



Universidad Autónoma de Querétaro

Facultad de Ciencias Naturales



**Inferencia y comparación entre redes de regulación génica de células
provenientes de cáncer endometrial y otros tipos celulares coexistentes en un
tumor primario**

Tesis individual

Que como parte de los requisitos para obtener el grado de

Licenciado en **Microbiología**

Presenta:

Jorge Luis Gómez Hernández

Dirigido por:

Dr. Roberto Carlos Álvarez Martínez

Dr. Roberto Carlos Álvarez Martínez

Presidente

Firma

Dra Etzel Garrido Espinosa

Secretaria

Firma

Dr. Jesús Espinal Enríquez

Asesor

Firma

Dra. Mónica Figueroa Cabañas

Asesora

Firma

Centro Universitario
Querétaro, Qro.
Junio, 2024
México

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.

RESUMEN

El cáncer endometrial es una enfermedad multifactorial que afecta principalmente a mujeres de la tercera edad, tanto en México como en el resto del mundo (Bahena-González et al., 2023). Esta enfermedad, al igual que otros tipos de cáncer, se caracteriza por la presencia e interacción entre células cancerígenas y células del propio organismo dentro de un mismo tejido (heterogeneidad tumoral), lo que genera un microambiente tumoral que puede influir la progresión de la enfermedad y su respuesta a distintos tratamientos (Dagogo-Jack & Shaw, 2017; Makker et al., 2021). Considerando estos puntos, el presente trabajo utiliza datos de secuenciación con resolución de una sola célula (scRNA-seq), provenientes de un tumor primario de cáncer endometrial en una mujer de 70 años, con el objetivo de identificar algunos de los posibles agentes genéticos que hacen posible el desarrollo y supervivencia de esta enfermedad. Para ello, se implementaron una serie de análisis bioinformáticos utilizando el lenguaje de programación R (R Core Team, 2024), con el fin de obtener resultados confiables y reproducibles. Entre los procedimientos realizados, se destacan: la identificación de las células cancerígenas y otros tipos celulares presentes en la muestra analizada, la generación de redes pesadas de co-expresión de genes, la identificación de genes asociados con macrófagos y el estado canceroso de las células, el análisis de enriquecimiento funcional de estos genes y la evaluación de las características topológicas de las redes de genes obtenidas. Los resultados obtenidos son consistentes con investigaciones previas. Específicamente, se identificaron genes asociados con mecanismos de biogénesis ribosomal, eventos proliferativos, de aumento en la tasa de traducción proteica y de evasión a la respuesta inmune; lo que se ha descrito con anterioridad en distintos tipos de cáncer.

Palabras clave: cáncer endometrial, enriquecimiento funcional, heterogeneidad celular, redes pesadas de co-expresión de genes, scRNA-seq, topología de red

SUMMARY

Endometrial cancer is a multifactorial disease that primarily affects elderly women, both in Mexico and around the world (Bahena-González et al., 2023). This disease, like many other types of cancer, is characterized by the presence and interaction of cancerous cells and the body's own cells within the same tissue (tumor heterogeneity), creating a tumor microenvironment that can influence the progression of the disease and its response to different treatments (Dagogo-Jack & Shaw, 2017; Makker et al., 2021). Considering these points, this study utilizes single-cell resolution sequencing data (scRNA-seq) obtained from a primary endometrial cancer tumor in a 70-year-old woman, with the aim of identifying some of the possible genetic agents that enable the development and survival of this disease. To achieve this, a series of bioinformatic analyses were carried out using the R programming language (R Core Team, 2024) to obtain reliable and reproducible results. Among the procedures carried out, the following stand out: the identification of cancerous cells and other cell types present in the analyzed sample, the generation of weighted gene co-expression networks, the identification of genes associated with macrophages and the cancerous state of the cells, the functional enrichment analysis of these genes, and the evaluation of the topological characteristics of the obtained gene networks. These results are consistent with previous research. Specifically, genes associated with ribosomal biogenesis, proliferative events, upregulation of protein translation, and immune response evasion, which have been previously described in various types of cancer, were identified.

Key words: endometrial cancer, functional enrichment, cellular heterogeneity, weighted gene co-expression networks, scRNA-seq, network topology

A mis padres, Antonio y Karla.

A mi hermano, Carlos.

AGRADECIMIENTOS

Personalmente agradezco a:

Mi director de tesis, Roberto Álvarez, pues su guía y consejo me acompañaron a lo largo de la carrera e hicieron posible este trabajo.

Mis sinodales, Etzel Garrido, Jesús Espinal y Mónica Figueroa, por su apoyo en la revisión de este trabajo.

Mi familia, Antonio, Karla y Carlos, por apoyarme desde el primer día en que me recibieron en sus vidas.

Mis amigos, Joaquín, Juan y Octavio, por ofrecerme su apoyo y amistad durante tantos años.

Gabriela, por ser siempre mi apoyo y motivación para dar el siguiente paso.

TABLA DE CONTENIDOS

RESUMEN	2
SUMMARY	3
AGRADECIMIENTOS	5
TABLA DE CONTENIDOS	6
ÍNDICE DE FIGURAS	8
1. INTRODUCCIÓN	9
2. ANTECEDENTES	10
2.1 Cáncer.....	10
2.2 Cáncer endometrial.....	11
2.3 Heterogeneidad tumoral.....	12
2.4 Single-cell RNA-sequencing.....	13
2.5 Redes de co-expresión de genes.....	15
3. OBJETIVOS	16
3.1 Objetivo general.....	16
3.2 Objetivos específicos.....	16
4. MATERIALES Y MÉTODOS	16
4.1 Obtención de datos.....	17
4.2 Pre-procesamiento y control de calidad.....	17
4.3 Procesamiento pre-remoción de dobletes.....	20
4.4 Remoción de dobletes.....	24
4.5 Procesamiento post-remoción de dobletes.....	25
4.6 Identificación de células cancerosas.....	25
4.7 Preparación de datos para análisis de redes pesadas de co-expresión.....	26
4.8 Pre-procesamiento y control de calidad para WGCNA.....	27
4.9 Construcción de redes pesadas de co-expresión.....	28
4.10 Análisis de módulos asociados a cáncer y macrófagos.....	29
5. RESULTADOS	32
5.1 Creación de objeto Seurat.....	32
5.2 Control de calidad.....	32
5.3 Procesamiento preliminar.....	33

5.3.1 Obtención de genes variables.....	33
5.3.2 Reducción de dimensiones.....	33
5.3.3 Identificación de comunidades preliminar.....	33
5.4 Remoción de dobles.....	33
5.5 Procesamiento subsecuente y asignación de tipos celulares.....	34
5.6 Identificación de células cancerosas.....	36
5.7 Pre-procesamiento para el algoritmo WGCNA.....	37
5.7.1 Identificación de outliers.....	37
5.7.2 Normalización de datos.....	37
5.8 Construcción de redes pesadas.....	38
5.8.1 Selección de valores umbrales.....	38
5.8.2 Construcción de redes e identificación de módulos.....	38
5.9 Identificación de módulos asociados a cáncer.....	39
5.10 Enriquecimiento funcional.....	40
5.11 Análisis topológico y estructural de las redes.....	42
6. DISCUSIÓN.....	43
7. CONCLUSIONES.....	49
8. PERSPECTIVAS.....	50
9. REFERENCIAS.....	50

ÍNDICE DE FIGURAS

Figura 1. Control de calidad.....	19
Figura 2. Remoción de células.....	20
Figura 3. Normalización de los datos de expresión.....	21
Figura 4. Identificación de genes variables.....	22
Figura 5. Evaluación de los PCAs.....	23
Figura 6. Visualización de reducción dimensional UMAP.....	24
Figura 7. Niveles de expresión.....	26
Figura 8. Identificación de outliers.....	27
Figura 9. Normalización de los datos de expresión por clúster.....	28
Figura 10. Valores umbrales.....	29
Figura 11. Red de módulo rosa.....	31
Figura 12. Red de módulo cian.....	32
Figura 13. Remoción de doblete.....	34
Figura 14. Células por tipo celular.....	35
Figura 15. Visualización de clústeres.....	36
Figura 16. Tipos celulares.....	37
Figura 17. Dendograma de módulos.....	39
Figura 18. Correlación de módulos y características.....	40
Figura 19. Resultados del análisis de enriquecimiento funcional.....	41
Figura 20. Agrupamiento y visualización de los términos enriquecidos provenientes del análisis de los genes del módulo rosa.....	42

1. INTRODUCCIÓN

El cáncer endometrial es el sexto tipo de cáncer con mayor incidencia en mujeres a nivel mundial y el quinto en México, siendo el responsable por la muerte de 1,164 mujeres anualmente en el territorio nacional (Bahena-González et al., 2023). Es una enfermedad multifactorial que se desarrolla en la capa interna del útero, llamada endometrio, la cual es diagnosticada mayoritariamente en mujeres postmenopáusicas a raíz de signos y síntomas no patognomónicos (Bahena-González et al., 2023; Brown et al., 2023; Hanahan, 2022; Makker et al., 2021).

Esta enfermedad, al igual que muchos otros tipos de cáncer, está caracterizada por la presencia de diversos tipos celulares que interactúan cercanamente entre sí, pues coexisten en un mismo tejido llamado tumor, lo cual genera un microambiente específico a este (Dagogo-Jack & Shaw, 2017; Makker et al., 2021; Vitale et al., 2021). En particular, este tipo de afección es capaz de inducir el padecimiento en células previamente sanas y de manipular las células a su alrededor para sobrevivir y proliferar, lo que no solamente explica la ya mencionada presencia de varios tipos celulares, sino, que también puede afectar las respuestas ante los tratamientos (Brown et al., 2023; Hanahan, 2022; Vitale et al., 2021).

Comprender los mecanismos genéticos subyacentes del cáncer endometrial es crucial para entender con mayor profundidad esta enfermedad y desarrollar posibles estrategias terapéuticas más efectivas. La secuenciación de ARN de una sola célula (scRNA-seq) ha emergido como una herramienta capaz de identificar genes clave para la transformación maligna y la comunicación intratumoral (Dagogo-Jack & Shaw, 2017; Vitale et al., 2021)

Este trabajo hace uso de datos de scRNA-seq provenientes de una muestra de un tumor primario de cáncer endometrial con el objetivo de identificar genes asociados con el estado canceroso de las células. Los análisis ejecutados para esta tarea abarcan la determinación de tipos celulares presentes en un tumor primario, la construcción de redes pesadas de co-expresión y el análisis de los genes asociados a características de interés.

Los resultados obtenidos no solo respaldan investigaciones previas, sino que también revelan genes posiblemente involucrados en vías de señalización manipuladas por el cáncer. De esta manera se profundiza en el estudio del cáncer endometrial y se abren posibles avenidas para la identificación de agentes genéticos involucrados en esta enfermedad.

2. ANTECEDENTES

2.1 Cáncer

El cáncer es una enfermedad, o el término utilizado para referirse a un grupo de enfermedades, caracterizado por una proliferación celular descontrolada con el potencial de esparcirse a partes del cuerpo no afectadas (Brown et al., 2023; Centers for Disease Control and Prevention, 2021; National Cancer Institute, 2021). Estas enfermedades son asociadas de manera general con la formación de nuevos tejidos llamados tumores a partir de las células anómalas; sin embargo, existen tipos de cáncer que no forman tumores, e incluso hay tumores considerados benignos al ser incapaces de esparcirse (Brown et al., 2023; National Cancer Institute, 2021). Cabe recalcar que, aunque las diversas definiciones de este tipo de afección causan debate actualmente, existe un acuerdo general en varias de las alteraciones biológicas necesarias en la progresión de estos padecimientos (Brown et al., 2023).

De forma general, el cáncer es descrito como el resultado de mutaciones, daños y/o alteraciones en genes específicos o regiones completas del ADN con diversas funciones regulatorias (Brown et al., 2023; Hanahan, 2022; National Cancer Institute, 2021). Estas perturbaciones pueden ser causadas por fallas aleatorias en procesos biológicos naturales, como la replicación del genoma durante la división celular, el resultado de interacciones nocivas entre agentes químicos y el material genético de las células, exposición a estímulos físicos como radiación por luz ultravioleta, etc.; no obstante, el resultado es el mismo: modificaciones genéticas que permiten a las células actuar de forma independiente dentro del organismo mientras evaden una respuesta por parte del hospedero (Brown et al., 2023; Hanahan, 2022). Ahora bien, pese a que las causas de los distintos cánceres pueden asemejarse, las características entre cada caso pueden variar tan enormemente que resulta necesario clasificar y categorizar este tipo de enfermedad para su estudio y tratamiento (Carbone, 2020).

En la actualidad se han definido cientos de tipos de cáncer con base en el tipo de tejido u órgano del cual se origina la enfermedad (National Cancer Institute, 2021); sin embargo, los esfuerzos en investigación para cada tipo pueden variar en función de distintos factores (Ferlay et al., 2021). La prevalencia, incidencia, mortalidad y capacidad de diagnóstico y tratamiento de los distintos tipos de cáncer, son factores que afectan la cantidad de recursos destinados para su respectiva investigación (Ferlay et al., 2021; Sung et al., 2021). Es por estos motivos que algunos tipos de cáncer, como cáncer de seno y cáncer de próstata, son estudiados intensivamente; sin embargo, más investigación en otros tipos de cáncer es

necesaria para entender esta clase de enfermedades y ofrecer terapias efectivas y seguras. Uno de estos casos menos estudiados es el cáncer endometrial.

2.2 Cáncer endometrial

El cáncer endometrial es un tipo de cáncer uterino que se origina en la capa interna del útero llamada endometrio, es el sexto cáncer con mayor incidencia global, el quinto en México y es la causa de muerte de 1,164 mujeres anualmente en el territorio mexicano (Bahena-González et al., 2023; PDQ® Adult Treatment Editorial Board, 2024; The American Cancer Society medical and editorial content team, 2019). Existen una variedad de clasificaciones de cáncer endometrial diferenciadas con base en los tipos celulares de los cuales se origina la enfermedad; sin embargo, la mayor parte de los casos (~75%) corresponden a adenocarcinomas endometrioides (PDQ® Adult Treatment Editorial Board, 2024).

