p < 0.05$ integrado en el código.

En este se presenta el nombre del gen y a continuación si es 1,0 o -1. Es este momento en el que análisis de todas las bases de datos se une. Para la comprobación de los genes estadísticamente significativos. Existen muchas maneras para la identificación de DEGS comunes, en este caso se ha decidido usar una herramienta de generación de diagramas de Ven, de acceso libre y gratuito; que a su vez genera un archivo de texto con los nombres de los genes. El programa puede ser encontrado en

<https://bioinformatics.psb.ugent.be/webtools/Venn/> . En este se suben los archivos decide previamente descargados y se determina cuantos genes up-regulated y down-regulated son capaces de entrar a la siguiente fase del estudio

3.10.2 Ontología Génica y KEEG pathway

Se ingresan la lista de genes a la página de DAVID del NCBI, que es un programa en línea (<https://david.ncifcrf.gov/summary.jsp>), el cual proporciona GO TERMS en tres categorías: función biológica, componente celular y función molecular. Estos son presentados en forma de grupos, agrupando la mayor cantidad de genes que participen en los mismos procesos biológicos. Los grupos son seleccionados por el valor de p, aquellos que tienen relevancia génica son los que son $p < 0.05$, con un score de enriquecimiento mayor a > 1 . Mientras mas alto el valor más posibilidad hay que este se desplace hacia los extremos de la regulación sea positiva o negativa. DAVID cuenta con un grupo de bases de datos que son utilizados para el entendimiento y la formación de términos ontológicos y para el enriquecimiento KEEG (B.T. Sherman, 2021) (Huang DW, 2009).

3.10.3 Construcción de red PPI y selección de modulo hub

Realizado el análisis de go terms y KEEG pathways, se seleccionan los genes que tengan mayor significancia génica. Estos genes, ya como proteínas son llevados a la plataforma STRING que permite la creación de módulos y redes de interacción, según coexpresion, fusión génica, concurrencia filogenética, homología, interacción experimental. Todos estos datos se descargan como archivos. tsv y se suben al programa CYTOSCAPE, en el cual se termina de formar la red de interacción.

3.11 Consideraciones éticas

Todos los datos y recursos utilizado en este estudio tuvieron un propósito netamente académico. La base de datos del NCBI es completamente de acceso libre. Los datos obtenidos de la base Geo-NCBI son completamente anónimos y de dominio público. El estudio no viola ningún derecho a la confidencialidad ni vulnera derechos de ningún individuo. Ninguno de los datasets tomados de

muestras de pacientes utilizados en este estudio fue llevado a cabo por la autora de esta investigación. Por lo que, no se requiere la aplicación de un consentimiento informado o solicitud de acceso a la base de datos. Este trabajo está realizado en base a un protocolo de titulación previamente aprobado por el comité de titulación de la Universidad.

Capítulo IV - RESULTADOS

4.1 Resultados generales

4.1.1 Recolección de datos:

El número total de datasets tomados para el estudio son 2 series. Cada serie cuenta con un número de variables y observaciones totales. La serie de datos GSE169255 cuenta con 33538 observaciones de 18 variables. Se entiende como variable a cada muestra que se presenta en la serie de datos. De las 18 variables, 6 corresponden a muestras de miometrio no fibroide, 6 a muestras de miometrio fibroide y 6 a muestras de leiomiomas. La serie de GSE199849 cuenta con 19922 observaciones de 154 variables. De las variables, 31 muestras son de miometrio fibroide, 50 representa fibromas, 23 muestras son de endometrio fibroide, 14 de pseudocapsula fibrosa, 10 muestras a miometrio no fibroide y 8 a muestras de endometrio no fibroide.

El total de muestras recolectadas son 193, de las cuales 16 se corresponden a miometrio no fibroide, 56 a muestras de fibromas, 37 a miometrio fibroide, 23 muestras a endometrio fibroide, 14 a pseudocapsula fibroide y 8 muestras a endometrio no fibroide. El total de observaciones o conteos de todas las series de datos es 53460.

4.1.2 Normalización de datos

Cuando se aplica la normalización de datos por medio de `calcNormFactors` las muestras presentan factores de normalización cercanos a 1.

	group	lib.size	norm.factors	group16
M1	1	16503807	1.0814373	myometrium.nonfibroid
M2	1	16450623	1.0344670	myometrium.nonfibroid
M3	1	16776900	1.0844986	myometrium.nonfibroid
M4	1	19004514	0.9520087	myometrium.nonfibroid
M5	1	19775949	1.0720769	myometrium.nonfibroid
M6	1	17134442	1.1403418	myometrium.nonfibroid
MF1	1	21126188	0.9998438	myometrium.fibroid
MF2	1	19448140	1.0198845	myometrium.fibroid
MF3	1	19102316	0.9289079	myometrium.fibroid
MF4	1	17252477	1.0671597	myometrium.fibroid

Figura 6 Base de datos GSE169255 Normalizados. Se presenta el grupo de datos donde la primera columna corresponde al nombre de muestra tamaño de librería, datos normalizados cercanos a 1 y el tipo de tejido al que pertenece.

	group	lib.size	norm.factors	group19
D100F	1	13018700	0.8551154	fibroid.fibroid
D108E	1	15423651	1.2132647	endometrium.nonfibroid
D108M	1	12073826	1.1847885	myometrium.nonfibroid
D109F	1	16144224	0.8625350	fibroid.fibroid
D109M	1	12484078	0.8424721	myometrium.fibroid
D10F	1	24674726	0.9919486	fibroid.fibroid
D111M	1	10969783	0.9138482	myometrium.nonfibroid
D113F	1	20138701	0.9040888	fibroid.fibroid
D116M	1	14112534	0.9303945	myometrium.nonfibroid
D119F	1	10626125	0.9877186	fibroid.fibroid

Figura 7 Base de datos GSE199849 Normalizados. Se presenta el grupo de datos donde la primera columna corresponde al nombre de muestra tamaño de librería, datos normalizados cercanos a 1 y el tipo de tejido al que pertenece.

En la figura 6 y figura 7 se observan los nombres de muestras, el tamaño de cada muestra, el factor normalizado y el tipo de muestra al que corresponde. Se presentan las primeras 10 muestras de cada base de datos. Figura 6 corresponde a la base de datos GSE169255. Figura 7 corresponde a la base de datos GSE199849.

Para demostrar de mejor manera la normalización de datos. Los datos se presentan en diagramas de cajas donde se presenta una comparación de los datos no normalizados y normalizados.

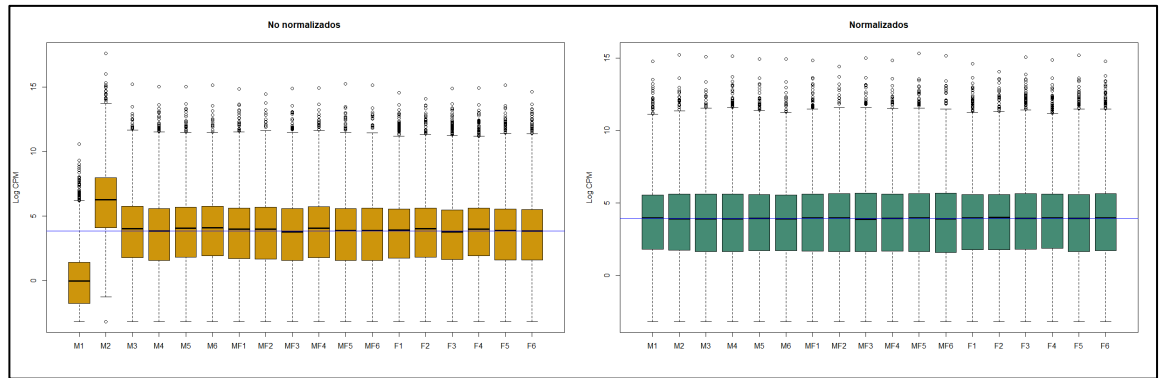


Figura 8 Diagrama de Cajas Base de datos GSE169255. A) presenta la distribución de log-CPM en los datos no normalizados B) se presenta la normalización de datos de cada muestra después de aplicar calcNormFactor por medio del método TMM

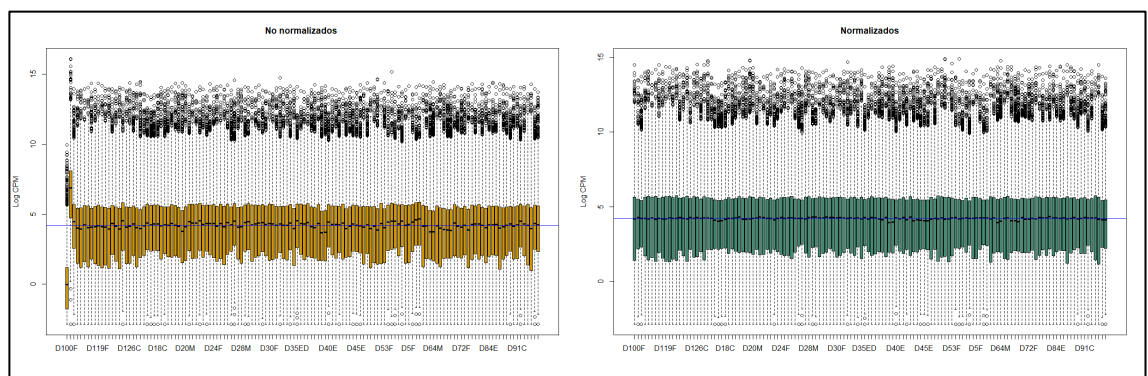


Figura 9 Diagrama de Cajas Base de datos GSE199849. A) presenta la distribución de log-CPM en los datos no normalizados B) se presenta la normalización de datos de cada muestra después de aplicar calcNormFactor por medio del método TMM

4.1.2 Agrupación no supervisada de muestras

Las figuras 10 y 11 presentan los mapas de calor de cada base de datos de los 250 genes más variables de cada base. Se puede observar en el color Key ubicado en la zona superior derecha que el color rojo refleja genes up-regulated y la zona verde genes down-regulated. Se puede observar que en los gráficos que las células tipo fibromas contienen la mayor cantidad de genes variables.

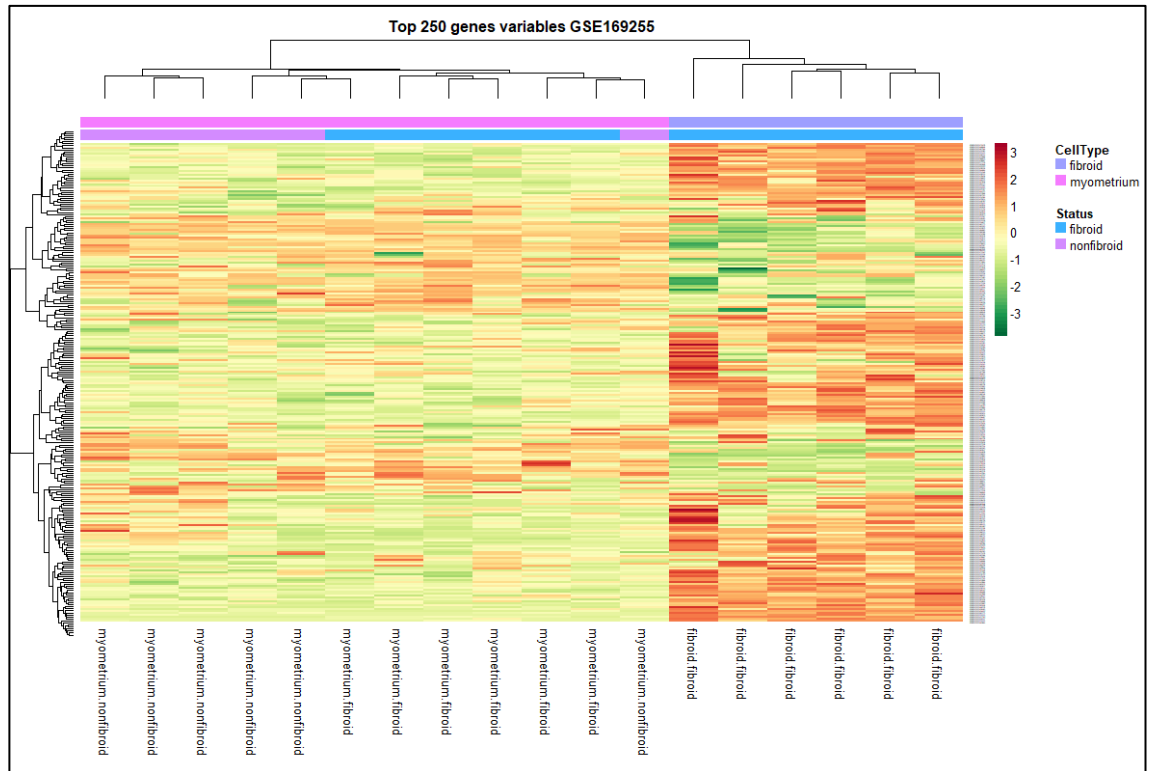


Figura 10 Mapa de calor de la base GSE169255. En rojo intenso se presentan los genes mayormente variables hacia up-regulated y en verde intenso los genes down-regulated



Figura 11 Mapa de calor de la base GSE199849. En rojo intenso se presentan los genes mayormente variables hacia up-regulated y en verde intenso los genes down-regulated

4.1.3 Análisis de expresión diferencial

A continuación, se presentan los gráficos de tendencia de la media de cada base de datos. En los gráficos de la izquierda se observa la aplicación de voom y cómo afecta la tendencia generando una parábola, cada punto en la gráfica representa un gen. En las gráficas del lado derecho se observa la aplicación de la Ebayes, en la cual se ve la normalización de la media. Y como genera una línea recta horizontal.

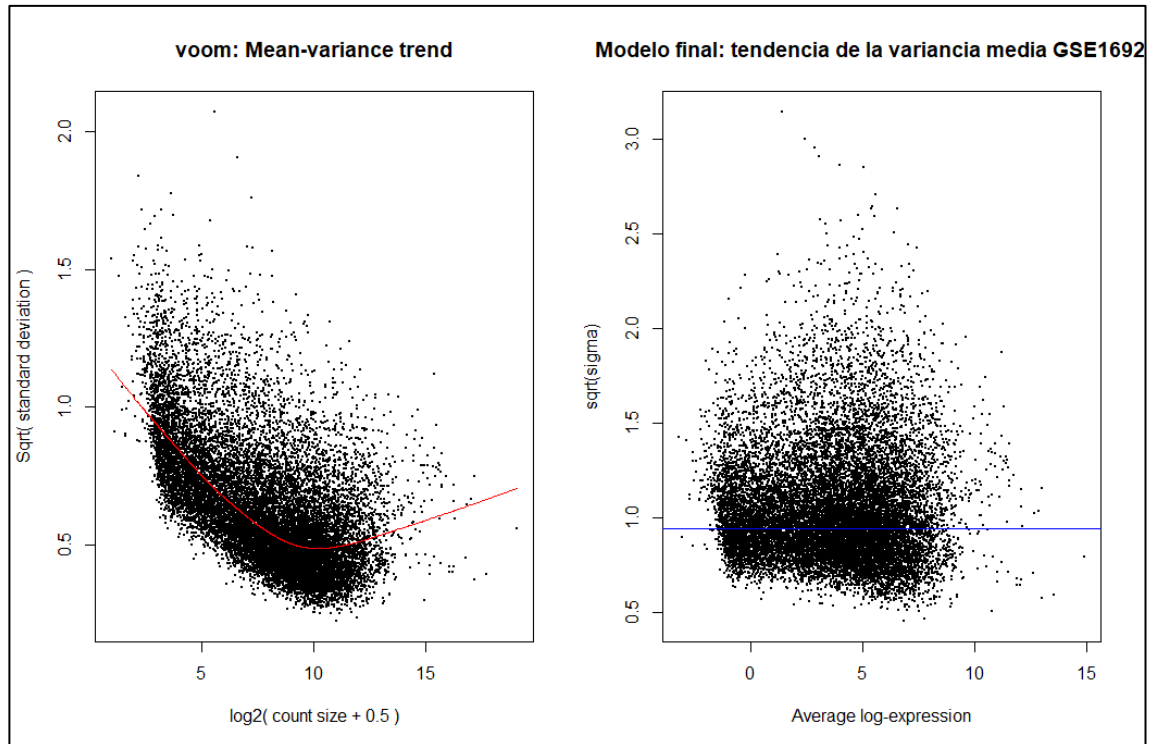


Figura 12 Modelo de la varianza de GSE169255. En la izquierda el modelo voom en el cual la media se presenta como una hipérbole. A la derecha el modelo final en el cual se aplica Ebayesiana y la tendencia de la media se vuelve una línea recta.

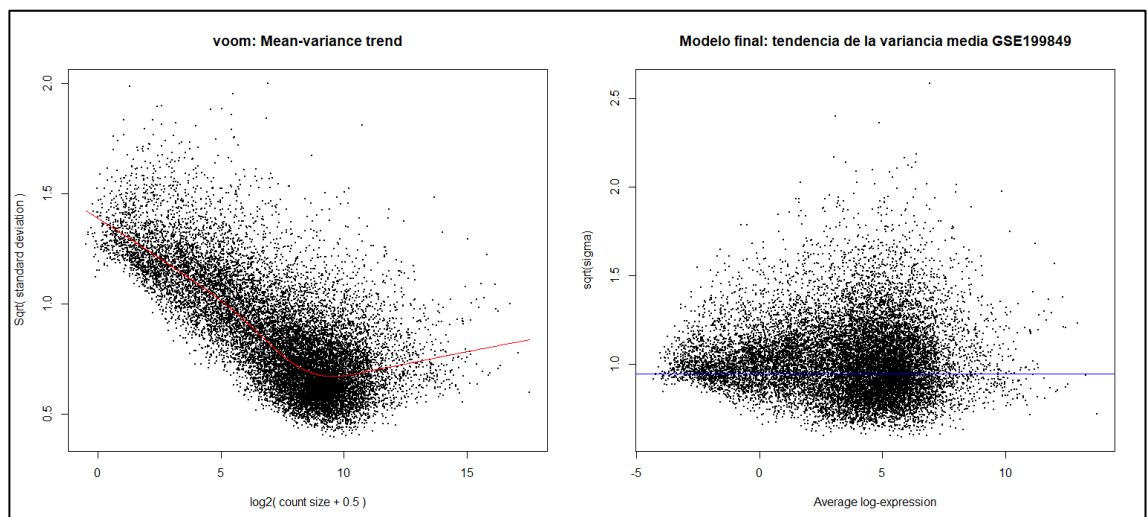


Figura 13 Modelo de la varianza de GSE199849. En la izquierda el modelo voom en el cual la media se presenta como una hipérbole. A la derecha el modelo final en el cual se aplica Ebayesiana y la tendencia de la media se vuelve una línea recta.

4.2 Identificación de DEGS comunes

Se presentan los gráficos de diagramas de ven respecto a los genes diferencialmente expresados que son comunes entre las bases de datos. En las intersecciones se presenta la cantidad de genes que son similares entre las muestras. Un total de 444 DEGS down-regulated fueron sustraído entre todas las muestras.

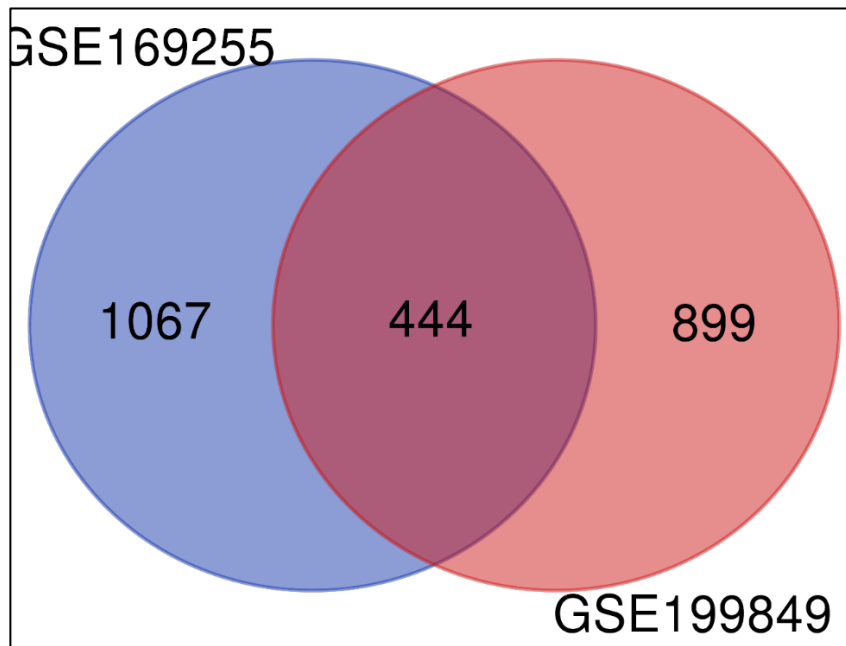


Figura 14 Diagrama de Venn de los genes down-regulated. En la intersección el número de genes similares entre los dos datasets.

De los genes up-regulated 595 son compartidos por GSE169255 y GSE199849.

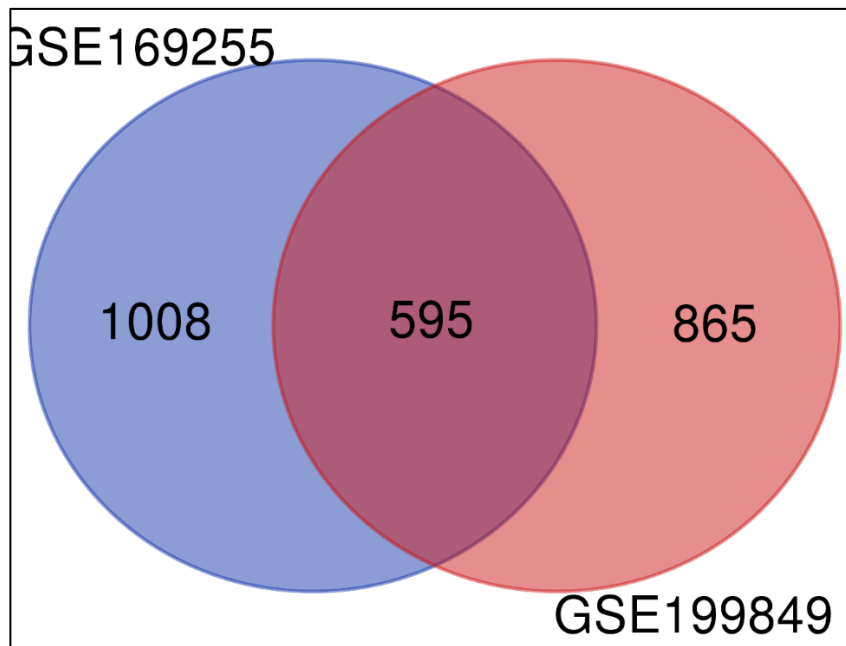


Figura 15 Diagrama de Venn de los genes up-regulated. En la intersección el número de genes similares entre los dos datasets.

4.3 Ontología Génica y KEEGs pathway

En cuanto a los términos ontológicos, se observa que de los 595 genes up-regulated presentados el 92.8% tiene asociación directa con procesos biológicos, 97.2% con componentes celulares y 92.8% con morfología celular. En cuanto a clúster, se escogió aquellos que tuviesen puntuación de enriquecimiento mayor a 1. Y que en el conteo de genes contase con un conteo mayor a 15 genes participantes. Bajo estos criterios se seleccionaron 4 grupos de clústers. En los clústeres se observa datos de 0 o 1. 0 corresponde a que el gen no está asociado ni reportado como influyente o asociado con la función génica y 1 como un gen ya reportado que se encuentra asociado a la función génica

El clúster numero 1 tiene un score de enriquecimiento de 11.55, en el intervienen la generación, estructuración y organización de la matriz extracelular, en verde se observan los genes reportados positivamente con la función de matriz extracelular, en amarillo los genes reportados en la organización de matriz extracelular y en naranja aquellos reportados en la estructura de la matriz extracelular.



Figura 16 Clúster 1 genes up-regulated. En verde los genes asociados a función matriz extracelular, amarillo organización de matriz extracelular y naranja estructura de la matriz extracelular.

El clúster número 2 tiene un score de 6.26, en el interviene la constitución estructural de la matriz extracelular, trímero de colágeno y confección de la fuerza tensil de la matriz extracelular, en verde se observan los genes reportados positivamente con constitución estructural de la matriz extracelular, en amarillo los genes reportados en trímero de colágeno y en naranja aquellos reportados en confección de la fuerza tensil de la matriz extracelular.

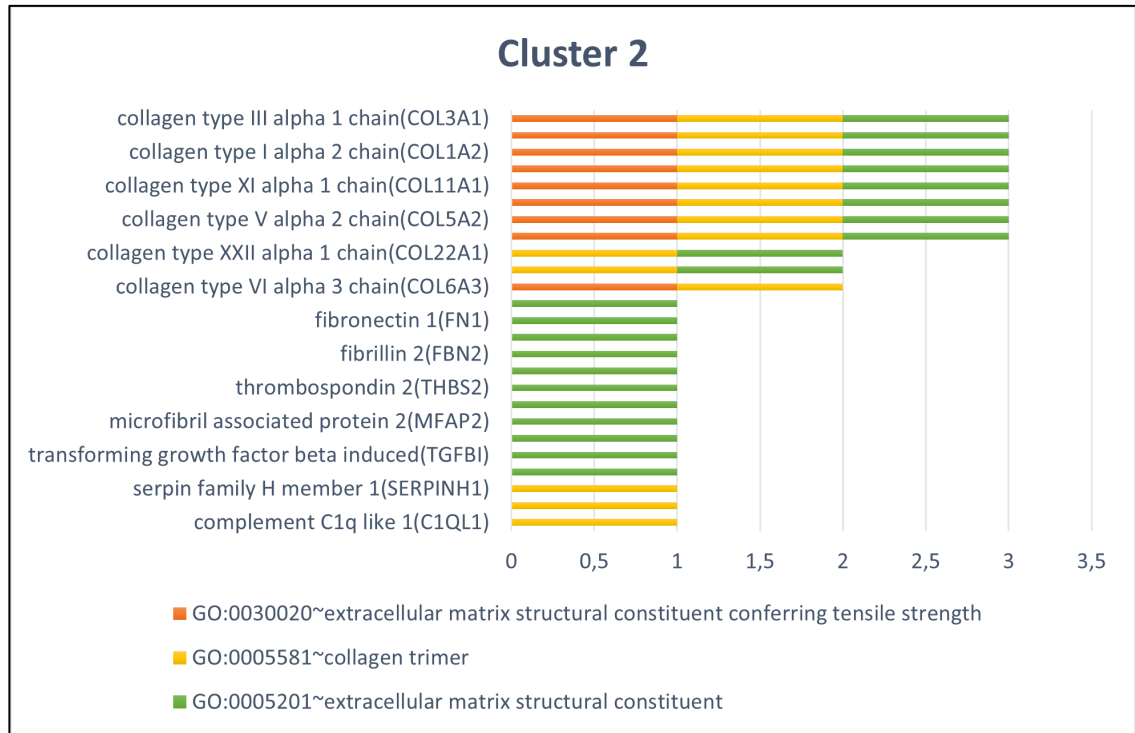


Figura 17 Clúster 2 genes up-regulated. Verde constitución estructural de matriz extracelular. Amarillo en trímero de colágeno y Naranja fuerza tensil de la matriz extracelular

El clúster número 3 tiene un score de 0.85, en el interviene regulación de la transcripción del promotor de la ARN polimerasa II, cromatina, regulación positiva de la transcripción del promotor de la ARN polimerasa II, unión de ADN específica de secuencia, actividad del factor de transcripción-unión de ADN específica de secuencia, actividad del activador transcripcional-unión específica de secuencia de la región reguladora de la transcripción de la ARN polimerasa II, unión de ADN específica de secuencia de la región proximal del promotor del núcleo de la ARN polimerasa II, actividad del factor de transcripción de la ARN polimerasa II, unión de ADN específica de secuencia y regulación de la transcripción del promotor de la ARN polimerasa II.