Los distintos casos de cáncer endometrial pueden ser categorizados en dos tipos de manera general: el tipo 1 y el tipo 2 (Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024). Los casos clasificados como tipo 1 se definen por un engrosamiento endometrial anormal inducido por estímulos desregulados de estrógeno, una composición basada en células con un grado de desdiferenciación bajo o nulo y prognosis favorable (Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024). Por otro lado, el tipo 2 de cáncer endometrial se caracteriza por la presencia de células con grados de desdiferenciación moderados y altos, tejido endometrial atrofiado y la ausencia de estímulos hormonales descontrolados (Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024; The American Cancer Society medical and editorial content team, 2019). Cabe destacar que la *Fédération Internationale de Gynécologie et d'Obstétrique* (FIGO) ha diseñado la estadificación de esta enfermedad con cuatro posibles estadios con ocho sub-estadios.

El estadio I, en el cual el tumor está confinado al cuerpo del útero, se divide en los estadios IA y IB dependiendo del grado de invasión en el miometrio (PDQ® Adult Treatment Editorial Board, 2024). Se considera que un tumor endometrial se encuentra en estadio II cuando hay una invasión del estroma del cuello uterino sin diseminación fuera del útero (PDQ® Adult Treatment Editorial Board, 2024). El estadio III se caracteriza y se divide en IIIA, IIIB, IIIC1 y IIIC2 con base en la diseminación (metástasis) local o regional del tumor desde tejidos cercanos como la vagina, hasta estructuras como nódulos linfáticos paraaórticos (PDQ® Adult Treatment Editorial Board, 2024).

Los signos y síntomas más comunes con los que se puede comenzar el proceso de diagnóstico de cáncer endometrial son dolor abdominal, dolor al orinar y durante el coito,

sangrado menstrual abundante, sangrado esporádico y sangrado post menopáusico; sin embargo, debido a que estos fenómenos no son patognomónicos del cáncer endometrial, es necesario realizar estudios para confirmar la presencia o ausencia de la enfermedad (Bahena-González et al., 2023; Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024). Las técnicas empleadas principalmente para obtener diagnósticos certeros incluyen ecografías transvaginales, tomografías computarizadas, obtención de imágenes por resonancia magnética, histeroscopías y biopsias (Bahena-González et al., 2023; Makker et al., 2021).

El tipo de cáncer endometrial, así como el grado de desdiferenciación celular que presenta la enfermedad, son dos de los factores que determinan en gran medida los tratamientos disponibles para las pacientes (Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024). Usualmente, el tratamiento consta de una histerectomía con salpingooforectomía bilateral con opción a recibir radioterapia adyuvante, esto dependiendo de si se presenta una invasión del miometrio o se detectan células desdiferenciadas grado tres (American Cancer Society, 1930; Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024). Adicionalmente, casos que presentan tumores tipo 2, metástasis, son recurrentes y/o son inoperables, son tratados con combinaciones de regímenes de radioterapia, quimioterapia, terapia hormonal y/o cirugía en caso de ser posible (Makker et al., 2021; PDQ® Adult Treatment Editorial Board, 2024). No obstante, se debe considerar que la progresión de esta enfermedad y la respuesta a los tratamientos depende en gran medida de factores específicos de cada paciente, por lo que el estudio de la variabilidad existente entre cada paciente y su respectivo cáncer resulta indispensable.

2.3 Heterogeneidad tumoral

Con el advenimiento de técnicas cada vez más precisas que permiten el estudio del cáncer como grupo de enfermedades, algunos esfuerzos se han enfocado en el entendimiento de los tumores formados por estos padecimientos como un ecosistema complejo (Dagogo-Jack & Shaw, 2017; Vitale et al., 2021). Los resultados de estos trabajos señalan la importancia de elementos como la heterogeneidad genética, distribución espacial de las células, temporalidad, etc., como agentes determinantes de la composición celular única en los tumores de cada paciente (Dagogo-Jack & Shaw, 2017; Vitale et al., 2021). Sin embargo, la sola identificación de estos componentes es insuficiente para determinar los posibles mecanismos y efectos que la heterogeneidad intratumoral (HIT) tiene en el cáncer.

A fin de estudiar los agentes causantes de la HIT, se han identificado dos cualidades generales no excluyentes que los elementos anteriormente mencionados pueden conferir a

los tumores: plasticidad y tolerancia o evasión (Ramón y Cajal et al., 2020; Vitale et al., 2021). Por un lado, la plasticidad conferida en este contexto se refiere a la capacidad de generar una respuesta rápida a estímulos por parte de las células que conforman el tumor, sean malignas o no (Vitale et al., 2021). Esta plasticidad permite al tumor desarrollar poblaciones celulares con diferentes capacidades proliferativas e invasivas que, en caso de enfrentarse a los estímulos correctos, pueden desencadenar en respuestas adaptativas que empeoren el padecimiento o incluso disminuyan la efectividad de los tratamientos (Dagogo-Jack & Shaw, 2017; McGranahan & Swanton, 2017; Ramón y Cajal et al., 2020; Vitale et al., 2021).

Por otra parte, las cualidades de tolerancia o evasión que adquieren las células en tumores malignos, se refieren a la capacidad de las mismas para evitar, controlar o inhibir las respuestas inmunes del hospedero en contra del cáncer (McGranahan & Swanton, 2017; Vitale et al., 2021). Se ha demostrado la capacidad de algunas subpoblaciones presentes en los tumores de alterar su metabolismo, perfiles de transcripción y rutas de señalización, así como a las propias células inmunes, para poder evadir las respuestas de defensa del hospedero e incluso utilizarlas a su favor (Gonzalez et al., 2018; McGranahan & Swanton, 2017; Vitale et al., 2021).

Independientemente de cuál mutación o cambio esté involucrado en el desarrollo de la HIT, los efectos que esta tiene sobre los pacientes pueden observarse en forma de peores pronósticos y resultados de la progresión del cáncer (Dagogo-Jack & Shaw, 2017; McGranahan & Swanton, 2017; Ramón y Cajal et al., 2020; Vitale et al., 2021). Es por esto que se han desarrollado metodologías para entender y analizar la heterogeneidad específica presente en cada paciente, dentro de las cuales se destaca la secuenciación del transcriptoma de muestras biológicas con resolución de una sola célula.

2.4 Single-cell RNA-sequencing

La aplicación de tecnologías de secuenciación para la investigación en distintos tipos de cáncer no es algo nuevo; sin embargo, con los avances habidos en estas tecnologías, la información recabada y analizada ha cambiado considerablemente con el paso del tiempo (Dagogo-Jack & Shaw, 2017; Vitale et al., 2021; Zhao et al., 2019). En particular, las tecnologías de nueva generación capaces de secuenciar el transcriptoma completo de una sola célula a la vez (scRNA-seq), han contribuido enormemente en la investigación de los microambientes generados por cáncer (Camps et al., 2023; Dagogo-Jack & Shaw, 2017; Vitale et al., 2021).

Tal como el resto de tecnologías de secuenciación, los métodos de scRNA-seq tienen la finalidad de identificar, multiplicar y cuantificar moléculas de interés dentro de las muestras analizadas; no obstante, no sólo las moléculas secuenciadas difieren en esta técnica, sino que también las metodologías aplicadas. La variedad existente para secuenciar moléculas de ARN ofrece diferentes opciones según los objetivos de la investigación realizada y los recursos disponibles; sin embargo, todas las aproximaciones comparten algunos elementos (Jovic et al., 2022).

De manera general, se puede dividir el trabajo ejecutado en las técnicas de scRNA-seq en dos fases: una fase experimental y una fase de análisis de datos (Jovic et al., 2022). Por un lado, la fase experimental consta, a grandes rasgos, del aislamiento y captura de las células de interés, lisis celular, retrotranscripción, amplificación, preparación de librerías y secuenciación (Jovic et al., 2022; Kolodziejczyk et al., 2015). Por otra parte, el análisis de datos abarca desde la ejecución de medidas de control de calidad, hasta procesamientos especializados basados en el interés del trabajo; esta fase tiene mayor variabilidad según las herramientas computacionales utilizadas, pero todas buscan extraer e interpretar la información biológica (Huang et al., 2021; Jovic et al., 2022).

La fase experimental comienza con la disociación de las muestras de interés a fin de obtener una suspensión de células separadas entre sí; este proceso es llevado a cabo a través de combinaciones de tratamientos físicos, como la separación de tejidos con navajas, y procedimientos químicos encargados de degradar compuestos celulares que mantienen unidas a las células (Jovic et al., 2022; Koboldt, 2020). Posteriormente, se lleva a cabo uno de los procesos críticos en los protocolos de scRNA-seq, el aislamiento y captura de las células de interés. Este aislamiento y captura puede ser logrado con una variedad de herramientas y tecnologías; sin embargo, el uso de microfluidos es prevalente en el campo (Jovic et al., 2022; Salomon et al., 2019). Este aislamiento por microfluidos permite la generación de microgotas que, de manera individual, contienen una sola célula a la vez, una perla de gel con oligonucleótidos incrustados, varias enzimas necesarias y un microambiente propicio para la supervivencia de moléculas de ARN y ADN (Jovic et al., 2022; Salomon et al., 2019; Zheng et al., 2017). Después de la encapsulación de las células en las microgotas, dichas células son lisadas de tal modo que las moléculas de ARN pueden anclarse a los extremos 3' de los oligonucleótidos incrustados en la perla de gel (Jovic et al., 2022; Kolodziejczyk et al., 2015; Salomon et al., 2019; Zheng et al., 2017). Una vez ancladas las secuencias de ARN en sus respectivas perlas, estas son retro transcritas con la finalidad de obtener moléculas de ADN complementario (cADN) que, finalmente serán amplificadas y subsecuentemente secuenciadas con alguna plataforma de secuenciación de nueva generación (Jovic et al., 2022; Salomon et al., 2019; Zheng et al., 2017). Ahora bien,

pese a la existencia de diversas adaptaciones y modificaciones en los muy variados protocolos de scRNA-seq habidos actualmente, todos tienen como finalidad la secuenciación del cADN obtenido para poder procesar los datos obtenidos.

El análisis subsecuente a la secuenciación inicia con la implementación de herramientas de control de calidad y de alineamiento (Hwang et al., 2018; Jovic et al., 2022). Por un lado, las herramientas de control de calidad ofrecen información a los usuarios respecto a las lecturas obtenidas, esta información permite discernir lecturas fiables entre, posiblemente, millones de datos que posteriormente son utilizados en el flujo de trabajo (Hwang et al., 2018; Jovic et al., 2022). Por otra parte, el alineamiento de las secuencias se refiere al proceso mediante el cual se mapean las lecturas a un genoma de referencia con el objetivo de identificar los genes cuyos transcritos fueron detectados y secuenciados (Hwang et al., 2018; Jovic et al., 2022). Cabe recalcar que, aunque el control de calidad y alineamiento sean prácticas comunes en el análisis de scRNA-seq, la mayor parte del análisis de datos es totalmente dependiente y personalizada a cada trabajo. En el caso particular de esta investigación, el trabajo se centra en la elaboración y análisis de redes pesadas de co-expresión de genes.

2.5 Redes de co-expresión de genes

Gracias al desarrollo y constantes mejoras de tecnologías cuyo objetivo es la obtención de perfiles transcriptómicos de organismos completos o muestras biológicas, considerables esfuerzos científicos han sido invertidos con el propósito de obtener conocimiento a partir de los datos de expresión génica (Chowdhury et al., 2019). Dentro de esta área de investigación, el análisis de co-expresión genética por medio de redes es uno de los métodos empleados más comúnmente (Sarmah & Bhattacharyya, 2021).

El análisis de redes de co-expresión de genes pretende identificar, agrupar y representar en un grafo genes con perfiles de expresión similares en diferentes muestras (Chowdhury et al., 2019; Montenegro, 2022). Este proceso permite a los usuarios determinar módulos de genes altamente correlacionados, inferir funciones putativas de los genes habidos en cada módulo, identificar genes con alta conectividad, asociar módulos de genes a características específicas de las muestras, etc. (Chowdhury et al., 2019; Langfelder & Horvath, 2008; Montenegro, 2022; Sarmah & Bhattacharyya, 2021).

Pese a las ramificaciones y modificaciones que han surgido a partir de este tipo de análisis, la elaboración de redes de co-expresión conserva algunos pasos a seguir en todas sus presentaciones (Montenegro, 2022). El primero de estos pasos es la eliminación de datos que pudieran sesgar los análisis como genes con niveles de expresión demasiado bajos o nulos, o muestras que no presenten expresión en la mayoría de los genes analizados

(Chowdhury et al., 2019; Montenegro, 2022). Posteriormente, sesgos inherentes a los datos de secuenciación, como la longitud de los genes y la profundidad de la secuenciación, son reducidos mediante el ajuste de los valores de expresión a través del uso de métodos de normalización (Montenegro, 2022; Sarmah & Bhattacharyya, 2021). A continuación, se lleva a cabo el cálculo de similitud en los patrones de expresión entre todos los posibles pares de genes y se obtiene una matriz de similitud; esta matriz contiene los coeficientes de correlación y funge como la base para el cálculo de las matrices de adyacencia que definen las redes (Montenegro, 2022; Sarmah & Bhattacharyya, 2021). Cabe recalcar que, aunque procedimientos más específicos como el análisis de redes pesadas de co-expresión, requieren mayor procesamiento para obtener información como los módulos de genes y su asociación a características específicas, los coeficientes de correlación y las matrices de adyacencia son la base para su correcta ejecución.

3. OBJETIVOS

3.1 Objetivo general

Inferir genes asociados al estado canceroso de células existentes en un tumor primario de cáncer endometrial con base en el análisis de redes pesadas de co-expresión.

3.2 Objetivos específicos

- Identificar tipos celulares habidos a partir de datos de scRNA-seq.
- Generar redes pesadas de co-expresión partiendo de datos procesados de scRNA-seq.
- Identificar módulos de genes asociados a características cancerígenas.
- Determinar funciones moleculares de genes pertenecientes a módulos de interés.
- Comparar y analizar las redes obtenidas de módulos asociados a células cancerosas y macrófagos.

4. MATERIALES Y MÉTODOS

Este trabajo se elaboró en su totalidad de manera computacional mediante el lenguaje de programación R versión 4.3.3 en el IDE (*Integrated development environment*) conocido como RStudio (R Core Team, 2024). La muestra biológica analizada a lo largo del trabajo, así como parte del código utilizado, provienen de la investigación realizada por Regner y

colaboradores (2021). La totalidad del código empleado en este trabajo puede ser consultado públicamente en <https://github.com/Jorge-LGH/Thesis.git>.

4.1 Obtención de datos

La obtención de datos biológicos sin procesar de un tumor primario de cáncer endometrial de tipo endometrioides en etapa IA, provenientes de una mujer caucásica de 70 años, se realizó mediante la descarga de tres archivos de la base de datos *Gene Expression Omnibus* (GEO) del *National Center for Biotechnology Information* (NCBI). Los archivos descargados consisten de dos tablas en formato tsv y una matriz dispersa alojada en un archivo con extensión mtx. La primera tabla cuenta con los identificadores únicos de las células secuenciadas en la muestra, mientras tanto, la segunda tabla contiene una serie de identificadores asociados a todos los genes detectados. Finalmente, la matriz dispersa representa el conteo de lecturas de cada gen por cada célula secuenciada. Todos los archivos descargados pertenecen a la misma muestra de tumor primario de cáncer endometrial, cuyos datos fueron depositados en la base de datos GEO por Regner y colaboradores (2021) bajo el código de acceso GSM5276935.

Además de los datos biológicos, tablas y archivos de texto plano empleados en la asignación de clústeres y la identificación de tipos celulares fueron extraídos de diversas fuentes. Se descargó una tabla con marcadores genéticos de *Homo sapiens* a partir de la base de datos *PanglaoDB* (Franzén et al., 2019). El archivo de texto con el ensamblaje del genoma procesado fue descargado del repositorio de Github elaborado durante el mismo trabajo del cual se obtuvieron los datos biológicos (Regner et al., 2021). La tabla con genes utilizados para discernir células estromales y células inmunes fue igualmente descargada del repositorio de Github previamente mencionado. Finalmente, datos biológicos de endometrio humano para la asignación de tipo celular fueron descargados de la base de datos GEO bajo el acceso GSE111976 (W. Wang et al., 2020).

4.2 Pre-procesamiento y control de calidad

Previo al control de calidad, se empleó la versión 4.4.0 del paquete Seurat (Hao et al., 2021) para generar un objeto multicapa de clase Seurat en R (R Core Team, 2024). Para la elaboración del objeto, se acoplaron los archivos con la información de los genes secuenciados, los códigos de identificación para cada célula y la matriz con el conteo de moléculas por gen de cada célula. Durante el proceso de elaboración del objeto, instancias de genes duplicados fueron identificadas y eliminadas automáticamente por medio de la función ``CreateSeuratObject`` del paquete Seurat (Hao et al., 2021).

En cuanto al control de calidad de las secuencias obtenidas, se trabajó con tres propiedades de las muestras que podrían ser utilizadas para detectar datos que podrían sesgar el análisis. Estas propiedades fueron: el total de secuencias obtenidas por muestra, el número de genes detectados en cada muestra y el porcentaje de genes mitocondriales presentes con respecto al total de genes en cada muestra (**Fig. 1**). El proceso de discriminación de las muestras se llevó a cabo a través del uso de la desviación media absoluta (MAD por sus siglas en inglés). En este caso, se implementó la función `isOutlier` del paquete `scater` (McCarthy et al., 2017) para detectar muestras que estuvieran por debajo de dos MADs en el caso de las secuencias totales y el número de genes detectados por cada muestra; en cuanto al porcentaje de genes mitocondriales, aquellas muestras que se encontraran con un porcentaje superior a dos MADs igualmente se denominaron como outliers. Finalmente, se removieron todas las células que fueron designadas como outliers del objeto Seurat previamente elaborado (**Fig. 2**).

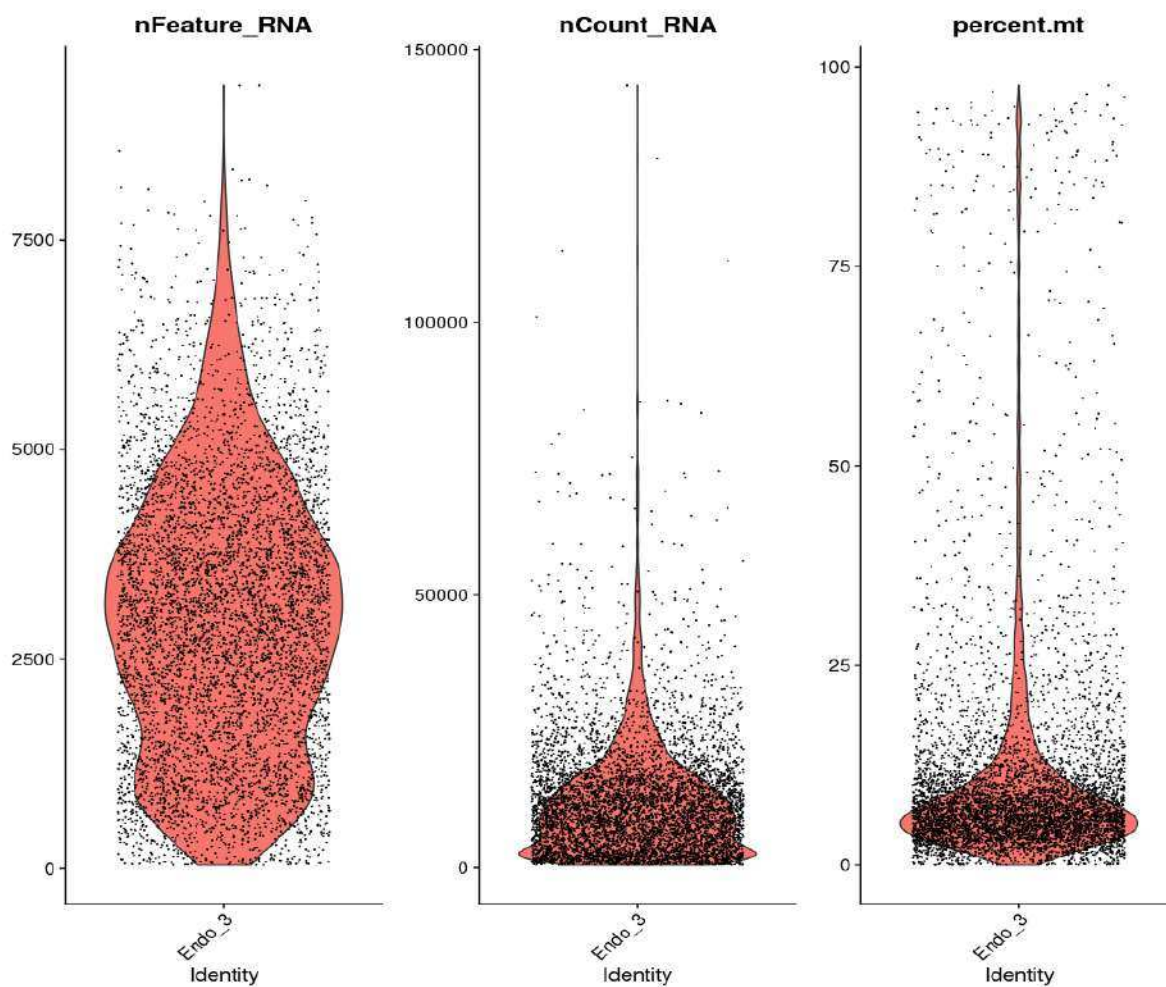


Figura 1. Control de calidad. Visualización del número de genes identificados, número de lecturas por célula y porcentaje de genes mitocondriales habidos en cada célula. Cada columna representa respectivamente las tres propiedades mencionadas.

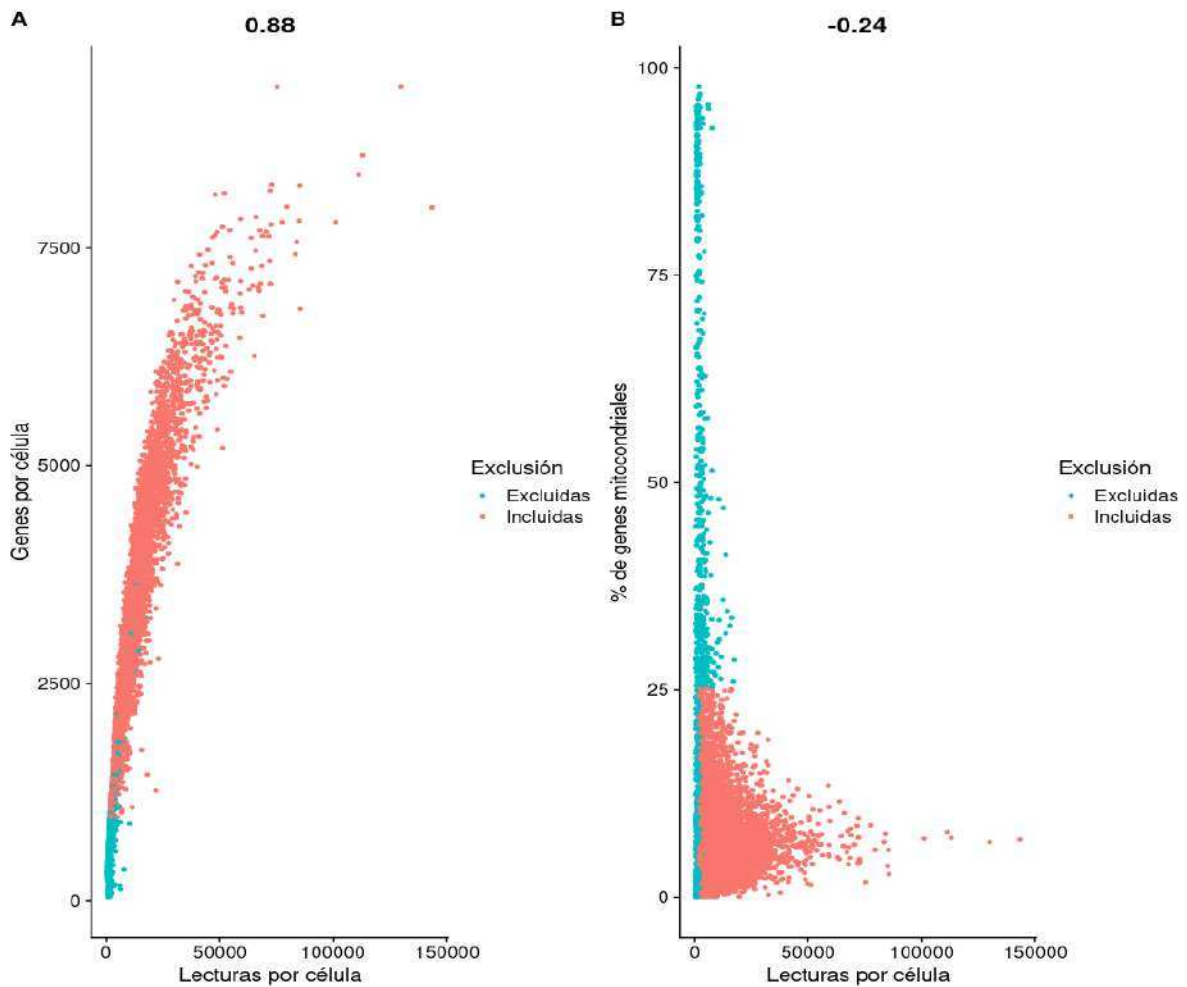


Figura 2. Remoción de células. Remoción de 929 células con base en las características de número de genes identificados, número de lecturas por célula y porcentaje de genes mitocondriales en cada muestra. Las células conservadas pueden visualizarse en un color salmón, mientras que las células removidas se observan de un color aqua. A: Scatter plot cuyos ejes X y Y representan el número de lecturas y el número de genes habidos en cada célula respectivamente. B: Scatter plot cuyos ejes X y Y representan el número de lecturas y el porcentaje de genes en cada célula respectivamente.

4.3 Procesamiento pre-remoción de dobles

Después del filtrado de muestras, se normalizaron los conteos de secuencias de cada célula por cada gen con la función `NormalizeData` del paquete Seurat (Hao et al., 2021) (**Fig. 3**). Esta función usa el método “LogNormal” basado en el uso del logaritmo natural para transformar los valores en una escala que se aproxime a una distribución normal (Hao et al., 2021). Una vez que se contaba con los datos normalizados, se procedió a determinar los 2,000 genes con mayor variabilidad con ayuda de la función `FindVariableFeatures` con el método de transformación estabilizadora de varianza (VST por su abreviatura en inglés) (**Fig. 4**).