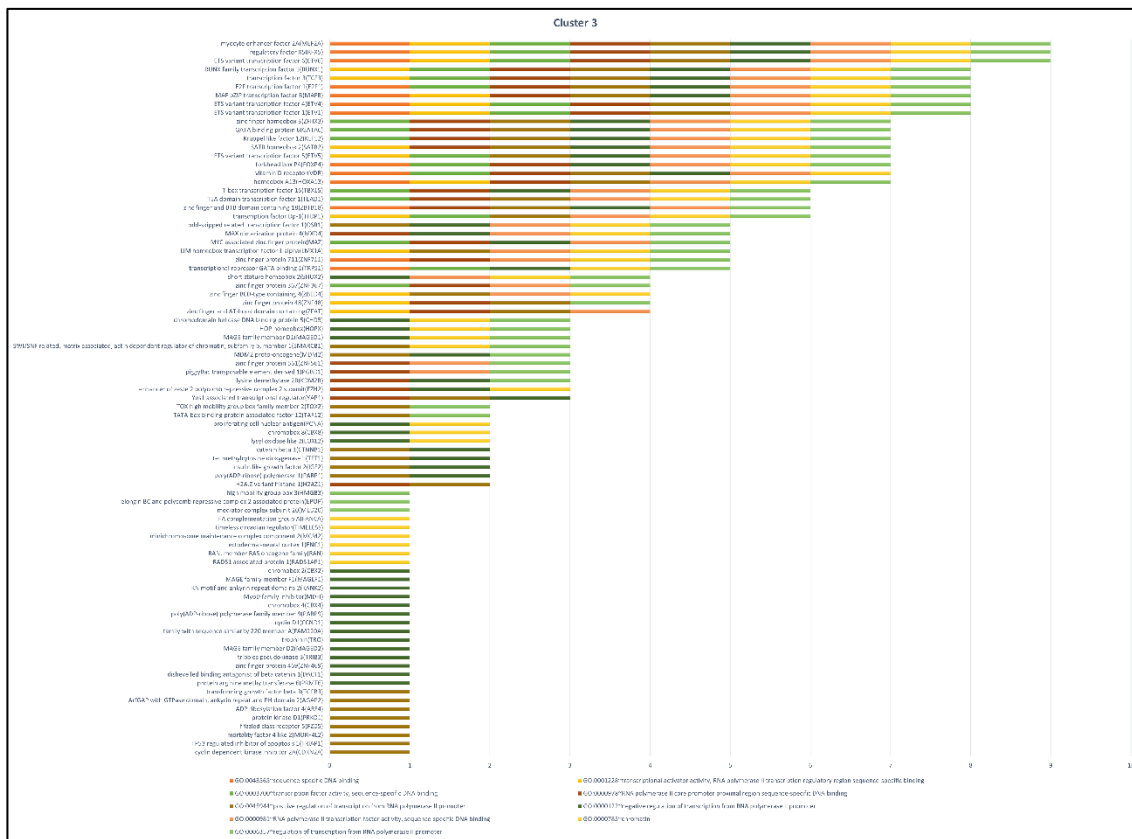


Figura 18 Clúster 3 genes up-regulated. Se observa la influencia de los genes en distintos niveles de transcripción y crecimiento celular.

Por último, el clúster número 4 tiene un score de 0.82, en el interviene unión de actina, citoesqueleto de actina y unión de filamentos de actina.

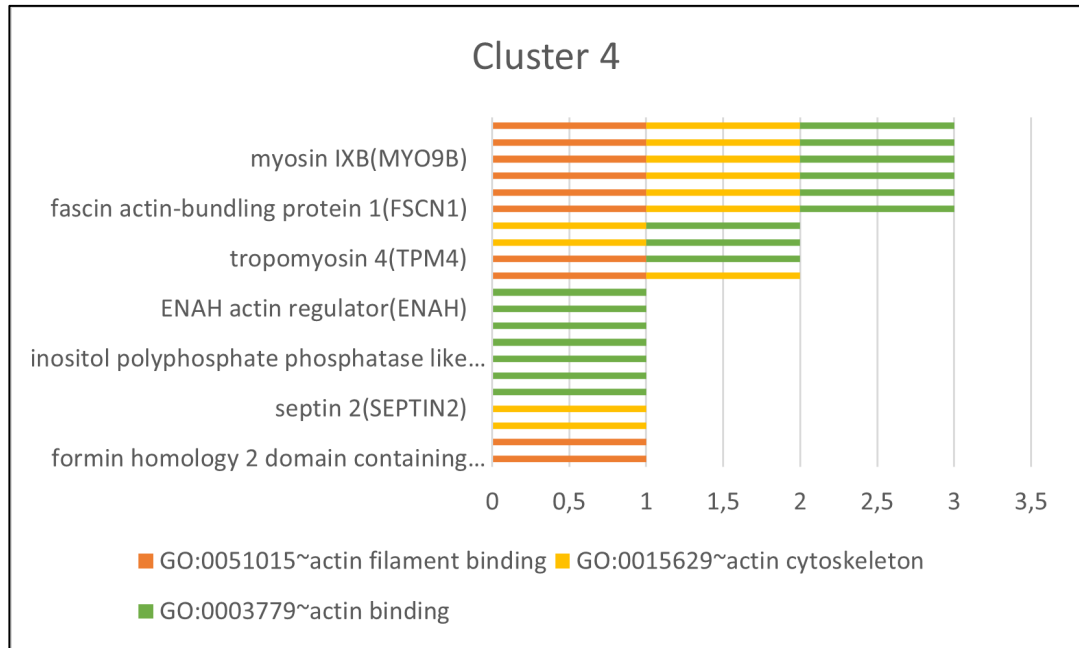


Figura 19 Clúster 4 genes up-regulated. Naranja unión de filamentos de actina, verde unión de actina y amarillo citoesqueleto de actina.

Los resultados previos recalcan que los DEGS up-regulated demuestran que todos participan de una manera global en las tres funciones ontológicas, sin embargo, se ven mayormente asociados a la formación de la matriz extracelular del citoesqueleto celular y en la regulación de la transcripción.

En el keeg pathway de los genes regulados up-regulated el 42.6% de los genes lo forman, es decir 255 genes. En el cual se puede observar que la mayor cantidad de vías se encuentran asociadas a la replicación del ADN, en la señalización de p53, que está asociada en la división y lisis celular, en la generación de distintos tipos de cáncer, en digestión y absorción de proteínas, en la adhesión de células y el ciclo celular

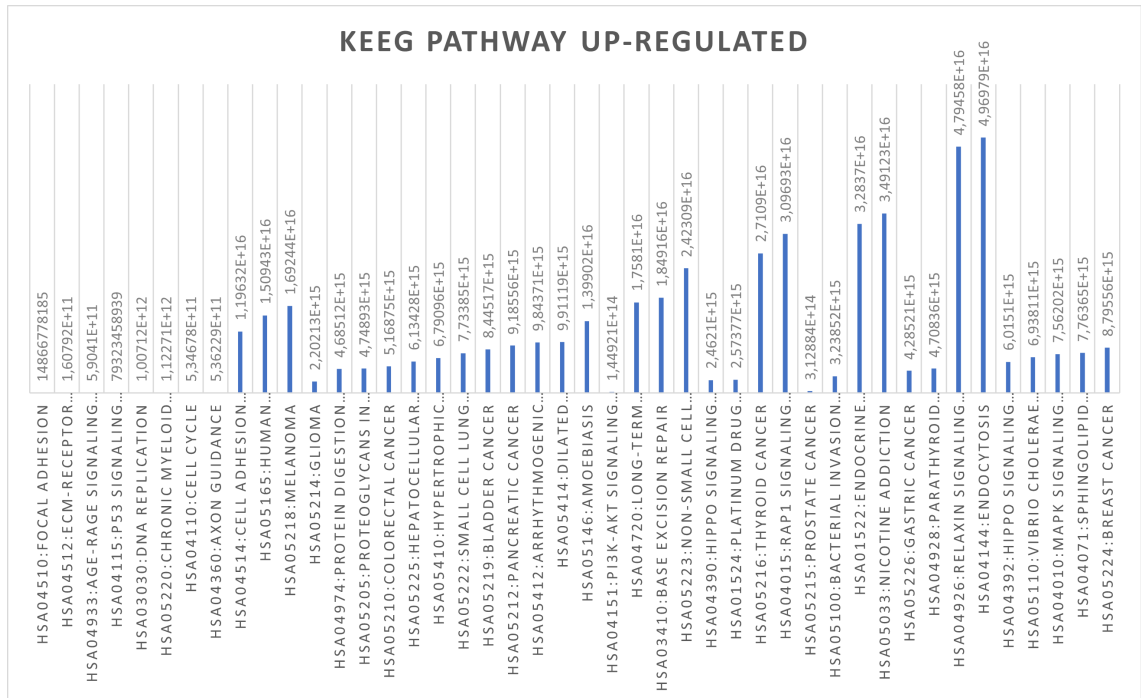


Figura 20 KEEG Pathway up-regulated. Intervienen vías asociadas replicación del ADN, división celular y generación de distintos cánceres.

En cuanto a los términos ontológicos, se observa que de los 444 genes down-regulated presentados el 89.9% tiene asociación directa con procesos biológicos, 93.9% con componentes celulares y 94.6% con morfología celular. En cuanto a cluster, se escogió aquellos que tuviesen puntuación de enriquecimiento mayor a 1. Y que en el conteo de genes contase con un conteo mayor a 15 genes participantes. Bajo estos criterios se seleccionó 1 grupo de clústeres.

En el figura 21, presentado a continuación se observa que las categorías presentadas en el clúster son: especificación de patrón anterior/posterior, regulación negativa de la transcripción del promotor de la ARN polimerasa II, regulación positiva de la transcripción del promotor de la ARN polimerasa II, unión de ADN de doble cadena específica de secuencia, cromatina, actividad del factor de transcripción, unión de ADN específica de secuencia, actividad del activador transcripcional-unión específica de secuencia de la región reguladora de la transcripción de la ARN polimerasa II, regulación de la transcripción del promotor de la ARN polimerasa II, actividad del factor de transcripción de ARN polimerasa II-unión de ADN específica de secuencia, Unión de ADN específica

de secuencia de la región proximal del promotor del núcleo de la ARN polimerasa II, Unión al ADN.

La agrupación tiene un score de enriquecimiento de 1.19.

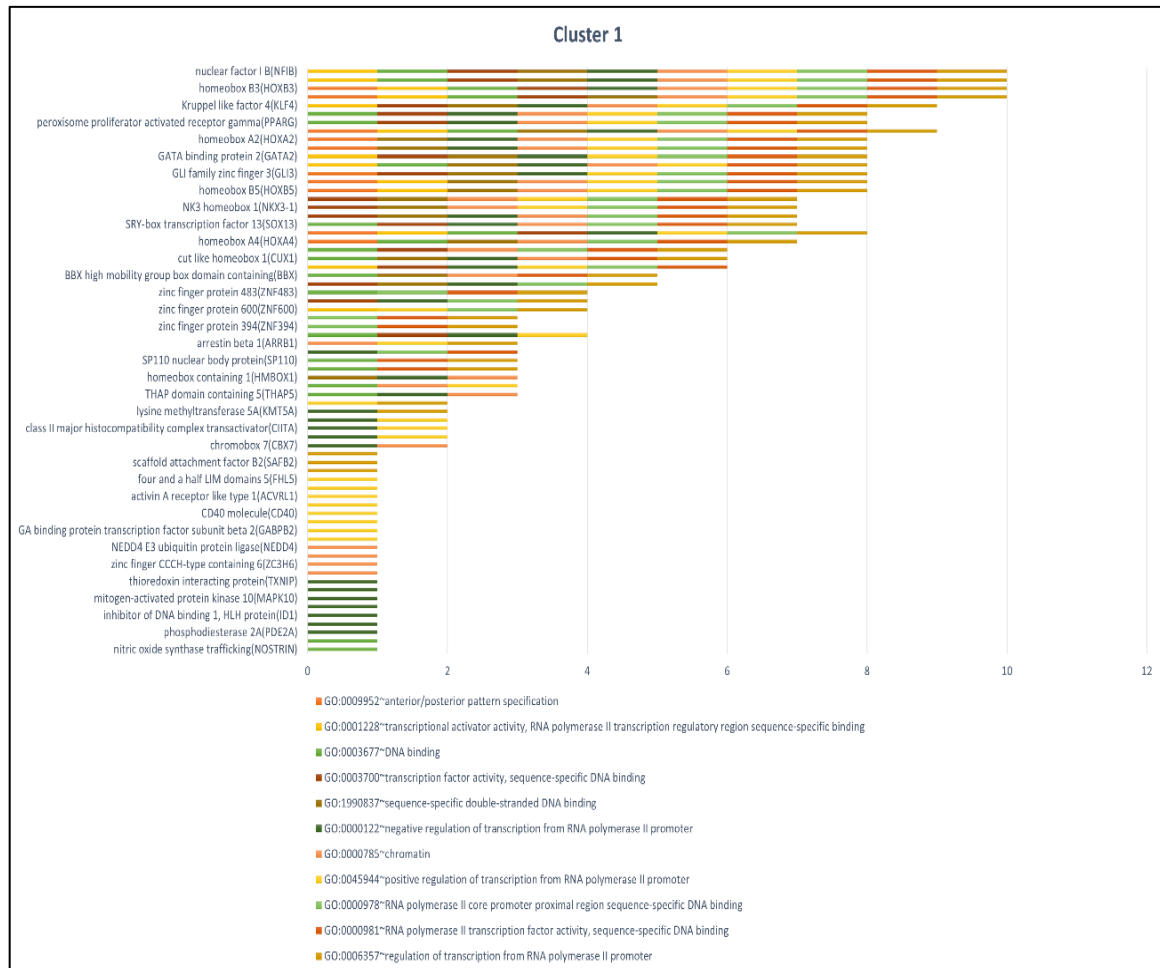


Figura 21 Clúster 1 genes down-regulated. se observa la categorización en dieferentes factores de transcripción y unión de ADN.