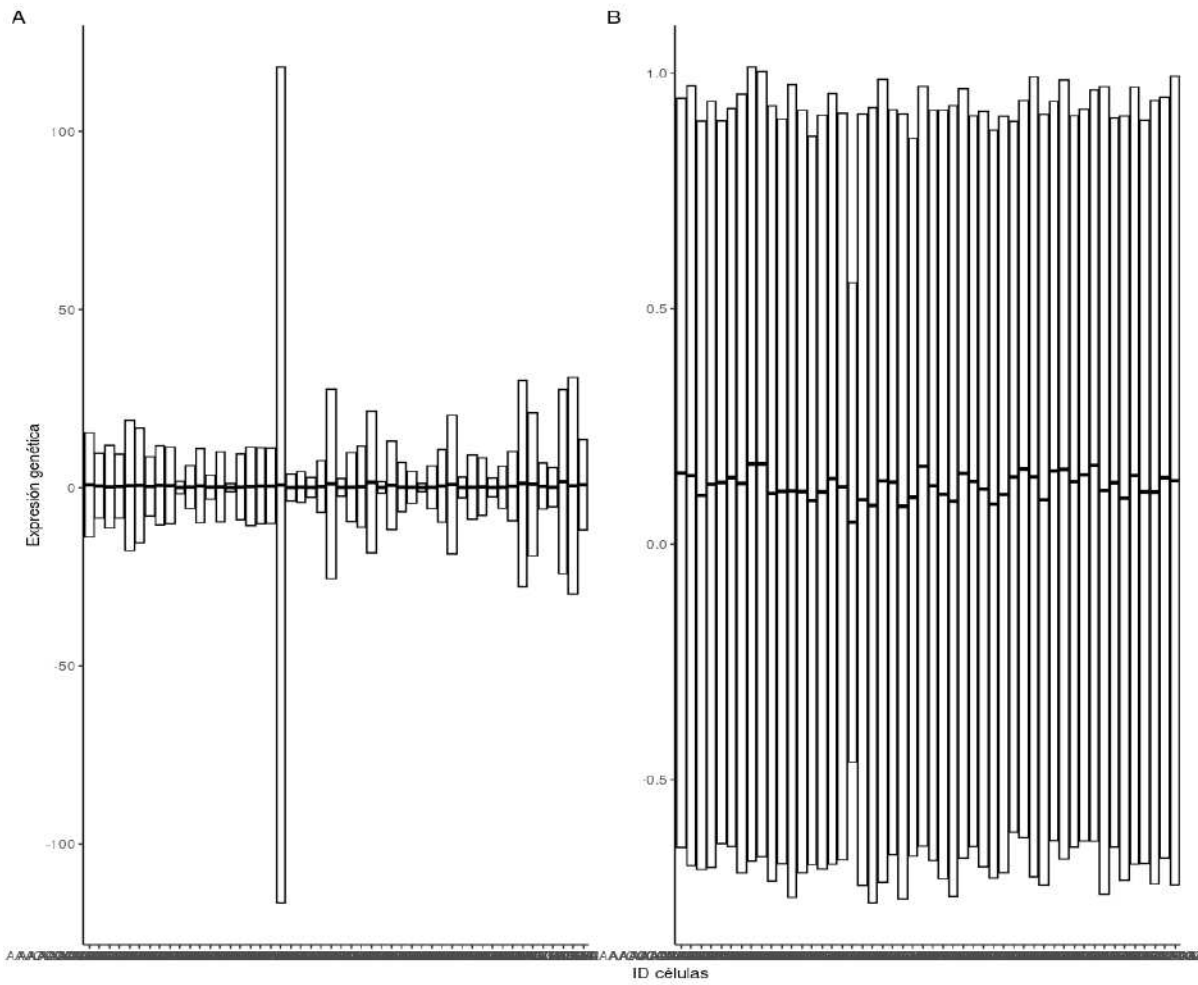


Figura 3. Normalización de los datos de expresión. A) Se muestra la distribución de los datos de expresión genética de 50 muestras previo a la realización de la normalización. B) Distribución de los datos de expresión genética de 50 células después de haber implementado la normalización.

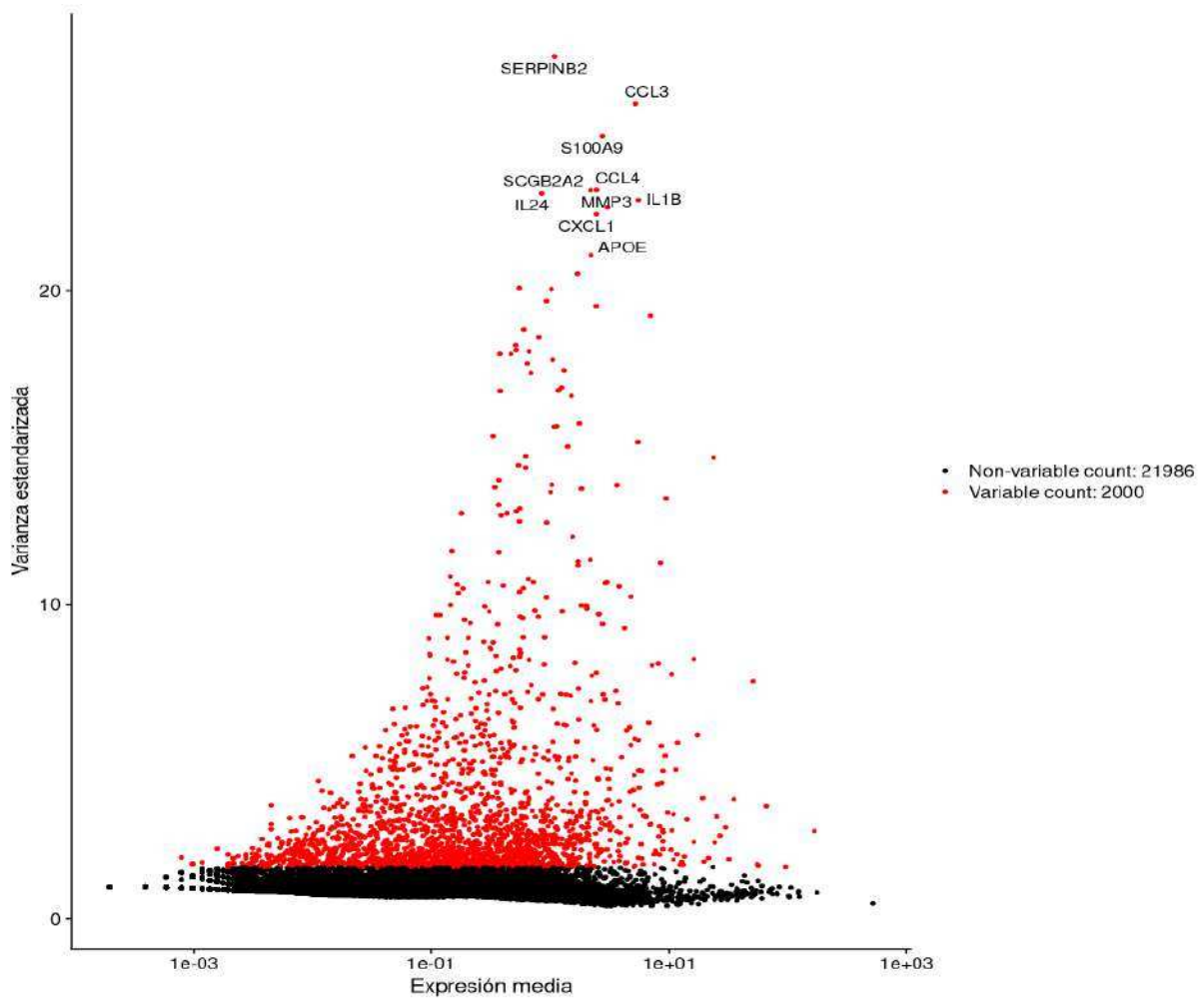


Figura 4. Identificación de genes variables. Identificación de los 2,000 genes con mayor variabilidad en sus perfiles de expresión. Los puntos rojos corresponden a los 2,000 genes seleccionados y se aprecian los nombres de los 10 genes con mayor variabilidad.

Tras haber definido los genes más variables, se efectuó la primera técnica de reducción de dimensiones: un análisis de componentes principales (PCA abreviado en inglés) sobre los valores de expresión de los genes con mayor variabilidad previamente identificados. Luego de haber obtenido los PC's, se determinó la significancia estadística de cada componente principal a través de un análisis *jackstraw* y se optó por emplear los 50 componentes principales calculados (**Fig. 5**). Tanto el PCA, como la determinación de significancia estadística, se llevaron a cabo con funciones del paquete Seurat (Hao et al., 2021).

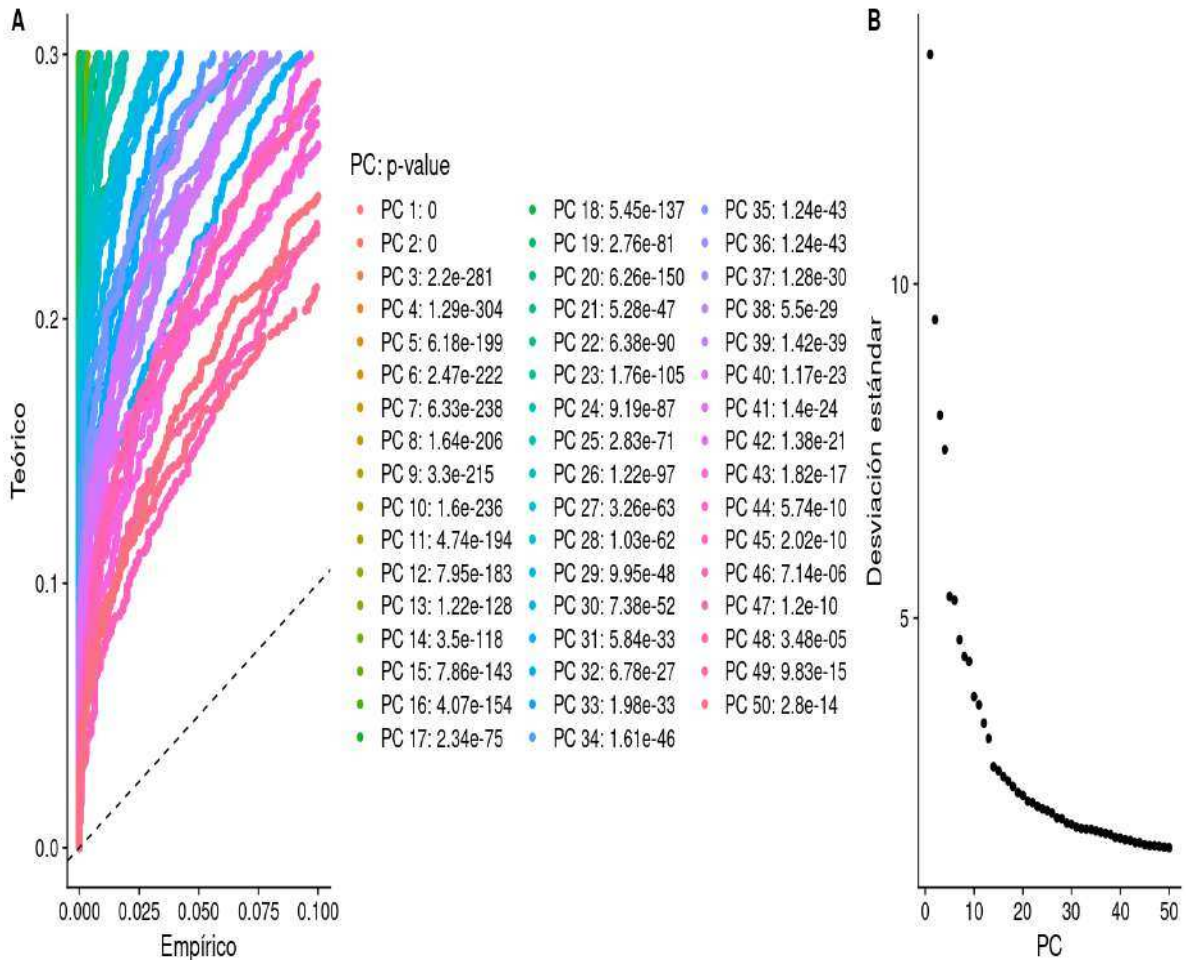


Figura 5. Evaluación de los PCAs. A) Visualización de la significancia estadística de todos los componentes principales calculadora con base en el análisis *jackstraw*. B) Representación de la desviación estándar correspondiente a cada componente calculado; la altura de cada punto en el *elbow plot* corresponde a la varianza explicada por cada componente.

Para finalizar el procesamiento previo a la remoción de dobletes, se generaron clústeres de células a partir de los datos obtenidos por el PCA. Este procedimiento se compone por dos partes: el cálculo de los k vecinos más próximos y la clusterización de los datos; las funciones `FindNeighbors` y `FindClusters` del paquete Seurat (Hao et al., 2021) efectúan respectivamente ambos pasos. Al concluir con la elaboración de clústeres, se aplicó la técnica de reducción de dimensiones conocida como UMAP (*Uniform Manifold Approximation and Projection*) para obtener una visualización más acertada de la distribución de los clústeres de células recién formados (**Fig. 6**).

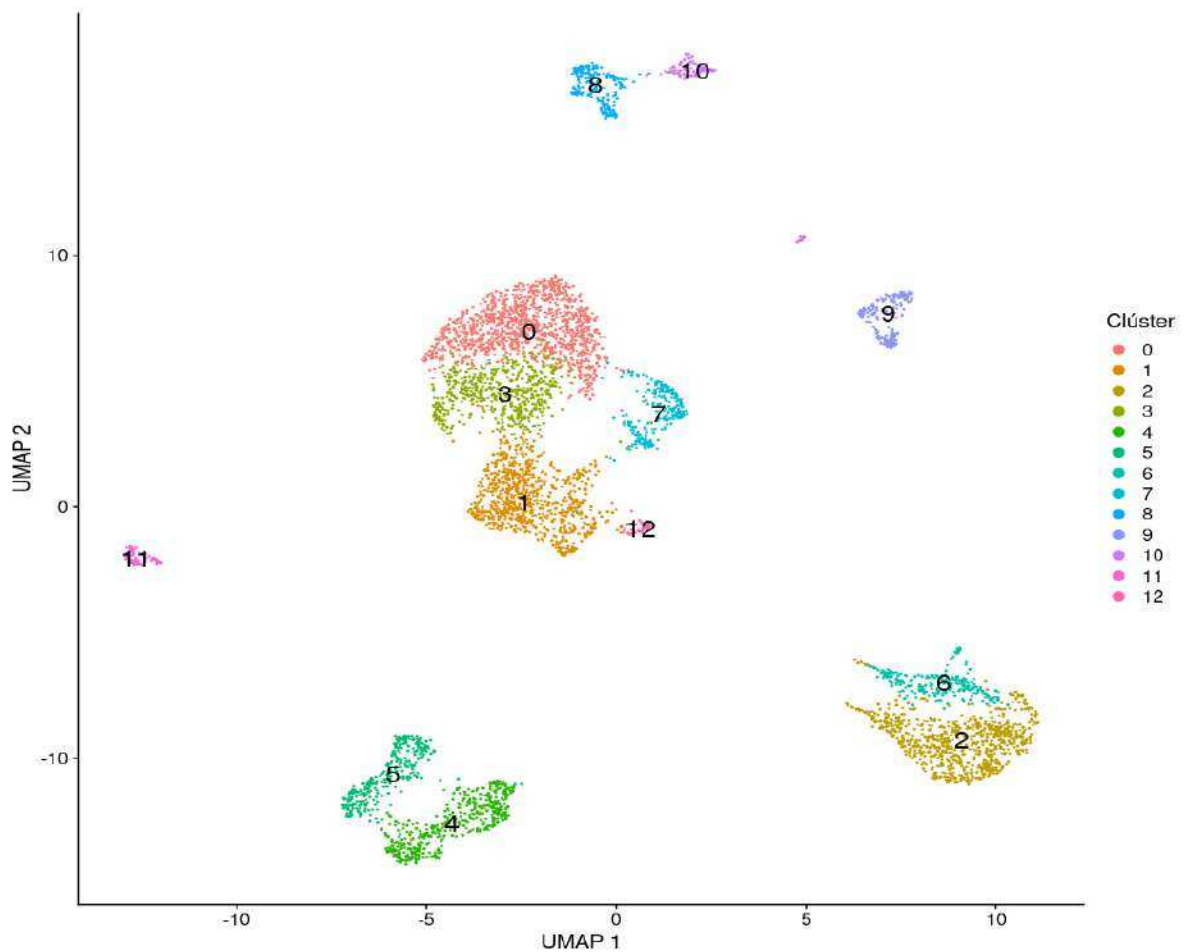


Figura 6. Visualización de reducción dimensional UMAP. Los colores representan los 13 clústeres a los cuales las células fueron asignadas.

4.4 Remoción de dobletes

De primera instancia, a través de la función `Improved_Seurat_Pre_Process` del paquete `DoubletDecon` (DePasquale et al., 2019), se extrajeron la matriz de expresión normalizada, los primeros 50 genes diferencialmente expresados de cada clúster y la lista de identificadores para cada clúster (todos provenientes del objeto tipo `Seurat` previamente elaborado) para ser transformados en objetos de clase `ICGS` y almacenados en una nueva variable. Una vez creada la variable, los datos almacenados en la misma fueron procesados con la función `Main_Doublet_Decon` para identificar presuntos clústeres de dobletes (DePasquale et al., 2019). El resultado de este procedimiento fue una nueva variable con los identificadores de las muestras consideradas como dobletes.

A continuación, se realizó un procedimiento adicional de detección de dobletes basado en los datos de expresión de las muestras con el paquete `DoubletFinder` (McGinnis et al.,

2019). El flujo de trabajo con este paquete consistió en la generación e inclusión de dobletes artificiales en los datos y una subsecuente discriminación de dobletes reales basado en la proporción de vecinos artificiales. El resultado fue una variable con los identificadores únicos de muestras consideradas como dobletes.

Finalmente, se consideraron como verdaderos dobletes para su remoción aquellas células cuyos identificadores únicos estuvieran presentes en los resultados de ambos métodos. Estos dobletes fueron removidos, junto con sus respectivos datos, del objeto tipo Seurat para seguir el trabajo.

4.5 Procesamiento post-remoción de dobletes

Los datos obtenidos hasta este momento fueron procesados de la misma manera en que se hizo previo a la remoción de dobletes. Se normalizaron los datos, se obtuvieron los 2,000 genes con mayor variabilidad, se efectuó una reducción de dimensiones mediante un PCA, la determinación de clústeres de células y una segunda reducción de dimensiones por medio de un UMAP. La única diferencia respecto al procesamiento previo fue un cambio de 0.5 a 0.7 en el parámetro de resolución en la función `FindClusters` del paquete Seurat (Hao et al., 2021).

Posteriormente, se realizó la asignación de tipo celular para cada clúster basado en los datos procesados por Wang y colaboradores (2020) disponibles en el repositorio de datos GEO. Para llevar a cabo esta asignación, los datos fueron utilizados como el conjunto de datos de referencia de expresión génica para asignar un tipo celular a cada clúster con ayuda de la función `SingleR` del paquete SingleR (Aran et al., 2019). Los tipos celulares obtenidos para todas las muestras de cada clúster fueron almacenados en el objeto tipo Seurat.

4.6 Identificación de células cancerosas

Para identificar las células cancerosas y su respectivo clúster, la función `FindMarkers` del paquete Seurat (Hao et al., 2021) fue empleada para evaluar la expresión diferencial de los genes WFDC2, MUC16 y KIT un clúster a la vez (**Fig. 7**). Todos aquellos clústeres que presentaron una expresión 0.5 veces mayor al promedio del resto en al menos uno de los genes ya mencionados, fueron asignados como muestras cancerosas tal como en el trabajo de Regner y colaboradores (2021); sin embargo, al considerar la posible expresión basal de WFDC2 en células ciliadas saludables y su distribución en el UMAP, los clústeres pertenecientes al tipo celular ciliado fueron removidos de las muestras consideradas como cáncer.

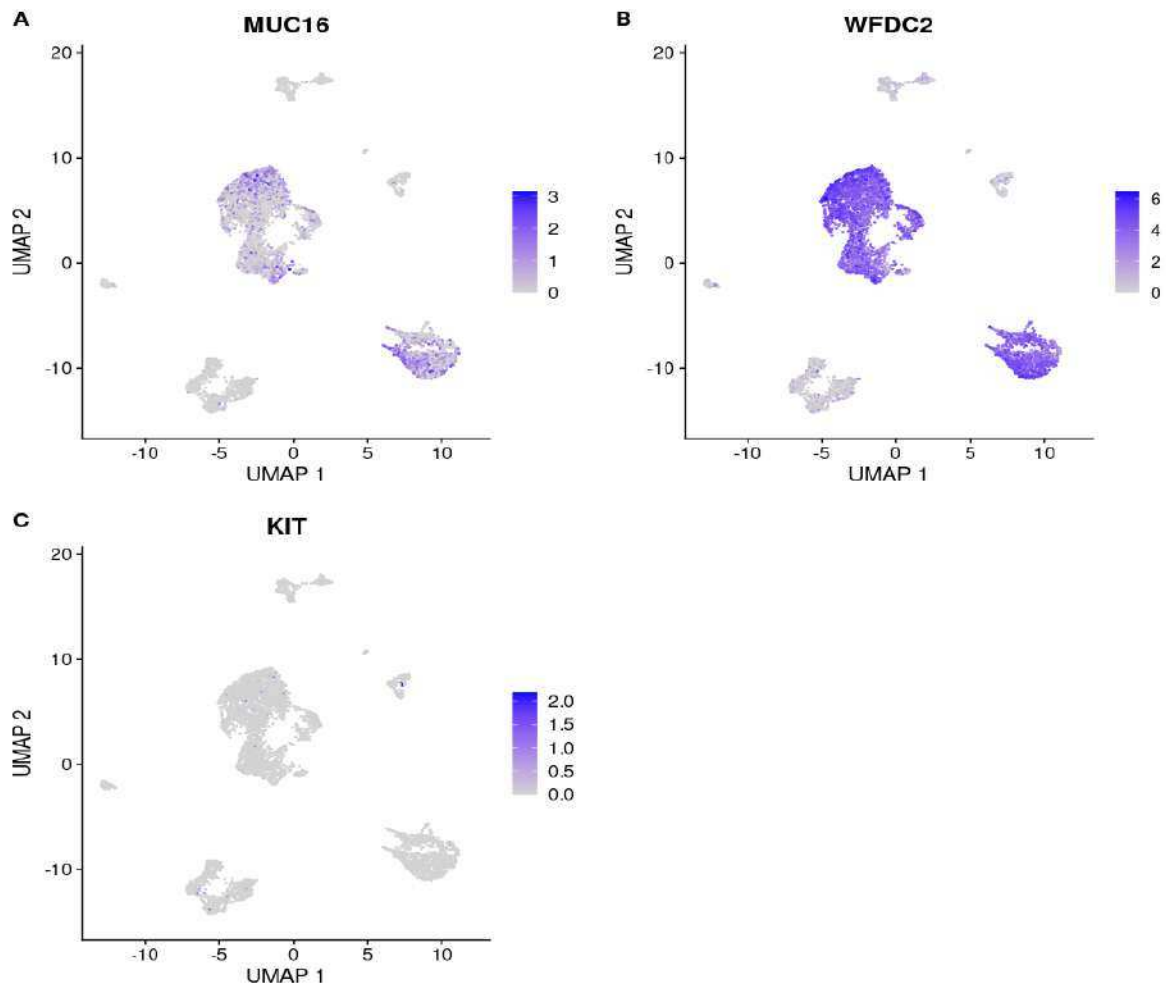


Figura 7. Niveles de expresión. Visualización de los niveles de expresión pertenecientes a los genes MUC16, WFDC2 y KIT sobre una proyección de la reducción dimensional UMAP. A) Niveles de expresión del gen MUC16. B) Niveles de expresión del gen WFDC2. C) Niveles de expresión del gen KIT.

4.7 Preparación de datos para análisis de redes pesadas de co-expresión

Con el objetivo de poder aplicar el análisis de redes pesadas de co-expresión (WGCNA por sus siglas en inglés) creado por Langfelder y Horvath (2008), se utilizaron los datos de expresión de cada clúster y se extrajeron secciones de sus metadatos.

Para llevar a cabo este proceso, se tomaron las matrices de expresión sin normalizar de cada clúster y se almacenaron como matrices independientes. Posteriormente, se efectuó una suma de renglones en cada matriz, de modo que el producto fuera un vector de tamaño igual al número de genes analizados. Finalmente, se unieron todos los vectores para generar una matriz con toda la información. Respecto a los metadatos, solamente se extrajeron los tipos celulares de cada clúster y si habían sido determinados como cancerosos o no.

4.8 Pre-procesamiento y control de calidad para WGCNA

Una vez con los datos en el formato adecuado, se procedió con el control de calidad según las recomendaciones de la herramienta WGCNA (Langfelder & Horvath, 2008). De primera instancia, se calculó la matriz de distancias entre los clústeres y se agruparon mediante una clusterización jerárquica. Posteriormente se efectuó un PCA en los datos. Tanto los resultados de la clusterización, como la reducción de dimensiones por medio del PCA, fueron graficados y se eliminaron aquellos clústeres considerados como outliers (**Fig. 8**). Este proceso fue llevado a cabo con funciones integradas al lenguaje R y la función `hclust` del paquete `fastcluster` (Müllner, 2013; R Core Team, 2024).

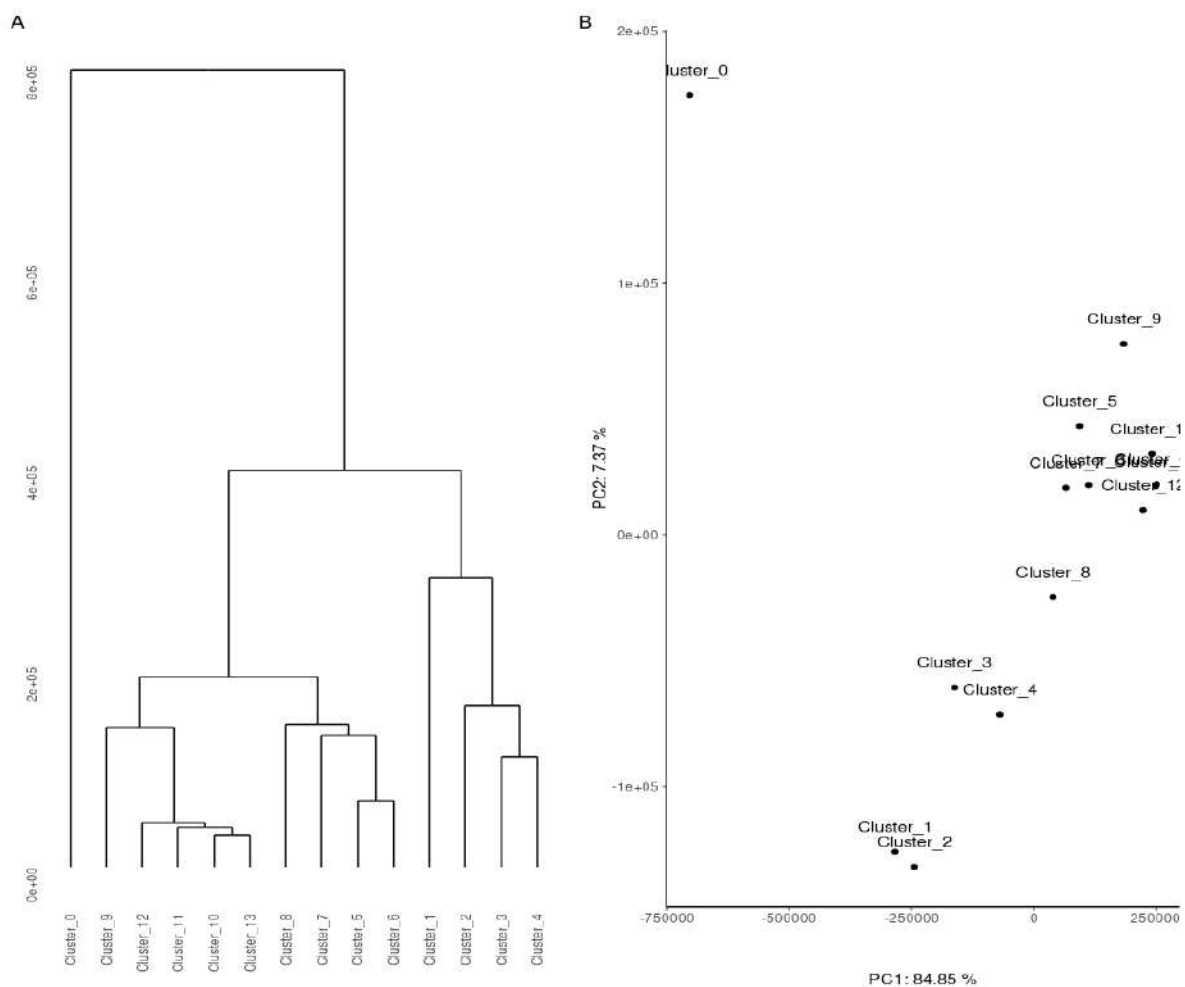


Figura 8. Identificación de outliers. A) Dendrograma de la clusterización jerárquica basada en las distancias entre las 14 muestras. B) Visualización de los dos primeros componentes principales del PCA efectuado sobre las 14 muestras.

Después de haber limpiado los datos de outliers, se generó un objeto tipo DESeq2 a partir de la matriz de expresión de todos los clústeres con ayuda de la función `DESeqDataSetFromMatrix` del paquete DESeq2 (Love et al., 2014). A continuación, se

removieron aquellos genes cuyos valores de expresión fueron menores a 10 en al menos 12 de los 13 clústeres restantes. Al objeto tipo DESeq2 resultante se le aplicó la función ``vsf`` del paquete DESeq2 (Love et al., 2014) para normalizar los valores de expresión (**Fig. 9**).