El resultado presentado demuestra que los DEGS down-regulated participan en la regulación y en la transcripción de los promotores de ARN y la unión a ADN. A pesar de participar en las tres funciones ontológicas, todas están ligadas a los procesos principales para la generación de proteínas desde el ADN.

El KEEG pathway de los genes down-regulated se encuentra conformado por el 45.7% del total de genes regulados aguas abajo. Se observa que las vías asociadas a estos genes se encuentran en relación con el metabolismo:

degradación de ácidos grasos, del piruvato, glucólisis y gluconeogénesis, y resistencia a la insulina, las vías metabólicas de los fármacos y drogas; vías metabólicas como activación de plaquetas, señalización de oxitocina, contracción del músculo liso y enfermedades degenerativas como Alzheimer y Parkinson.

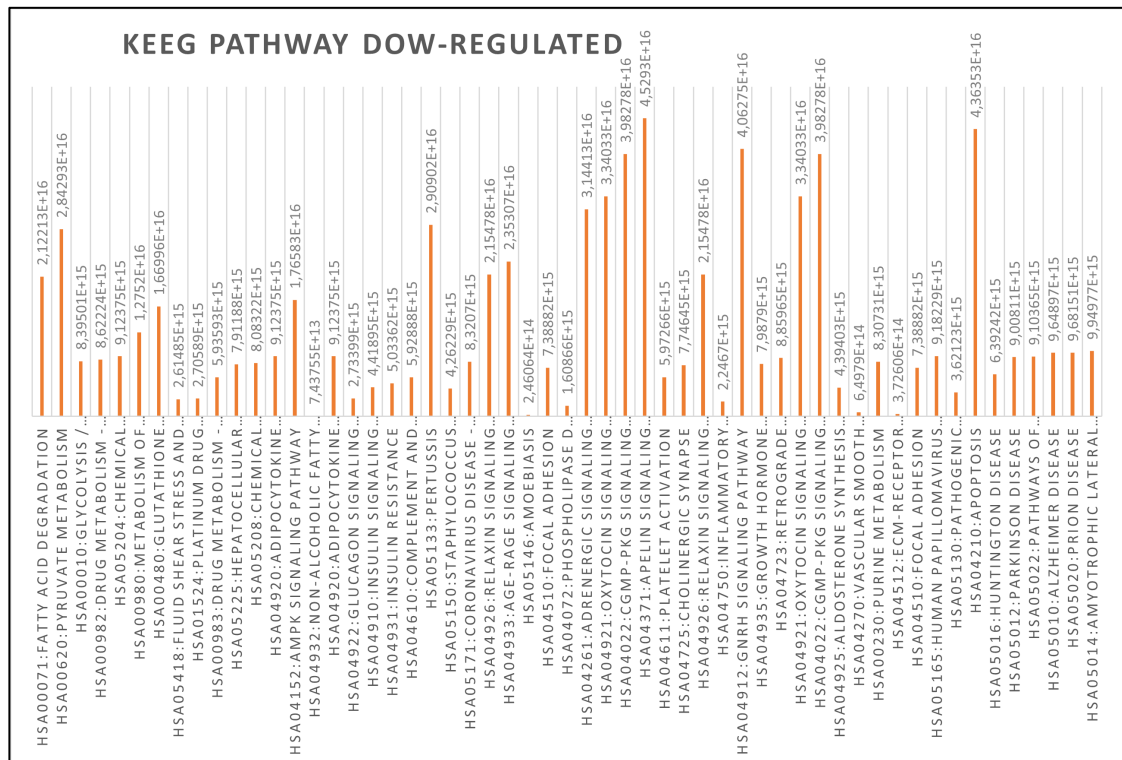


Figura 22 KEGG Pathway genes down-regulated. Las vías asociadas están en relación con el metabolismo como degradación de ácidos grasos, glucólisis y gluconeogénesis.

4.4 Construcción de red PPI y selección de módulo hub

En la construcción de la interacción proteína-proteína se observa un total de 109 nodos y 475 tipos de conexiones entre esos nodos. En la figura 23 se observa en morado, los nodos con mayor cantidad de interacciones dentro de su grupo de asociación. Estos son los seleccionados para el análisis global de todos los DEGS para la determinación del módulo hub. En la red proteína-proteína de los genes down-regulated se observan 43 nodos y 75 aristas de conexiones, en

café se observan los genes que son mayormente significativos por la cantidad de interacciones que presentan.

Por último, en la figura 25, se observa un global de 20 nodos y 52 aristas o conexiones entre los cuales se encuentran los genes mayormente conectados entre la red de los genes regulados aguas arriba y los genes regulados aguas abajo. Se presenta la interacción entre ellos y como se generan tres grupos de nodos que se relacionan entre ellos.

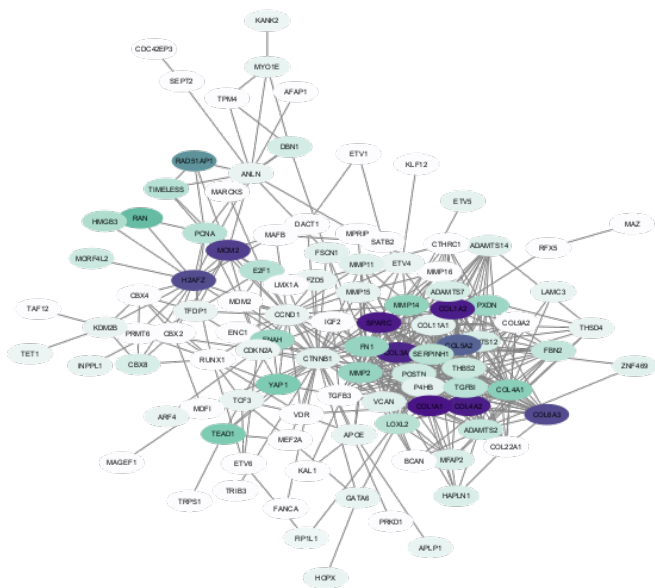


Figura 23 Red PPI up-regulated. En morado se observan los genes con mayor interacción en los módulos.

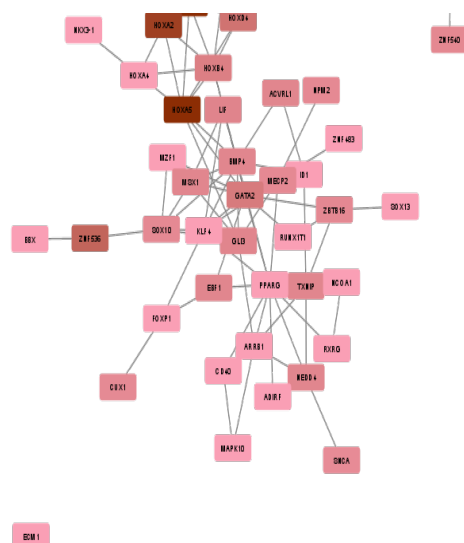


Figura 24 Red PPI down-regulated. En café se observan los genes con mayor interacción en los módulos. Se observan los genes HOX como mayormente asociados.

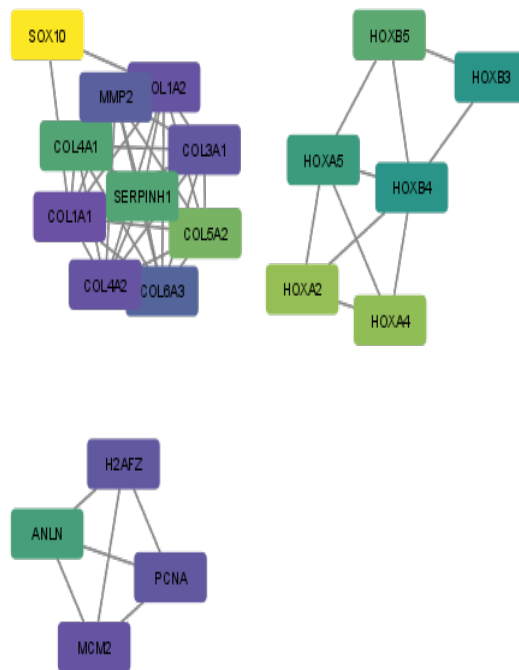


Figura 25 Red PPI final módulo g \acute{e} nicos seleccionados. Se puede observar la interacci \acute{o} n entre los genes determinados previamente.

Capítulo V - DISCUSIÓN

5. 1 Hallazgos

Los leiomiomas uterinos representan una de las enfermedades ginecológicas con mayor prevalencia en las mujeres en edad fértil. Con una prevalencia del 70% esta enfermedad afecta de gran manera a la calidad de vida de las mujeres. Su fisiopatología no se encuentra bien definida ni estudiada, por lo tanto, se vuelve necesario el entendimiento de los genes expresados diferencialmente y su función biológica para el entendimiento de la enfermedad. Los RNA seq es un nuevo tipo de secuenciación el cual busca determinar patrones de expresión de los genes. Se enfoca en conocer y estudiar el RNA mensajero que traduce y produce una proteína. Los datos fueron extraídos de 2 series de datos: GSE169255 y GSE199848.

En la figura 8 y 9 se observan los gráficos de cajas. En los gráficos a) de todas las figuras se puede observar como el rango el valor más bajo y el valor más alto no se encuentran en el mismo rango hacia la mediana, si bien esto no resulta un problema. La mediana de la caja se observa muy por debajo o por encima de la mediana calculada para toda la base de datos. Si se realizase el estudio con estos datos, se presentarían varios falsos positivos puesto que no existe una buena comparación entre observaciones y la mediana. Por ello se aplica la normalización de datos, en la imagen B) de todos los datos se observa una distribución normalizada de los conteos en la cual se ha aplicado el conteo por millón a cada muestra, el logaritmo del conteo por millón que permite modelizar la cantidad de conteos para todos los tamaños, en este caso se observa como todas las muestras tiene sus medianas cercanas al valor central de la mediana de la base de datos. Los datos que se utilizaron para continuar con el análisis son los datos normalizados, ya que permite contar con bases estandarizadas.

El análisis de los mapas de calor demuestra que existe gran variabilidad génica cuando se trata de muestras de tejido fibroso. Tanto en la figura 10 como 11 se puede observar que la alteración se dirige mayormente hacia genes regulados aguas arriba, lo que impulsa a deducir que la mayor cantidad de genes

alterados en la patología de los leiomiomas se dirigen hacia los genes up-regulated.

En la expresión diferencial (figuras 12 y 13) se observan dos tipos de gráficos: la tendencia de la media y el modelo final de la tendencia media. El primer gráfico catalogado nombrado como voom: demuestra la filtración de los genes con bajo nivel de expresión. Si bien la media sigue un patrón de hipérbola no estandarizado, los datos alrededor suyo no caen en desviaciones estándar que rompan el patrón de seguimiento a la parábola. Esto demuestra que la filtración de datos realizada en cada una de las bases fue efectiva lo que validó la diferenciación de los genes expresados diferencialmente. A su vez se determina una tendencia decreciente, este tipo de tendencias se asocia con experimentos con baja variación biológica, es decir, que la cantidad de genes expresados diferencialmente no llegan a ser de miles. En el gráfico de la derecha nombrado modelo final, se observa la estabilización de la media después de aplicar el método Bayesiano, donde cada punto negro representa un gen y la línea azul la tendencia de la media.

La identificación de DEGS comunes entre las dos series de datos presenta un total de 1039 genes, 444 gen correspondiente a Down-regulated y 595 genes correspondientes a genes up-regulated. Debido a que esta investigación plantea el reconocimiento de posibles biomarcadores y dianas terapéuticas se toman como resultados principales los genes compartidos por las dos muestras. Estos genes son los genes que tienen mayor representación génica en un total de 163 muestras de tejido fibroso y tejido normal.

5.2 Interpretación

En cuanto a la ontología génica se determina la importancia de los genes expresados diferencialmente en tres campos: componente celular, proceso biológico y morfología celular. En este estudio se determina que más del 90% de los genes up-regulated participan activamente en los tres campos de formación celular con mayor representación en los componentes celular, principalmente en la formación, estructuración y organización de la matriz extracelular. Esto puede ser comparado con el estudio hecho por (Paul, y otros, 2021) en el cual

determinan que en su análisis los genes regulados aguas arriba se presentan principalmente en la organización de la matriz y la estructura extracelulares. Los genes principales asociados a la formación de la matriz extracelular son los genes de la familia COL, asociada a la producción y generación de las cadenas de colágeno. (González, 2014) (Du & Taylor, 2016)

Otros genes importantes regulados aguas arriba son TGFBI (factor transformador de crecimiento beta inducido) y SERPINH1. SERPINH1 toma relevancia en este estudio ya que se encuentra asociado con la biosíntesis de colágeno, se ha considera un posible biomarcador para cáncer y nacimiento prematuro (NCBI National Center for Biotechnology Information, 2022). En el caso de TGFBI actúa en la unión de los colágenos tipo I, II, IV. Actuando principalmente en las interacciones colágeno-célula y como ligando para integrinas; importante para el control de sistema inmune (Gene Cards Suite, 2008). Por último, se observa la participación de genes para la unión y creación del citoesqueleto de actina de la célula, se expresa en el endometrio. (González, 2014) (Carbajo, y otros, 2022) (Du & Taylor, 2016)

Con respecto a los genes regulados a la baja se observa que su principal medio de acción es la regulación y transcripción de los promotores de ARN, por lo tanto, en la generación de proteínas desde el ADN. Resalta en este caso que se presenta la familia de genes HOX. Los genes HOX cumplen con varias funciones entre ellas el desarrollo embrionario, la angiogénesis, la proliferación y diferenciación celular. a su vez, estos se han visto ligados al desarrollo de diferentes tipos de canceres, especialmente en leucemia y cáncer cervicouterino. cuando existe una variación en la regulación, específicamente al alza o una expresión de nuevo, estos genes participan en la oncogénesis y progresión de esta (López, y otros, 2015). Esto podría explicar porque los leiomiomas no progresan a cáncer, ya que en la patogénesis del cáncer cervicouterino la familia de los genes HOX se regulan diferencialmente aguas arriba. Dos genes más determinados como importantes para la transcripción que son down-regulated son: NFIB y KLF4. En el caso de KLF4, se cree que interviene en la proliferación celular al actuar en el gen supresor de tumores P53 (Gene Cards Suite, 2015) (González, 2014) (Carbajo, y otros, 2022) (Du & Taylor, 2016).