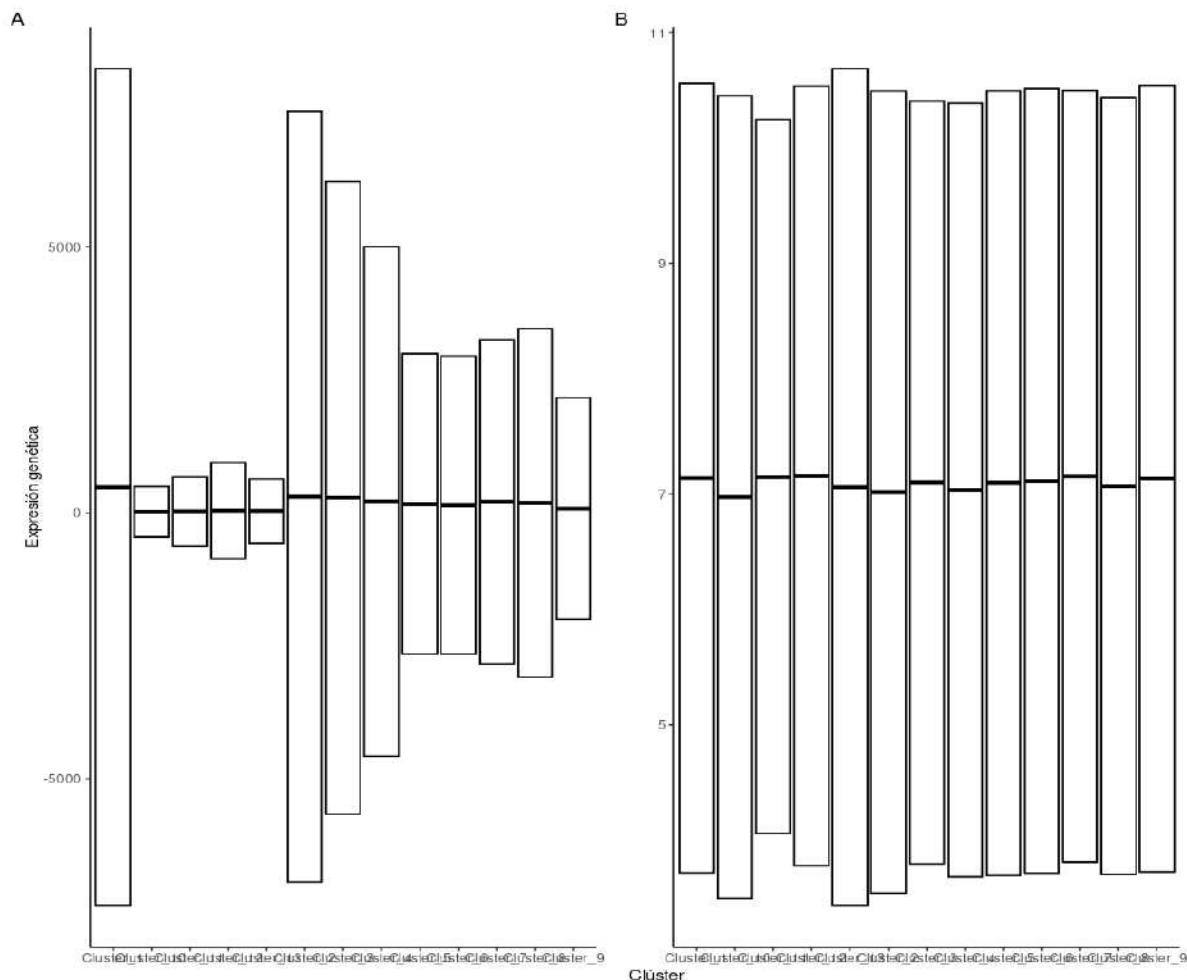


Figura 9. Normalización de los datos de expresión por clúster. A) Distribución de los datos de expresión genética en los 13 clústeres analizados. B) Distribución de los datos de expresión genética en los 13 clústeres analizados después de haber implementado la normalización.

4.9 Construcción de redes pesadas de co-expresión

Antes de iniciar la construcción de las redes de co-expresión, se estableció el valor específico al que se elevarían las correlaciones entre genes, con el fin de configurar una red de topología libre de escala (**Fig. 10**). Una vez seleccionado el valor umbral adecuado para alcanzar la topología deseada, se procedió a desarrollar la red pesada de co-expresión utilizando la función ``blockwiseModules`` del paquete WGCNA (Langfelder & Horvath, 2008). Paralelamente, en el curso de esta construcción, se identificaron módulos de genes con alta correlación a las características evaluadas y se determinaron sus respectivos eigengenes, los cuales corresponden al primer componente principal de cada módulo.

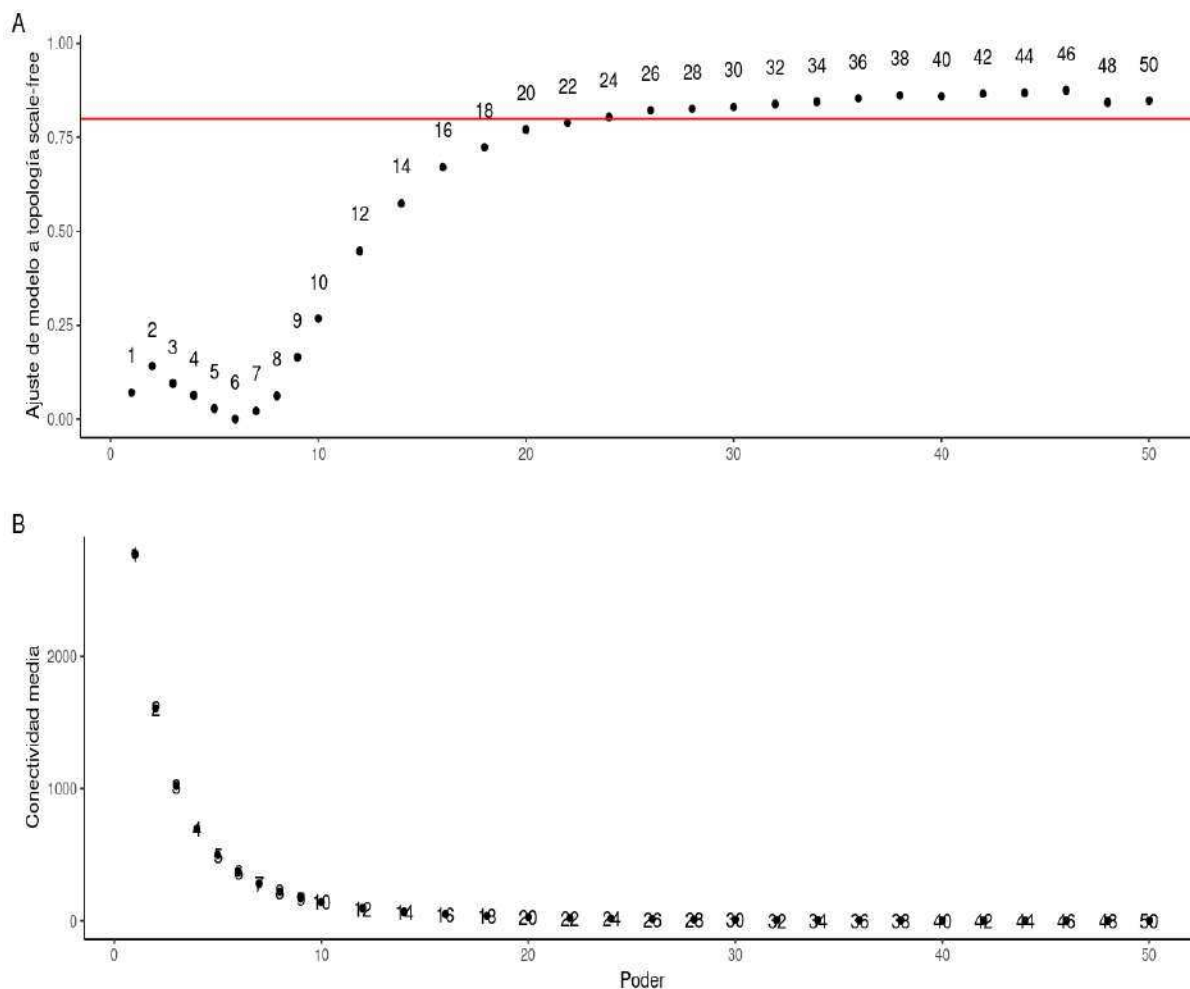


Figura 10. Valores umbrales. Valores umbrales para la construcción de las redes pesadas de co-expresión. A) Índice del ajuste de la red a una topología de escala libre dependiente del poder al que se elevan las correlaciones. B) Conectividad media entre nodos dependiente del poder al que se elevan las correlaciones.

4.10 Análisis de módulos asociados a cáncer y macrófagos

Después de obtener la red de co-expresión y los módulos de genes correspondientes, se llevó a cabo un análisis detallado de dichos módulos para explorar posibles vínculos con los tipos celulares específicos de cada clúster, y determinar si estos presentaban características cancerígenas. La evaluación de la asociación se basó en la significancia estadística derivada de las correlaciones entre los eigengenes de cada módulo y la condición cancerígena de cada tipo celular y clúster. Mediante el uso de la función `CorLevelPlot` del paquete CorLevelPlot (Blighe, 2024), se realizó una inspección visual que facilitó la identificación y elección de un clúster asociado a macrófagos y uno a células cancerosas. Los clústeres destacados para esta asociación fueron el cian y rosa, vinculados respectivamente con macrófagos y células cancerosas.

A continuación, se procedió a identificar los genes dentro de los módulos seleccionados para realizar un análisis de enriquecimiento genético. Este análisis se efectuó mediante herramientas del paquete `enrichR` (Jawaid, 2023), el cual proporciona acceso a la plataforma digital `Enrichr` (Kuleshov et al., 2016). Aunado a este proceso, se optó por utilizar el plugin `ClueGO` (Bindea et al., 2009), de la plataforma `Cytoscape`, para visualizar y agrupar los términos provenientes del enriquecimiento funcional de los genes del módulo rosa.

Para los módulos cian y rosa, se calcularon las Matrices de Traslape Topológico (TOM, por sus siglas en inglés) utilizando los datos de expresión génica de cada módulo a través de la función `TOMsimilarityFromExpr` del paquete `WGCNA` (Langfelder & Horvath, 2008).

Posteriormente, se utilizaron las matrices TOM de ambos módulos para generar representaciones en forma de objetos tipo `igraph`, mediante la función `graph_from_adjacency_matrix` del paquete `igraph` (Csárdi et al., 2024) (**Fig. 11** y **Fig. 12**). Finalmente, características topológicas como la centralidad de los nodos y la fuerza de cada nodo (peso total de sus enlaces), fueron evaluadas en cada red. Aquellos nodos de interés en la red del módulo rosa, fueron sujetos a una investigación bibliográfica en busca de posibles explicaciones de su relevancia topológica.

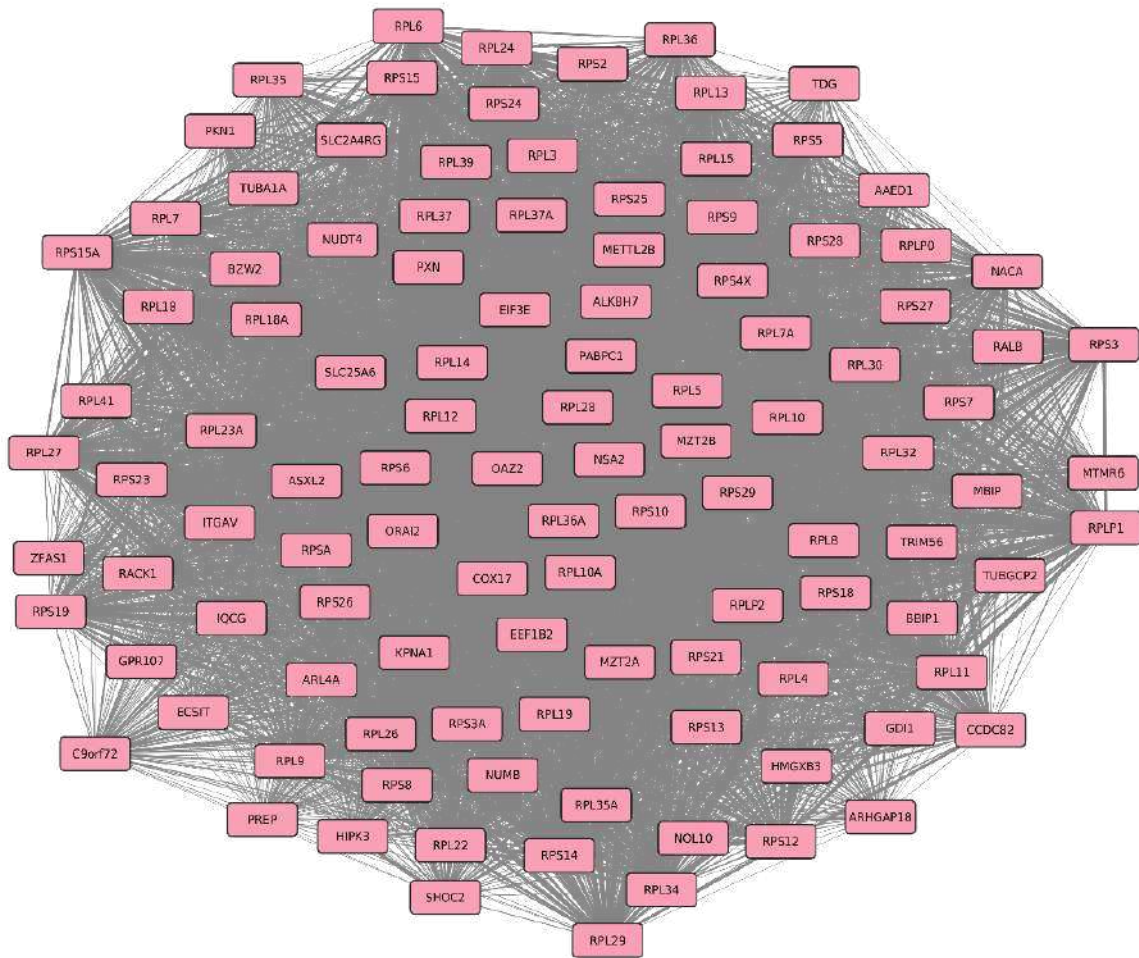


Figura 11. Red de módulo rosa. Representación gráfica de la red construida a partir de los genes del módulo rosa.