En cuanto a las vías clave consideradas en los genes up-regulated y down-regulated. Se contrasta que los genes up-regulated sus vías clave se orientan más hacia el tumor génesis, al estar presentes en las vías de varios tipos de cáncer, también de vías de crecimiento celular como las vías asociadas a la infección por el virus del papiloma humano (HSA05165) y la vía de endocitosis y de vasodilatación por medio de la vía de la relaxina. Por otro lado, los genes down-regulated se encuentran asociadas con vías mayormente metabólicas, con mayor prevalencia la vía de apoptosis. Se entiende que al haber una regulación a la baja de la vía apoptosis, este proceso se inhibe dentro de las células lo que da el crecimiento de los tumores benignos presentes en leiomiomatosis uterina (Shaw, 2011). También se presenta una regulación a la baja de otros procesos metabólicos como la vía de la apelina que es una de las vías para el vaso relajación, así como de la vía de la oxitocina (Creative Diagnostics., 2015). Por último, otra de las vías a recalcar es la vía de señalización de GNRH, que fisiológicamente regula la producción y liberación de FSH y LH, lo cual explica la metrorragia que presentan las pacientes y los dolores pélvicos no cíclicos (KEGG: Kyoto Encyclopedia of Genes and Genomes, 2009) (González, 2014).

En las redes de interacción proteína-proteína adquieren relevancia los genes que se coexpresan con mayor regularidad y por lo tanto tienen mayor interacción entre sí para la formación de leiomiomas, en el caso de esta investigación se presentan tres tipos de redes, que a pesar de presentarse como aisladas forman parte importante para el entendimiento de la fisiopatología. La presencia de la familia COL y su interacción con SERPHIN1 demostraría cómo se da la formación de los tumores al aumentar la matriz extracelular de las células y por lo tanto la unión entre ellas (Vargas, Rodríguez, & Acosta, 2013). Otra de las redes de interacción es de los genes HOX que como ya se ha mencionado antes, estas al ser reguladas a la baja no permiten la progresión de los tumores benignos hacia malignidad. Por último, la regulación aguas arriba de los genes H3AF2, PCNA Y MCM2 que actúan en el aumento de la síntesis de la cadena principal del ADN (González, 2014) (Carbajo, y otros, 2022).

Se observa en el análisis final de las redes 3 tipos de redes con el mismo nivel de co-expresión pero no se encuentran asociados. En gran manera esto puede estar representado por los subtipos de leiomiomas que ya han sido determinados en estudios previos externos.

Capítulo VI - CONCLUSIONES Y LIMITACIONES

6.1 Conclusiones

- El mínimo de muestras necesitadas para el entendimiento y realización del estudio son 15 variables.
- Uno de los aspectos relevantes de este estudio es la necesidad de tener muestras de miometrio sano y leiomiomas uterinos que no sean de la misma paciente, ya que permite tener una apreciación más amplia de los genes que pueden estar involucrados en un contexto global.
- Se determinó la presencia de 444 genes regulados a la baja y 595 regulados al alza. Cerca de la mitad de los genes expresados diferencialmente en los dos extremos de la regulación intervienen en la producción de leiomiomas uterinos, por lo tanto, se necesita de una homeostasis entre estos dos patrones para la generación de la enfermedad.
- Los genes up-regulated se encuentran asociados con el crecimiento y la proliferación de las células para la creación de tumores benignos, a través de la creación de citoesqueleto y aumento de la replicación celular.
- La regulación de la familia HOX explica porque estos tumores no progresan a malignidad. Para la generación de cáncer cervicouterino la familia HOX debe estar regulada al alza y producir una supresión de los genes oncosupresores. Al ser regulados a la baja los genes oncosupresores se mantienen en un estado basal, impidiendo que los oncogenes tengan un desarrollo y transcripción no controlada.
- La actuación de las vías metabólicas reguladas a la baja también puede explicar la presencia de ciertos síntomas que se producen en los leiomiomas uterinos.
- Las vías metabólicas asociadas con glucólisis, gluconeogénesis y de la señalización de la insulina dan indicios que hay sintomatología y signos en mujeres con leiomiomas que no han sido estudiados, no han tomado relevancia

o que se han estudiado de manera externa y no como parte de la enfermedad en sí, para ello serviría realizar un nuevo estudio en el cual se investigue el cuadro clínico general en mujeres con leiomiomas uterinos.

- Los genes expresados al alza la mayoría este compuesto por la familia COL. Se pueden determinar dos conclusiones a partir de este análisis. En primer lugar, se determina que la proliferación de células para la producción de tumores benignos está dada por el aumento de la matriz extracelular generando un medio adecuado para esta función. Otra posible conclusión está ligada al conocimiento previo de 4 subtipos de leiomiomas, uno de los cuales es conocido como COL, debido a su sobreexpresión de la familia COL. Por lo tanto, se podría determinar que una parte de las muestras de utilizadas en este estudio son de subtipo COL no diagnosticado. También se podría determinar que el subtipo COL es más prevalente de lo que se ha estudiado en investigaciones con universos reducidos. Esto anima a la producción de más investigaciones que pueden ser realizadas en la mujer ecuatoriana con esta patología.

- Las vías de interacción proteína-proteína a pesar de no relacionarse tienen un mismo nivel de importancia que puede ser estudiado a futuro, para la determinación de nuevas vías fisiopatológicas.

- Los genes de la familia COL, SERPINH1 y la familia HOX son estudiados como biomarcadores para otras enfermedades principalmente para ciertos tipos de cáncer. Una orientación de la medicina personalizada a esta área puede permitir usar estos genes como biomarcadores para el diagnóstico temprano de leiomioma uterino. También al estar influenciadas en vías estructurales o de metabolización se vuelven dianas terapéuticas para posibles estudios a futuro.

6.2 Limitaciones

Las limitaciones de esta investigación se basan principalmente en el tamaño muestral, ya que mínimamente es necesario tener 15 para poder obtener resultados que puedan ser valorables. Esto pudo ser determinado ya que al usar un dataset conocido como GSE173334 que contaba únicamente con 9 muestras,

los genes expresados diferencialmente al ser comparados entre los 3 datasets, se reducían a menos de 100 genes. Dos posibles razones para esto, es la poca información que se obtiene de pocas muestras que no representan un universo muestral y por lo tanto el número de observaciones se limita a la comparación dentro del dataset. A continuación, se presentan los gráficos determinados de este dataset, los mismos que no fueron incluidos en el capítulo de resultados al ser eliminados del proceso.

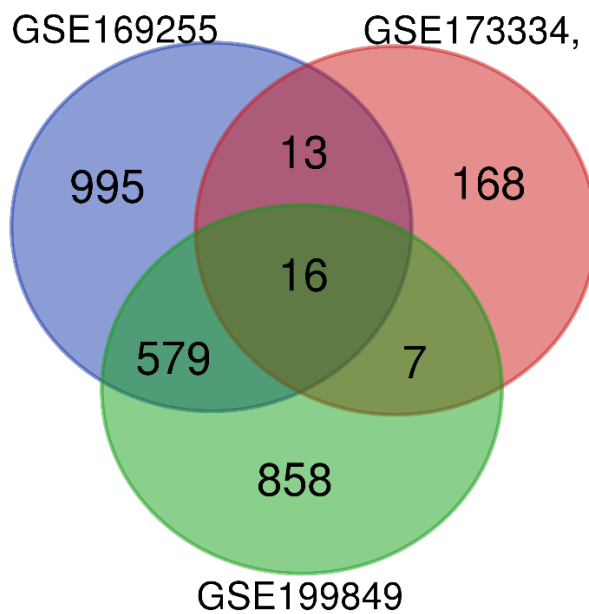


Figura 6. 1 Determinación de genes up-regulated en tres datas. Tomando en cuenta el dataset GSE173334 que cuenta con 9 muestras de pacientes

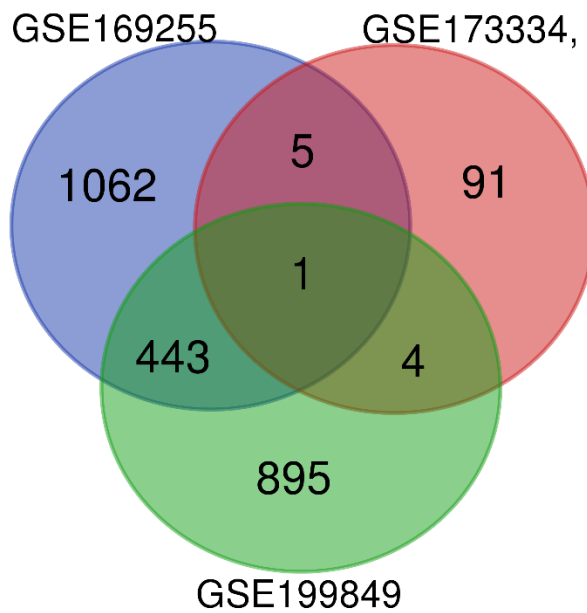


Figura 6. 2 Determinación de genes down-regulated en tres datas. Tomando en cuenta el dataset GSE173334 que cuenta con 9 muestras de pacientes

La segunda razón puede estar determinada al procesamiento de datos del estudio original al cual pertenecen. Esto no es solo una limitación en el caso de este dataset, sino de todo el estudio. Por ello, se normalizo las muestras tomando en cuenta el tipo de estudio original realizado en el cual se escogió aquellos que tengan expresión de RNA. Si bien esto dejo muchas muestras que podrían tener resultados similares a los estudios realizados previamente. Permitted a la determinación de nuevos genes, poco conocidos o nombrados por otros autores al no ser un enfoque de estudio. Por ello, esta investigación trato desde un inicio no verse influenciada por los subtipos de leiomiomas, por lo cual se escogieron datasets en los cuales no se especificaba el subtipo génico de las muestras.

6.3 Implicaciones de salud pública

A pesar de este ser un estudio que propone beneficiar a la comunidad científica, no se puede dejar de lado la salud en el Ecuador. Por ello, si bien estos datos no pueden ser extrapolados de manera certera a la realidad ecuatoriana, permiten poner un punto de inicio para estudios de interés sobre leiomiomas uterinos en el país. Los leiomiomas uterinos es una de las enfermedades en el

Ecuador con la que se cuenta con pocos estudios de prevalencia e incidencia. A partir, de este estudio se quiere demostrar que en el Ecuador existe capacidad tecnológica a nivel de análisis computacional y estadístico para la realización de este tipo de estudios, y más que nada se pretende impulsar a los desarrolladores de estudios genéticos, biológicos moleculares y bioinformáticos del país a tomar en consideración el estudio de los subtipos de leiomiomas uterinos prevalentes en el Ecuador y poder generar de esta manera una medicina personalizada para la mujer que lo necesite. Por otro lado, tomando en cuenta las conclusiones planteadas en el primer acápite de este capítulo, el trabajo en conjunto con médicos especialistas en ginecología para estudiar el tiempo de progresión de la enfermedad de intereses y un estudio más ampliado para determinar síntomas y signos que han sido infradiagnosticados.

Bibliografía

- Miriam Martini, M. C. (2014). PI3K/AKT signaling pathway and cancer: an updated review. *Annals Of Medicine*, 372-383. doi:10.3109/07853890.2014.912836
- Alba Machado-Lopez, C. S. (2021). Molecular and Cellular Insights into the Development of Uterine Fibroids. *International Journal of Molecular Sciences*, 8483. doi:10.3390/ijms22168483
- Phelippeau, J., & Fernandez, H. (2016). Fibromas Uterinos. *TRATADOS EMC: Ginecología - Obstetricia*, 52, 1-6.
- Ministerio de Salud Pública. (2013). *Encuesta nacional de salud y nutrición ENSANUT*. Quito, Ecuador. Obtenido de https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/ENSANUT/MSP_ENSANUT-ECU_06-10-2014.pdf
- Finkelstein, J., Guitron Leal, C., Chu, W., Krisher, J., Haas, J., Mehta, S., & Freire, W. (2019). Anemia and Iron, Vitamin B12, and Folate Deficiencies in Women of Reproductive Age in Ecuador: Results from the Ecuadorian National Health and Nutrition Survey (P10-027-19). *Current Developments in Nutrition*, 3.
- Stewart, E., Cookson, C., Gandolfo, R., & Schulze-Rat, R. (2017). Epidemiology of uterine fibroids: a systematic review. *BJOG : an international journal of obstetrics and gynaecology*, 124(10), 1501-1512. doi:<https://doi.org/10.1111/1471-0528.14640>
- Laganà, A., Vergara, D., Favilli, A., La Rosa, V. L., Tinelli, A., Gerli, S., . . . Vitale, S. G. (2017). Epigenetic and genetic landscape of uterine leiomyomas: a current view over a common gynecological disease. *Archives of gynecology and obstetrics*, 196(5), 855-867. doi:10.1007/s00404-017-4515-5.

- Chabbert-Buffet, N., Esber, N., & Bouchard, P. (2014). Fibroid growth and medical options for treatment. *Fertility and sterility*, *102*(3), 630-639. doi:10.1016/j.fertnstert.2014.07.1238
- Mehine, M., Kaasinen, E., Heinonen, H.-R., Makinen, N., Kampjarvi, K., Savilinna, N., . . . Aaltonen, L. (2016). Integrated data analysis reveals uterine leiomyoma subtypes with distinct driver pathways and biomarkers. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), 1315-1320. doi:10.1073/pnas.1518752113
- Mehine, M., Makinen, N., Heinonen, H.-R., Aaltonen, L., & Vatheristo, P. (2014). Genomics of uterine leiomyomas: insights from high-throughput sequencing. *Fertility and sterility*, *120*(3), 621-629. doi:10.1016/j.fertnstert.2014.06.050
- Mantilla, C., Suarez, I., Duque, A., & Navas, M. C. (2015). Mecanismos de señalización por β -catenina y su papel en la carcinogénesis. *CES Medicina*, *29*(1), 109-127.
- Zhang , S., Mo, Q., & Wang, X. (2019). Oncological role of HMGA2 (Review). *International journal of oncology*, *55*(4), 775-788. doi:10.3892/ijo.2019.4856
- Valencia, C., & Arteaga, C. (2017). Síndrome de leiomiomatosis hereditaria y cáncer de células renales: revisión de la literatura. *Urología Colombiana*, *26*(2), 126-134. doi:10.1016/j.uroco.2017.03.013
- Wang, X., Li, W., Wei, K., Xiao, R., Wang, J., Ma, H., . . . Li, C. (2018). Missense mutations in COL4A5 or COL4A6 genes may cause cerebrovascular fibromuscular dysplasia: Case report and literature review. *Medicine*, *97*(30). doi:10.1097/MD.00000000000011538.
- Wise, L., & Laughlin-Tommaso, S. (2016). Epidemiology of Uterine Fibroids: From Menarche to Menopause. *Clinical obstetrics and gynecology*, *59*(1), 2-24. doi:10.1097/GRF.0000000000000164

- Laughlin-Tommaso, S. (2016). Alternatives to Hysterectomy: Management of Uterine Fibroids. *Obstetrics and gynecology clinics of North America*, 43(3), 397- 413. doi:10.1016/j.ogc.2016.04.001
- Bosquet, E. G. (2020). Embriología y anatomía del aparato genital femenino. En *González-Merlo. Ginecología* (págs. 1-19). España: Elsevier.
- Pöder, L. (2018). Evaluación ecográfica del útero . En *Callen. Ecografía en obstetricia y ginecología* (págs. 846-882). España: Elsevier.
- Erica, E. M., & Ayman, A.-H. (2018). Burden, Prevalence, and Treatment of Uterine Fibroids: A Survey of U.S. Women. *Journal of Womens Health*, 27(11), 1359–1367. doi:10.1089/jwh.2018.7076
- Mara, U., & Mohamed Ali, B. (2020). Uterine Fibroids in Menopause and Perimenopause. *Menopause*, 238-242. doi:10.1097/GME.0000000000001438
- Giuliani, E., As-Sanie, S., & Marsh, E. (2020). Epidemiology and management of uterine fibroids. *International Journal of Gynecology & Obstetrics.*, 3-9. doi:10.1002/ijgo.13102
- Abdullah, R., Massey, I., & Liu, N. (2021). The differences in characteristics of uterine leiomyomas and the diverse adverse. *JOURNAL OF OBSTETRICS AND GYNAECOLOGY*, 41(6), 841–847. doi:10.1080/01443615.2020.1846020
- Lewis, T., Malik, M., Britten, J., San Pablo, A., & Catherino, W. (2018). A Comprehensive Review of the Pharmacologic Management of Uterine Leiomyoma. *BioMed Research International*, 11. doi:10.1155/2018/2414609
- Jacques Donnez, Pablo Arriagada, Olivier Donnez & Marie-Madeleine. (2018). Emerging treatment options for uterine fibroids. *Expert Opinion on Emerging Drugs*, 23(1), 17-23. doi:10.1080/14728214.2018.1446943

- Morocho, A., Espinoza, C., & Villarruel, A. (2019). Nuevos biomarcadores en la evaluación del riesgo cardiovascular. *Revista Latinoamericana de Hipertensión*, 712-716.
- Sacristan, J. A. (2018). Medicina centrada en el paciente, medicina personalizada y medicina de precisión: ¿de qué estamos hablando? *Medicina Centrada en el Paciente*, 35-41. Obtenido de https://www.researchgate.net/profile/Jose-Sacristan/publication/326186622_Medicina_centrada_en_el_paciente_medicina_personalizada_y_medicina_de_precision_De_que_estamos_hablando/links/5b3ccd730f7e9b0df5edb03d/Medicina-centrada-en-el-paciente-medicina-pers
- Alonso, M., Laserna, E., Varo, G., Molina, M., & Orera, M. (2018). El papel del laboratorio clínico en la medicina personalizada: situación actual y retos futuros. *Revista del Laboratorio Clínico*, 202-208. doi:<https://doi.org/10.1016/j.labcli.2017.11.006>
- Chen, Y., Lun, A. T., & Smyth, G. K. (2022). From reads to genes to pathways: differential. *F1000Research*, 1-51. doi:<https://doi.org/10.12688/f1000research.8987.2>
- Ayala, G. (2022). Expresión diferencial con datos RNASeq. En G. Ayala, *Bioinformática Estadística Análisis estadístico de datos ómicos* (págs. 141-164). Valencia: Universidad de Valencia.
- Rodrigo, J. A. (2017). *Clustering y heatmaps: aprendizaje no supervisado*. Attribution 4.0 International. Obtenido de https://rpubs.com/Joaquin_AR/310338#:~:text=La%20distancia%20eucl%C3%ADdea%20entre%20las,cada%20una%20de%20las%20variables
- Anders, S. H. (2010). Differential expression analysis for sequence count data. *Genome Biology*.

- McCarthy, D. J. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25, 765-771. Obtenido de <http://bioinformatics.oxfordjournals.org/content/25/6/765>
- Páez, L. O. (2011). Descripción general de la Inferencia Bayesiana y sus aplicaciones en los procesos de gestión. *Universidad del Rosario*, 1-28. Obtenido de https://www.urosario.edu.co/Administracion/documentos/investigacion/laboratorio/miller_2_2.pdf
- B.T. Sherman, M. H. (2021). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*. doi:10.1093/nar/gkac194
- Huang DW, S. B. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.*, 44-57.
- Paul, E., Burns, G., Carpenter, T., Grey, J., Fazleabas, A., & Teixeira, J. (2021). Transcriptome Analyses of Myometrium from Fibroid Patients Reveals Phenotypic Differences Compared to Non-Diseased Myometrium. *International Journal of Molecular Sciences*. doi:10.3390/ijms22073618
- González, I. M. (2014). *Ontología de Alus*. Universitat Oberta de Catalunya. Obtenido de http://openaccess.uoc.edu/webapps/o2/bitstream/10609/30121/6/imallon_aTFC0114memoria.pdf
- Du, H., & Taylor, H. (2016). The Role of Hox Genes in Female Reproductive Tract Development, Adult Function, and Fertility. *Cold Spring Harbor Perspectives in Medicine*, 1-33. Obtenido de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4691806/>
- NCBI National Center for Biotechnology Information. (2022). *SERPINH1 serpin family H member 1*. (Gene ID: 871) Obtenido de [Homo sapiens (human)]: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=871>

- Gene Cards Suite. (2008). *TGFBI Gene - Transforming Growth Factor Beta Induced*. Obtenido de [Homo sapiens (human)]: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=TGFBI>
- Carbajo, M., Corachán, A., Juárez, E., Quiñotero, A., Pellicer, A., & Ferrero, H. (2022). Integrative analysis of the DNA methylome and transcriptome in uterine leiomyoma shows altered regulation of genes involved in metabolism, proliferation, extracellular matrix, and vesicles. *The Journal Of Pathology*. doi:10.1002/path.5920.
- López, R., Marrero, D., Romero, P., Villegas, V., Valdivia, A., Arreola, H., . . . Salcedo, M. (2015). El papel de los genes del desarrollo tipo HOX en el cáncer cervicouterino. *Rev Med Inst Mex Seguro Soc*, 188-193. Obtenido de <https://www.medigraphic.com/pdfs/imss/im-2015/ims152k.pdf>
- Gene Cards Suite. (2015). *KLF4 Gene - Kruppel Like Factor 4*. (NCBI Entrez Gene: 9314) Obtenido de [Homo sapiens (human)]: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=KLF4>
- Shaw, M. M. (2011). The AMP-activated protein kinase (AMPK) signaling pathway coordinates cell growth, autophagy, & metabolism. *Nature cell biology*, 1016-1023. Obtenido de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3249400/>
- Creative Diagnostics. (2015). *Apelin Signaling Pathway*. Obtenido de [Homo sapiens (human)]: <https://www.creative-diagnostics.com/apelin-signaling-pathway.htm>
- KEGG: Kyoto Encyclopedia of Genes and Genomes. (2009). *PATHWAY: hsa04912*. (ko04912) Obtenido de GnRH signaling pathway - Homo sapiens (human): <https://www.genome.jp/entry/hsa04912>
- Vargas, V., Rodríguez, J., & Acosta, G. (2013). Leiomiomatosis uterina. Aspectos epidemiológicos, fisiopatogénicos, reproductivos, clínicos y terapéuticos. *Rev Hosp Jua Mex*, 173-182. Obtenido de <https://www.medigraphic.com/pdfs/juarez/ju-2013/ju133e.pdf>

ANEXOS

Anexo 1

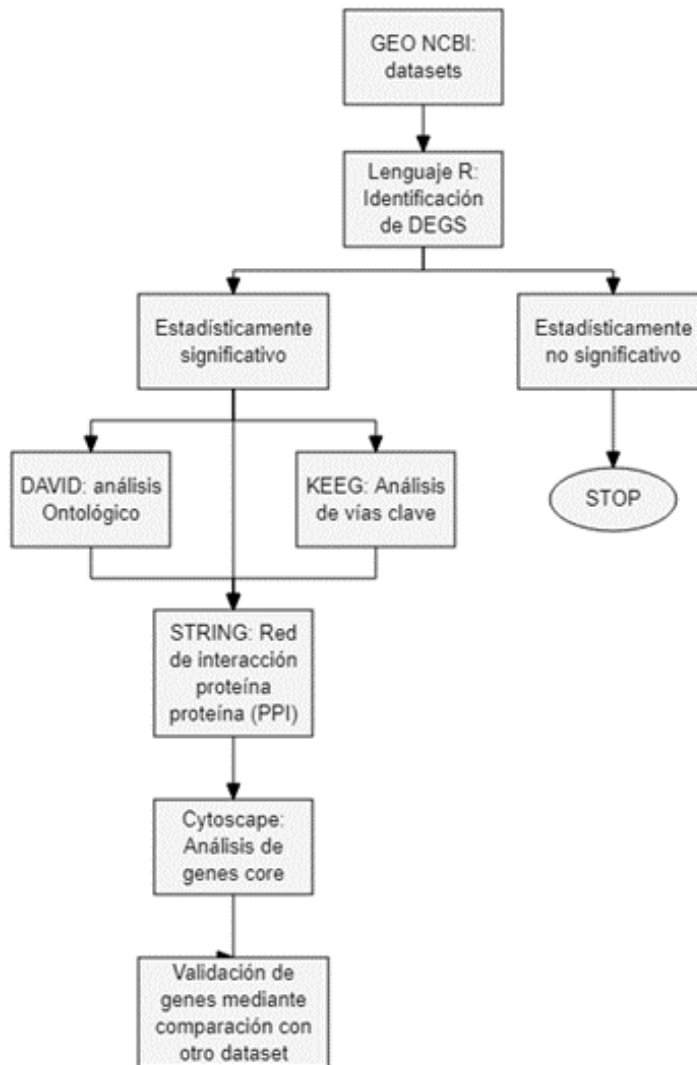


Figura 26 Flujograma de trabajo de investigación

Anexo 2

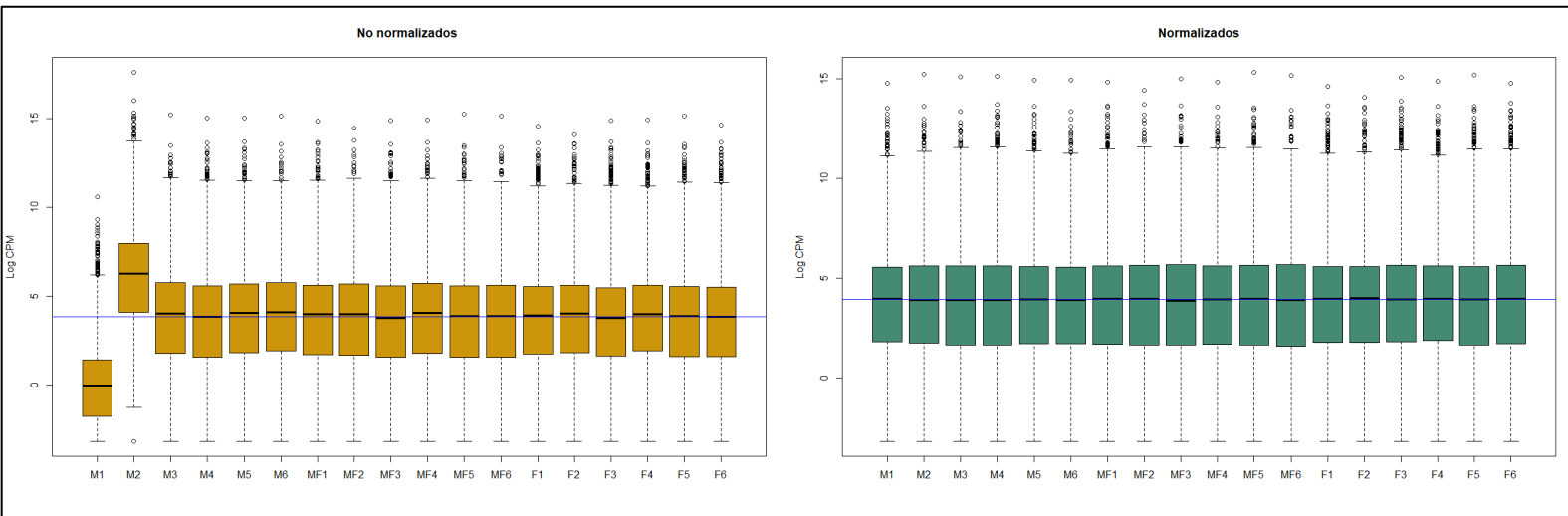


Figura 27 Diagrama de Cajas Base de datos GSE169255. A) presenta la distribución de log-CPM en los datos no normalizados B) se presenta la normalización de datos de cada muestra después de aplicar `calcNormFactor` por medio del método TMM