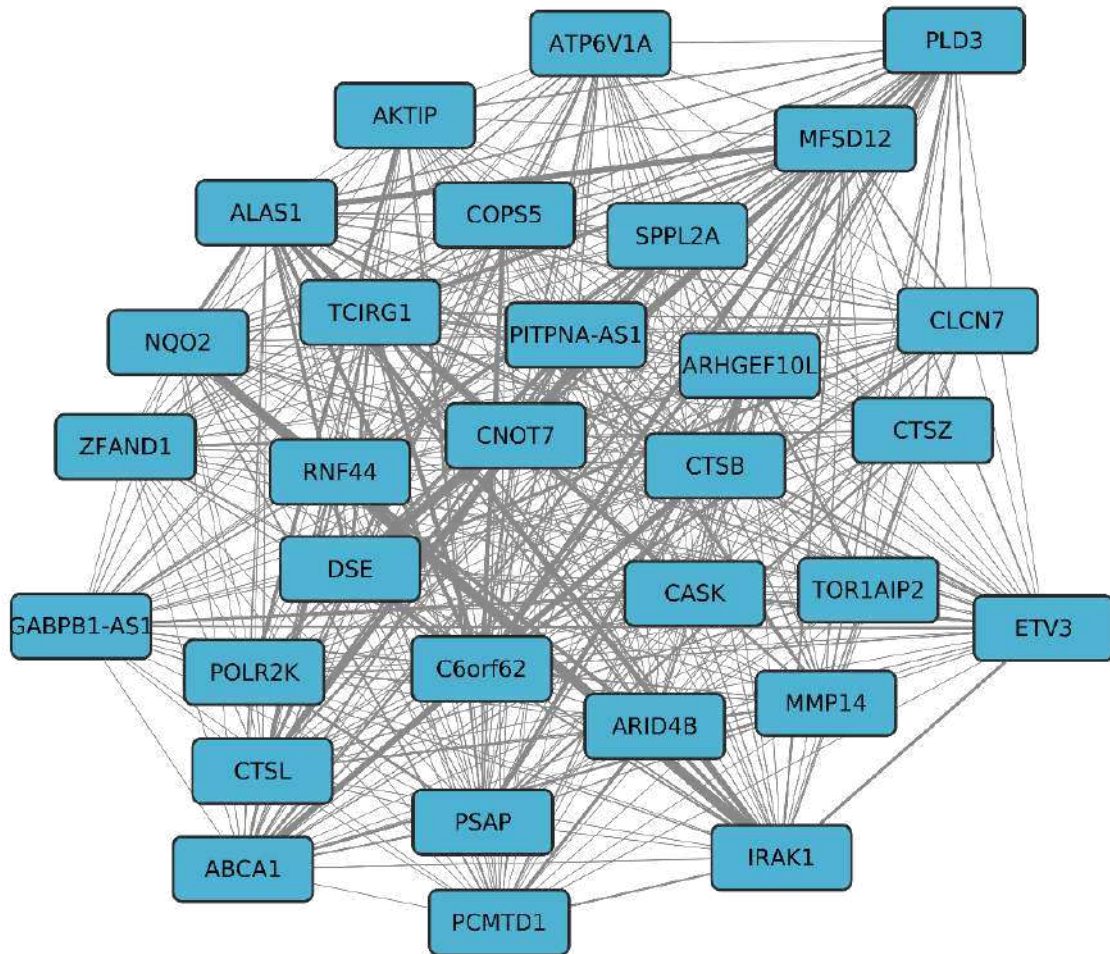


Figura 12. Red de módulo cian. Representación gráfica de la red construida a partir de los genes del módulo cian.

5. RESULTADOS

5.1 Creación de objeto Seurat

El filtrado inicial de los datos biológicos crudos permitió la elaboración de un objeto tipo Seurat preliminar. Este objeto únicamente contaba con una matriz dispersa de dimensiones 23,986 renglones x 6,054 columnas y los metadatos asociados a cada célula. Cabe aclarar que los ejes X y Y de la matriz representan todos los genes que pasaron los filtros impuestos y la totalidad de las células existentes respectivamente.

5.2 Control de calidad

Después de discriminar los datos con base en los criterios mencionados en la sección [4.2](#), un total de 929 muestras fueron descartadas tras no haber cumplido con uno o más de los

criterios establecidos (**Fig. 2**). Tras la remoción de las muestras seleccionadas, únicamente 5,125 células con sus respectivos metadatos fueron mantenidas dentro del objeto Seurat.

5.3 Procesamiento preliminar

5.3.1 Obtención de genes variables

Una vez efectuada la normalización de los datos, se identificaron los 2,000 genes con mayor variabilidad en sus perfiles de expresión (**Fig. 4**). En este caso, se puede apreciar que los 10 genes con mayor variabilidad son SERPINB2, CCL3, S100A9, CCL4, SCGB2A2, IL24, IL1B, MMP3, CXCL1 y APOE.

5.3.2 Reducción de dimensiones

Se calcularon los primeros 50 componentes principales de un PCA para efectuar una reducción dimensional. Tras evaluar la significancia estadística de cada componente mediante el análisis *jackstraw* y una inspección visual del *elbow plot* realizado, se optó por considerar los 50 componentes principales para el resto del trabajo debido a su aparente retención de información biológica (**Fig. 5**).

5.3.3 Identificación de comunidades preliminar

Tras haber calculado los k vecinos más próximos de los puntos en el espacio del PCA, la clusterización efectuada sobre los datos generó 13 clústeres. Posteriormente, se generó una visualización de los clústeres de células previamente generados al emplear y graficar la reducción dimensional UMAP (**Fig. 6**).

5.4 Remoción de dobletes

En cuanto a los dos métodos empleados para la detección de dobletes en las muestras respecta, ambos fueron capaces de designar presuntos dobletes; sin embargo, únicamente una muestra fue identificada como doblete por ambas técnicas (**Fig. 13**). El doblete en cuestión fue removido del resto del trabajo.

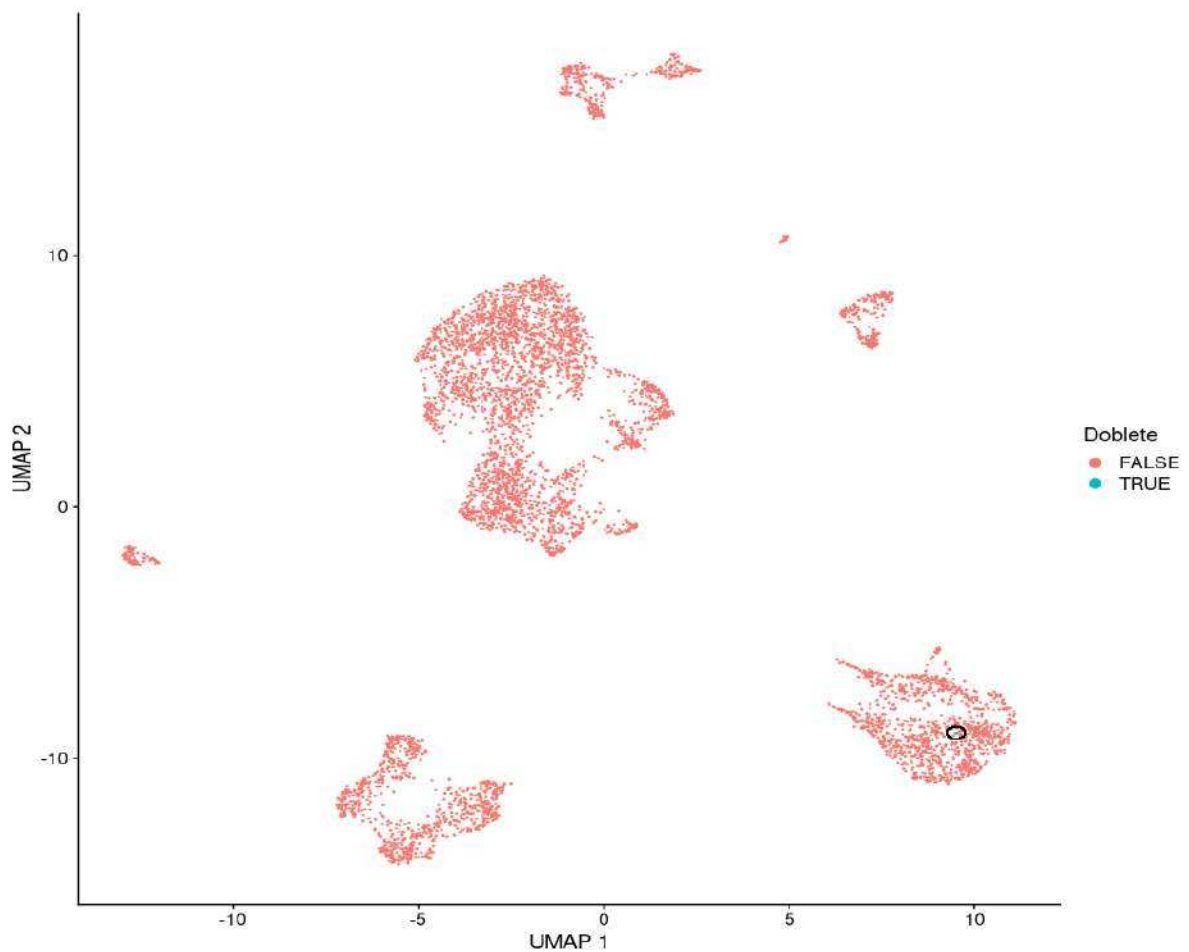


Figura 13. Remoción de doblete. Visualización de la única muestra designada como doblete. La muestra en cuestión se presenta encerrada en el círculo negro en la esquina inferior derecha.

5.5 Procesamiento subsecuente y asignación de tipos celulares

Al haber removido el único doblete identificado, el procesamiento descrito en la sección [4.5](#) dio como resultado el agrupamiento de las células en 14 clústeres diferentes (**Fig. 14** y **Fig. 15**), a diferencia de los 13 encontrados en el procesamiento preliminar (**Fig. 6**). En este caso, se encontraron siete tipos celulares: células ciliadas, células endoteliales, linfocitos, macrófagos, células de músculo liso, fibroblastos y células de epitelio no ciliado. A cada clúster le fue asignado alguno de los tipos celulares mencionados (**Fig. 15**). De manera particular, cabe resaltar que el grueso de las células fueron designadas como células no ciliadas y se ven agrupadas en un mismo espacio.

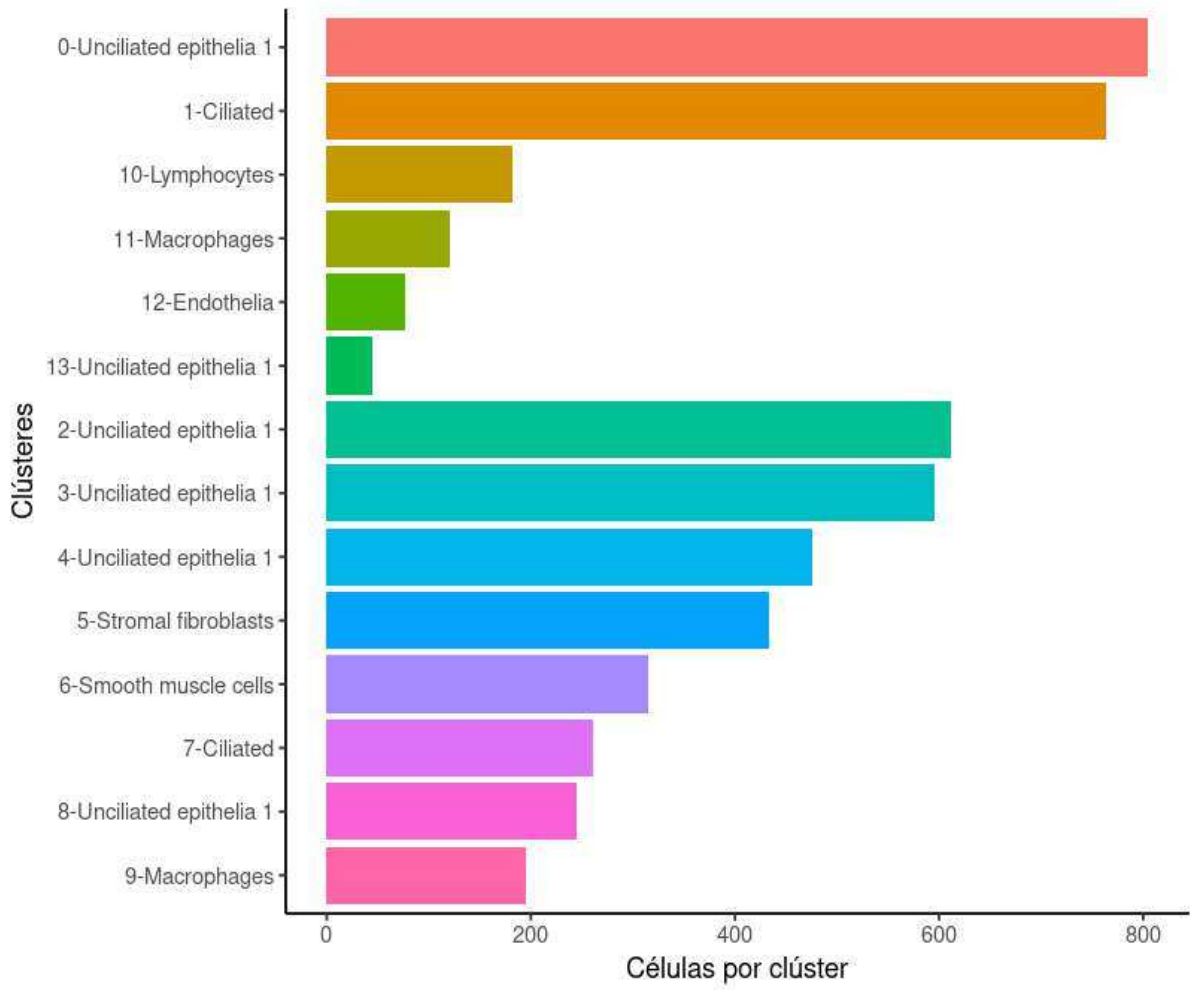


Figura 14. Células por tipo celular. Número de células asignadas a cada tipo celular.