Anexo 3

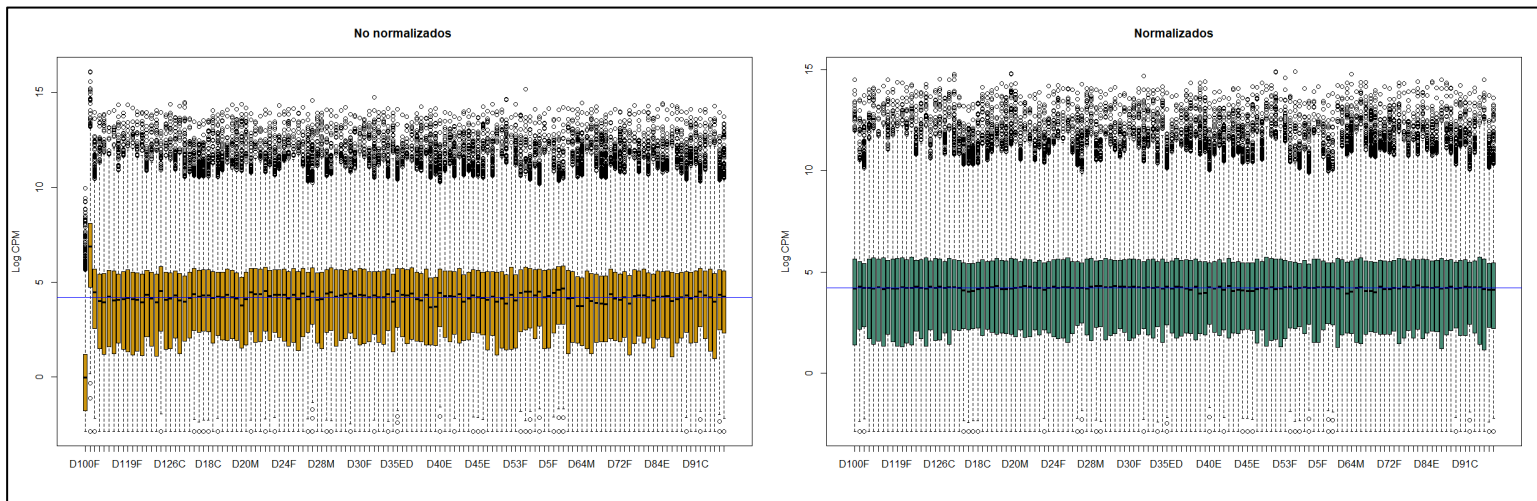


Figura 28 Diagrama de Cajas Base de datos GSE199849. A) presenta la distribución de log-CPM en los datos no normalizados B) se presenta la normalización de datos de cada muestra después de aplicar `calcNormFactor` por medio del método TMM

Anexo 4

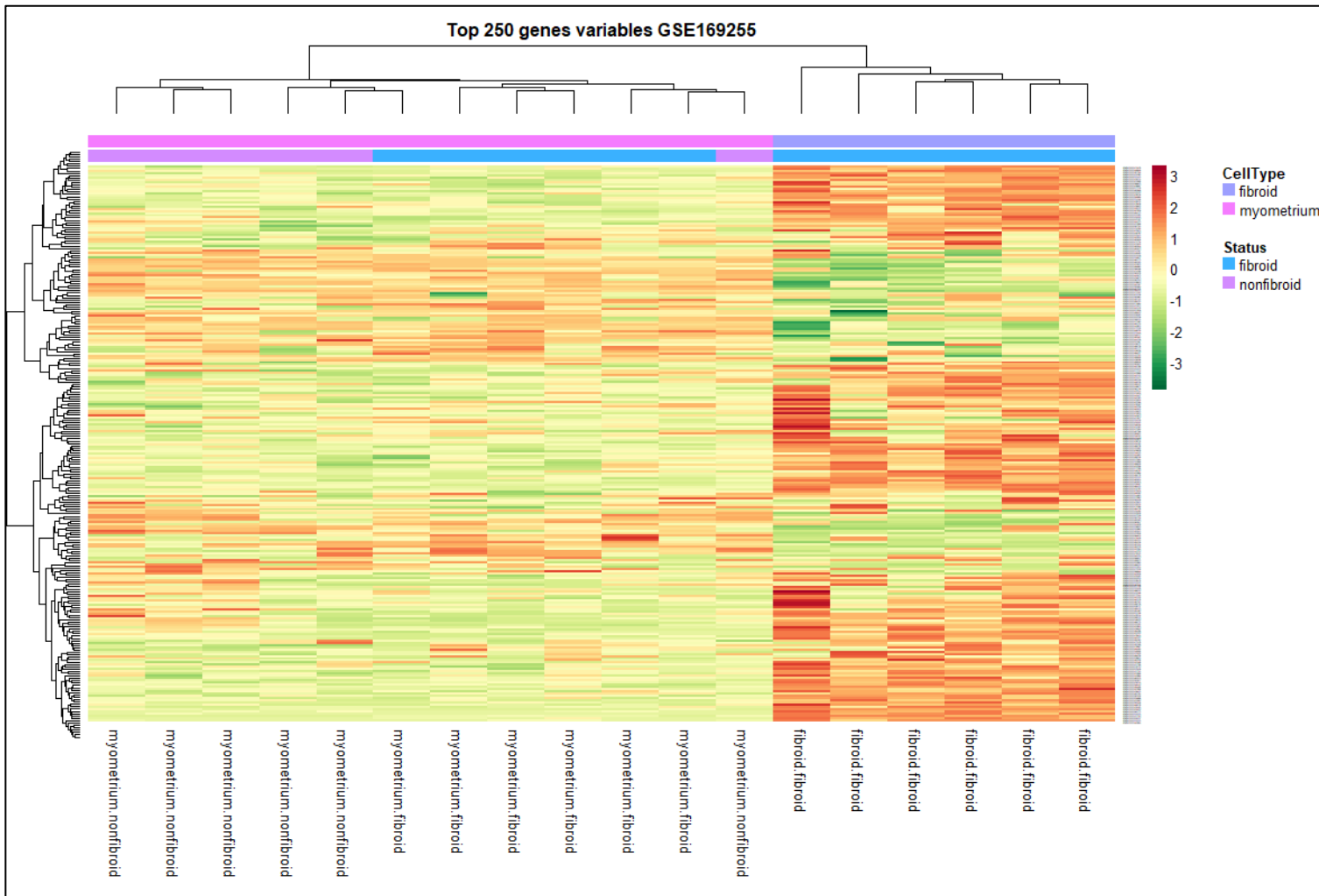


Figura 29 Mapa de calor de la base GSE169255. En rojo intenso se presentan los genes mayormente variables hacia up-regulated y en verde intenso los genes down-regulated

Anexo 5

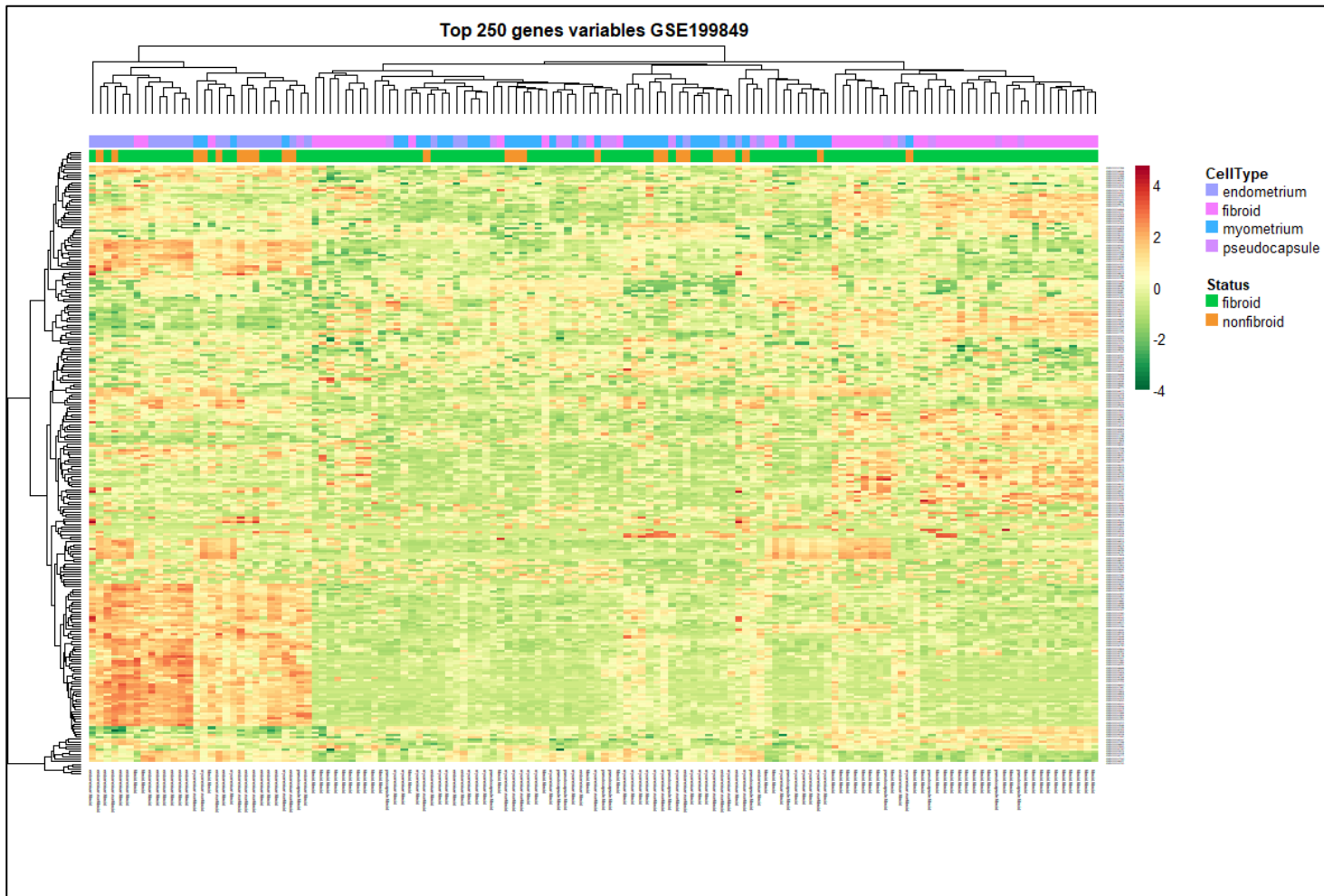


Figura 30 Mapa de calor de la base GSE199849. En rojo intenso se presentan los genes mayormente variables hacia up-regulated y en verde intenso los genes down-regulated

Anexo 6



Figura 31 Clúster 1 genes up-regulated. En verde los genes asociados a función matriz extracelular, amarillo organizacion de matriz extracelular y naranja estructura de la matriz extracelular.

Anexo 8

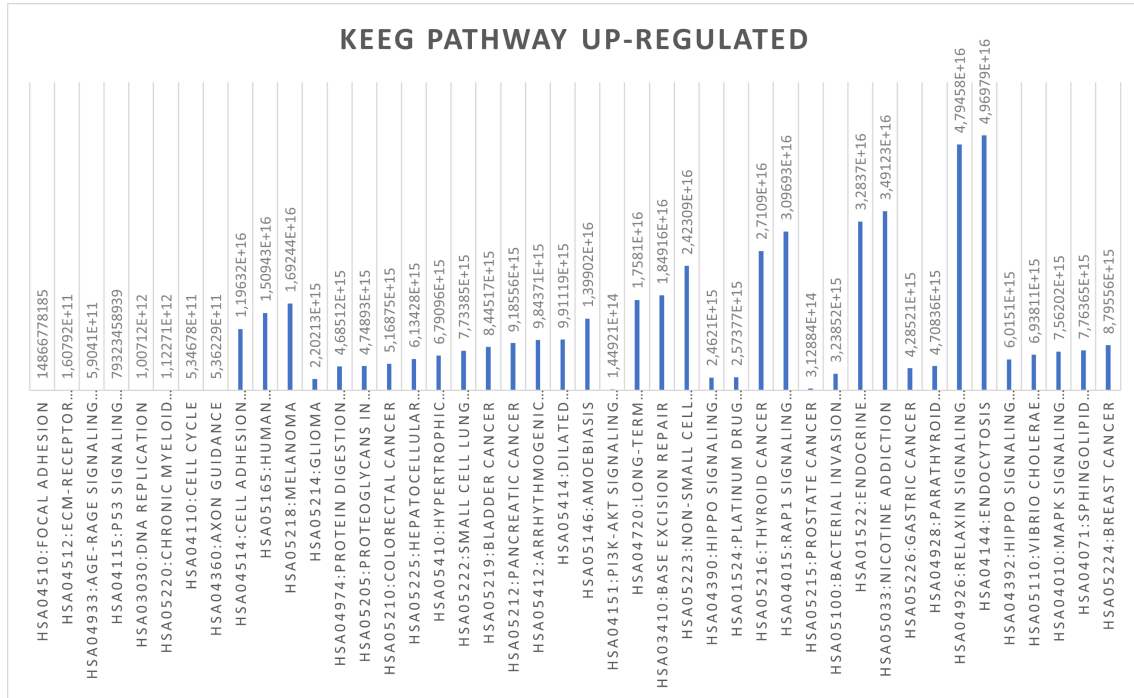


Figura 33 KEEG Pathway up-regulated. Intervienen vías asociadas replicación del ADN, división celular y generación de distintos cánceres.

Anexo 9



Figura 34 Clúster 1 genes dow-regulated. se observa la categorización en diefernetes factores de transcripción y unión de ADN.

Anexo 10

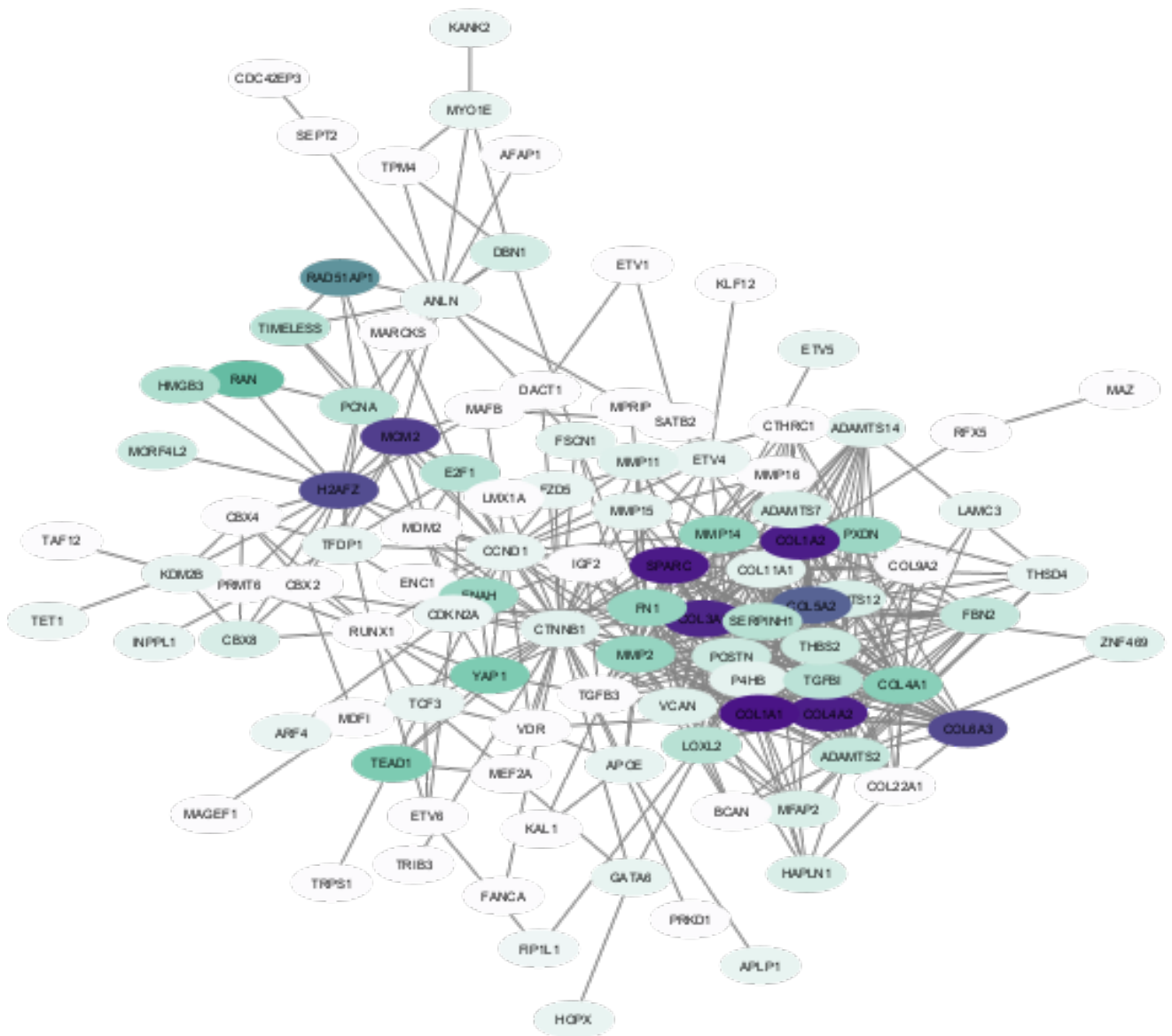


Figura 35 Red PPI up-regulated. En morado se observan los genes con mayor interacción en los módulos.

Anexo 11

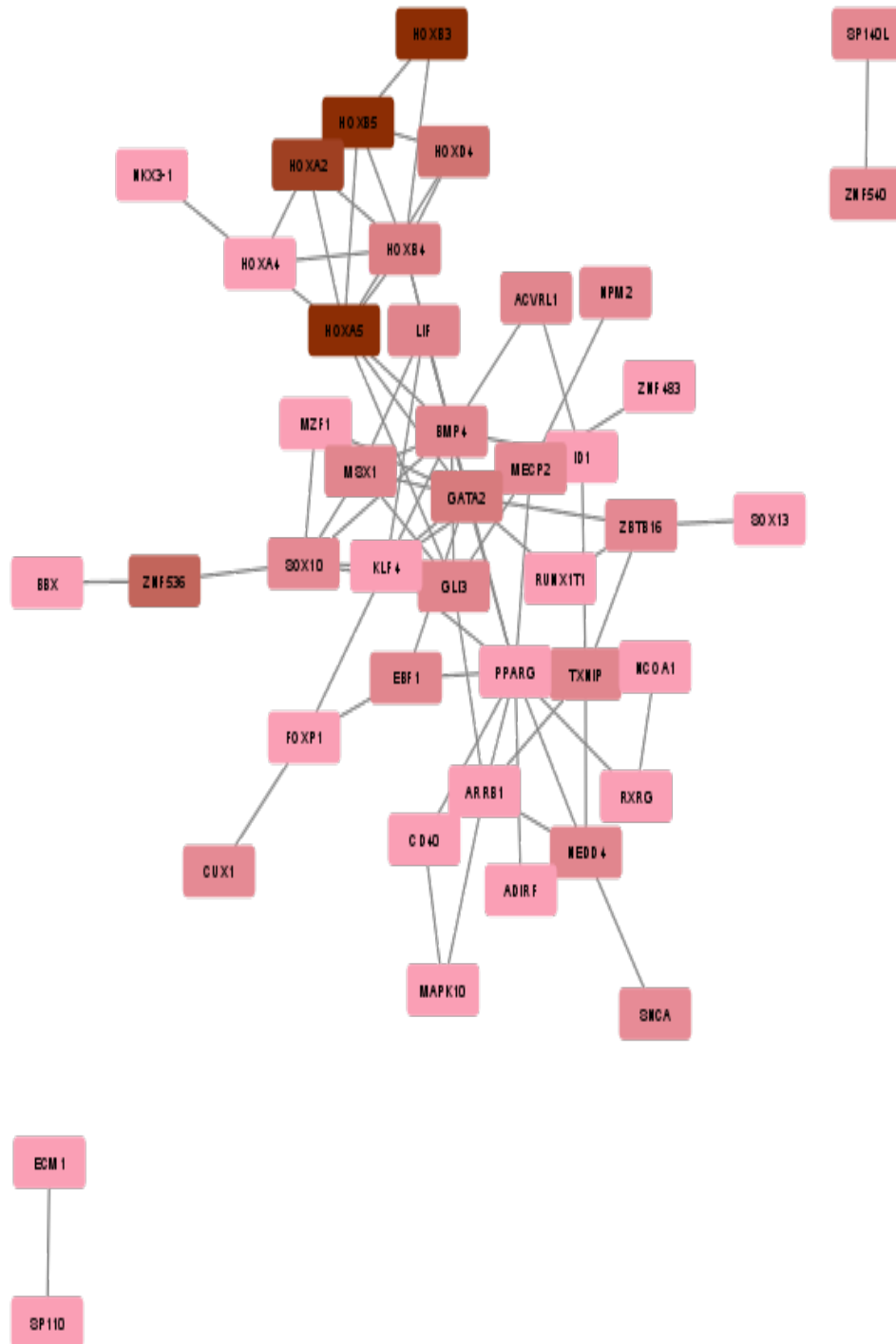


Figura 36 Red PPI down-regulated. En café se observan los genes con mayor interacción en los módulos. Se observa los genes HOX como mayormente asociados.

Anexo 12

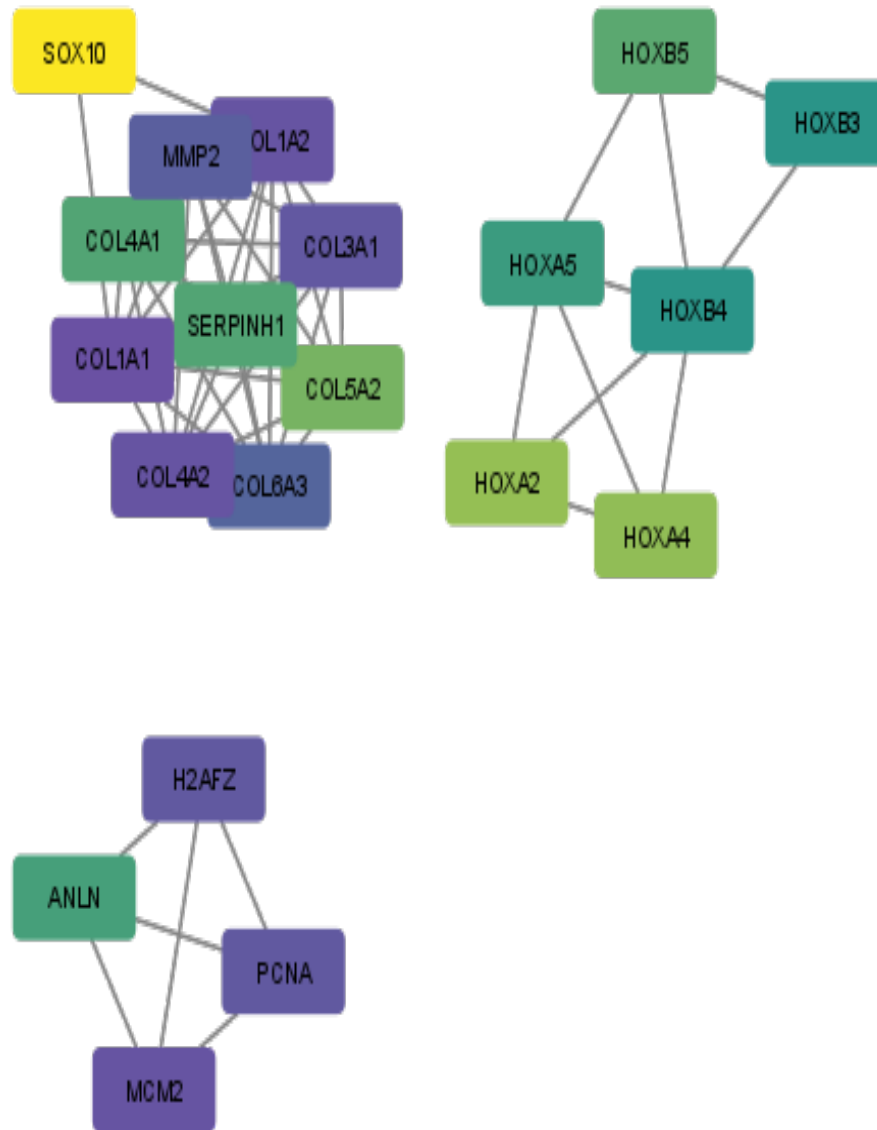


Figura 37 Red PPI final módulo g nicos seleccionados. Se puede observar la interacci3n entre los genes determinados previamente.

Tabla 3 Variables de investigación

Variable	Definición conceptual	Definición operacional	Categoría	Tipo	Indicador	Técnica o instrumentos
Perfiles de expresión génica	Medida de la actividad de expresión génica de miles de genes simultáneamente para crear una imagen global de la función celular	Módulos de secuencias o genéticas que se encuentren asociadas a fibrosis uterina	Cualitativa	Nominal	1 Presencia de secuencias génicas asociadas a fibrosis uterina 2 Ausencia de secuencias génicas asociadas a fibrosis uterina	T: análisis bioinformático
Genes expresados diferencialmente (DEGS)	Niveles de expresión de genes significativamente diferentes Fuerza impulsadora molecular y/o biomarcadores moleculares de diferentes fenotipos	Genes que tengan una mayor nivel de expresión génica respecto a fibrosis uterina	Cualitativa	Nominal	1 Estadísticamente significativo: $p < 0.5$ 2. No es estadísticamente significativo: $p > 0.5$	T: análisis bioinformático co I: análisis estadístico

Proteína	Clase importante de moléculas que se encuentran en las células vivas. Se compone de una o mas cadenas largas de aminoácidos cuya secuencia corresponde a la secuencia de ADN del gen que la codifica	Genes codificantes de proteínas que tienen relación con procesos biológicos y/o funcionalidad molecular de fibrosis uterina	Cualitativa	No nominal	1 Función molecular 2 Procesos biológicos 3 Componentes celulares	T: análisis bioinformático
Gen Hub	Gen que tiene muchas interacciones con otros genes, generalmente tiene un papel esencial en la regulación de genes y proceso	Genes con alta correlación en un módulo genético	Cualitativa	No nominal	Genes que presenten redes de conexión $P < 0.05$ 2 Genes que presenten redes de conexión $p > 0.05$	T: análisis estadístico, univariado y multivariado

	biológicos					
--	------------	--	--	--	--	--