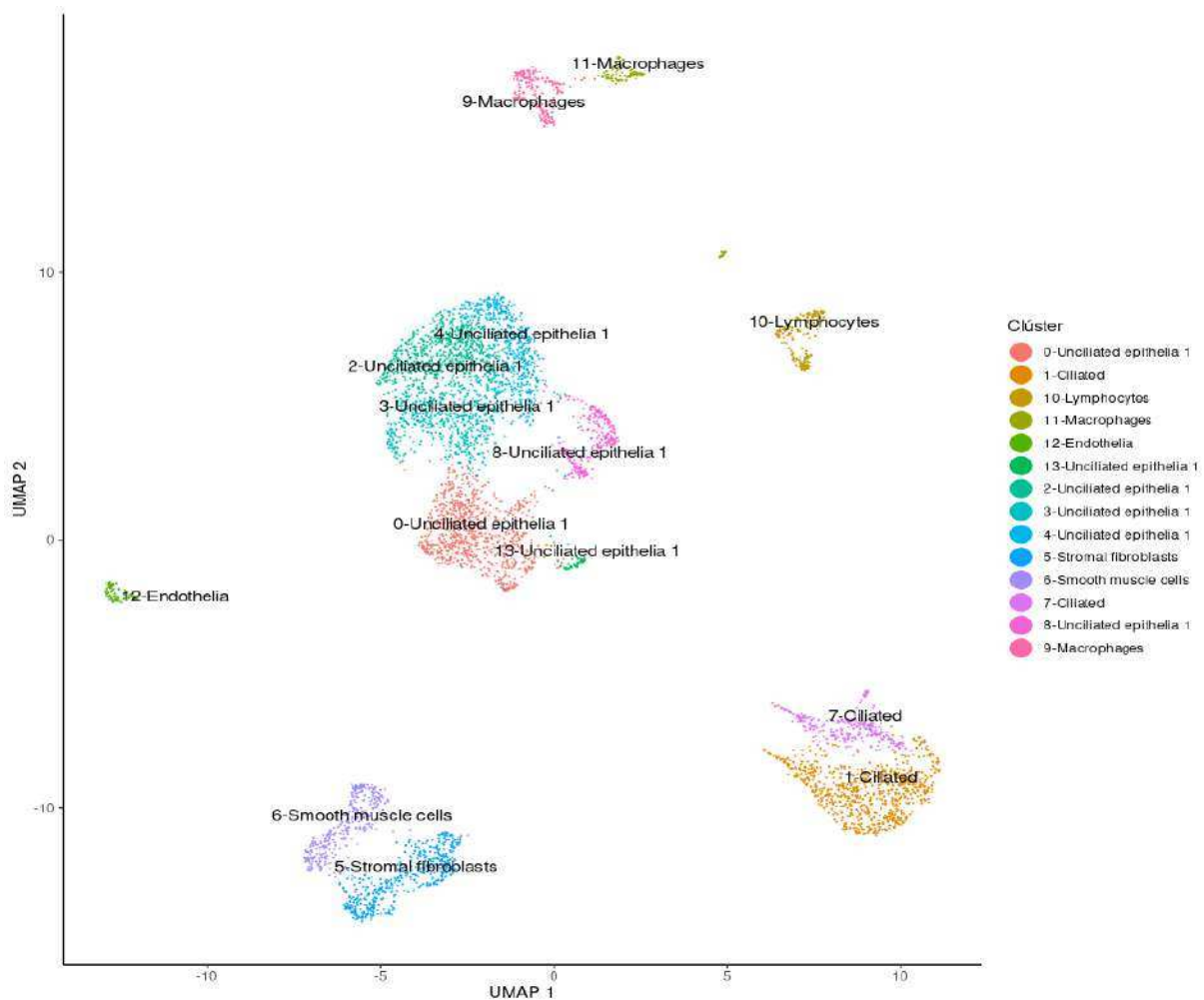


Figura 15. Visualización de clústeres. Representación gráfica de los 14 clústeres de células generados con los respectivos tipos celulares que les fueron asignados.

5.6 Identificación de células cancerosas

Las células pertenecientes a los clústeres 0, 1, 2, 3, 4, 7, 8 y 13 fueron designadas como cancerosas después de presentar niveles de expresión en los genes MUC16, WFDC2 y/o KIT igual o mayor al límite establecido en la sección 4.6 (Fig. 7); no obstante, tal como se mencionó anteriormente, ya que los clústeres 1 y 7 fueron designados como células ciliadas, estos fueron removidos de las muestras consideradas como cáncer (Fig. 16).

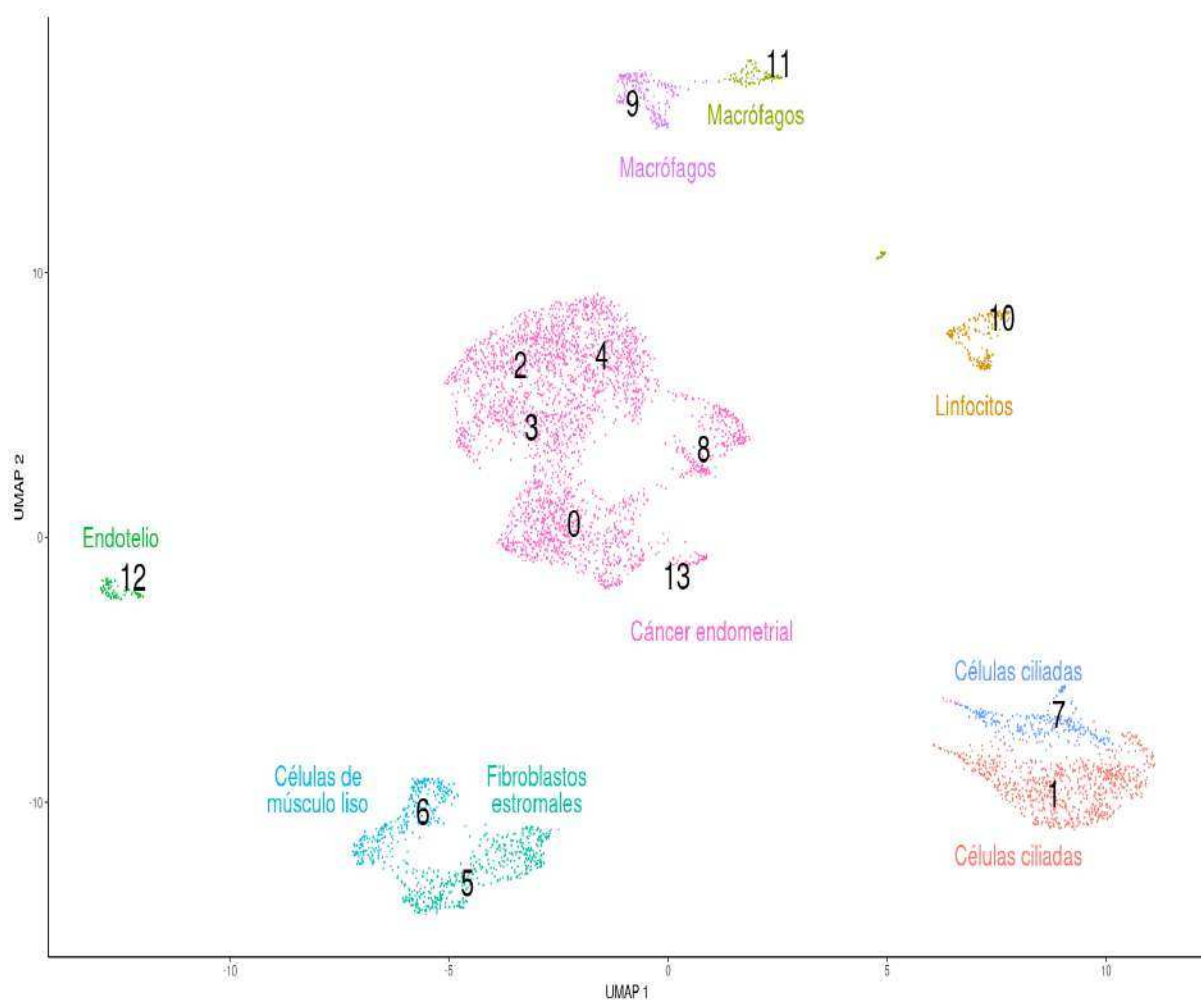


Figura 16. Tipos celulares. Representación de los clústeres designados como cancerosos y no cancerosos.

5.7 Pre-procesamiento para el algoritmo WGCNA

5.7.1 Identificación de outliers

Al observar los resultados de la clusterización jerárquica y una visualización de los primeros dos componentes del PCA efectuado sobre los datos, se designó al clúster 0 como un outlier y fue removido del resto del trabajo (**Fig. 8**).

5.7.2 Normalización de datos

Al evaluar los límites establecidos en los valores de expresión en la sección [4.8](#), 18,451 genes fueron descartados del trabajo. Posteriormente, se llevó a cabo la normalización de los datos pertenecientes a los 5,535 genes conservados a lo largo de las 13 muestras/clústeres (**Fig. 9**).

5.8 Construcción de redes pesadas

5.8.1 Selección de valores umbrales

Después de probar un rango de valores que abarcó desde 1 hasta 50, se optó por utilizar 26 como el valor al cual elevar las correlaciones entre los genes a fin de obtener una topología de escala libre (**Fig. 10**). Al mismo tiempo, la selección de este valor como el umbral al cual elevar las correlaciones, dio como resultado una conectividad media por nodo de 14.72.

5.8.2 Construcción de redes e identificación de módulos

Al efectuar la metodología descrita en la sección [4.9](#), se generó la red pesada de co-expresión para todos los genes en todas las muestras. Paralelamente, el análisis automático de la red identificó un total de 17 módulos de genes (**Fig. 17**). 16 de los módulos corresponden a genes altamente interconectados, mientras que el módulo restante fue formado a partir de los genes que no pudieron ser asignados a sus propios módulos.

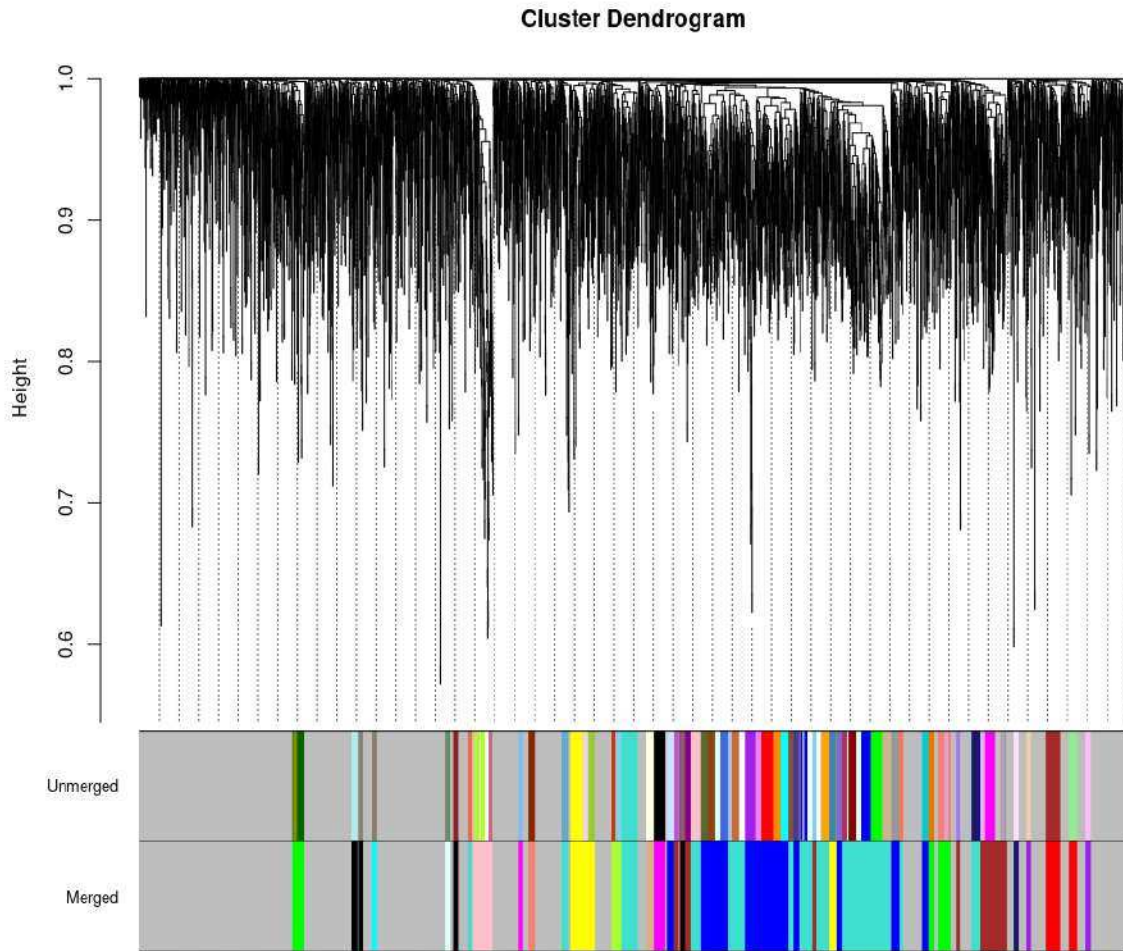


Figura 17. Dendrograma de módulos. Dendrograma representativo de la clusterización de genes a módulos. Aquellos grupos de genes con perfiles de expresión similares son unidos en módulos con colores distintos al gris.

5.9 Identificación de módulos asociados a cáncer

El cálculo de las correlaciones entre los eigengenes pertenecientes a los módulos y la condición cancerígena de cada tipo celular, permitió asociar seis módulos con el estado canceroso de las células (**Fig. 18**). Los módulos rosa y cian se encontraron positiva y significativamente correlacionados con el estado canceroso de las muestras; mientras tanto, los módulos canela, amarillo, cian claro y azul se encontraron negativa y significativamente correlacionados con este mismo estado. Por otro lado, los módulos cian, negro y amarillo-verde mostraron una correlación positiva a los macrófagos. Se tomaron los módulos rosa y cian, compuestos por 111 y 31 genes, para representar las células cancerosas y macrófagos respectivamente.

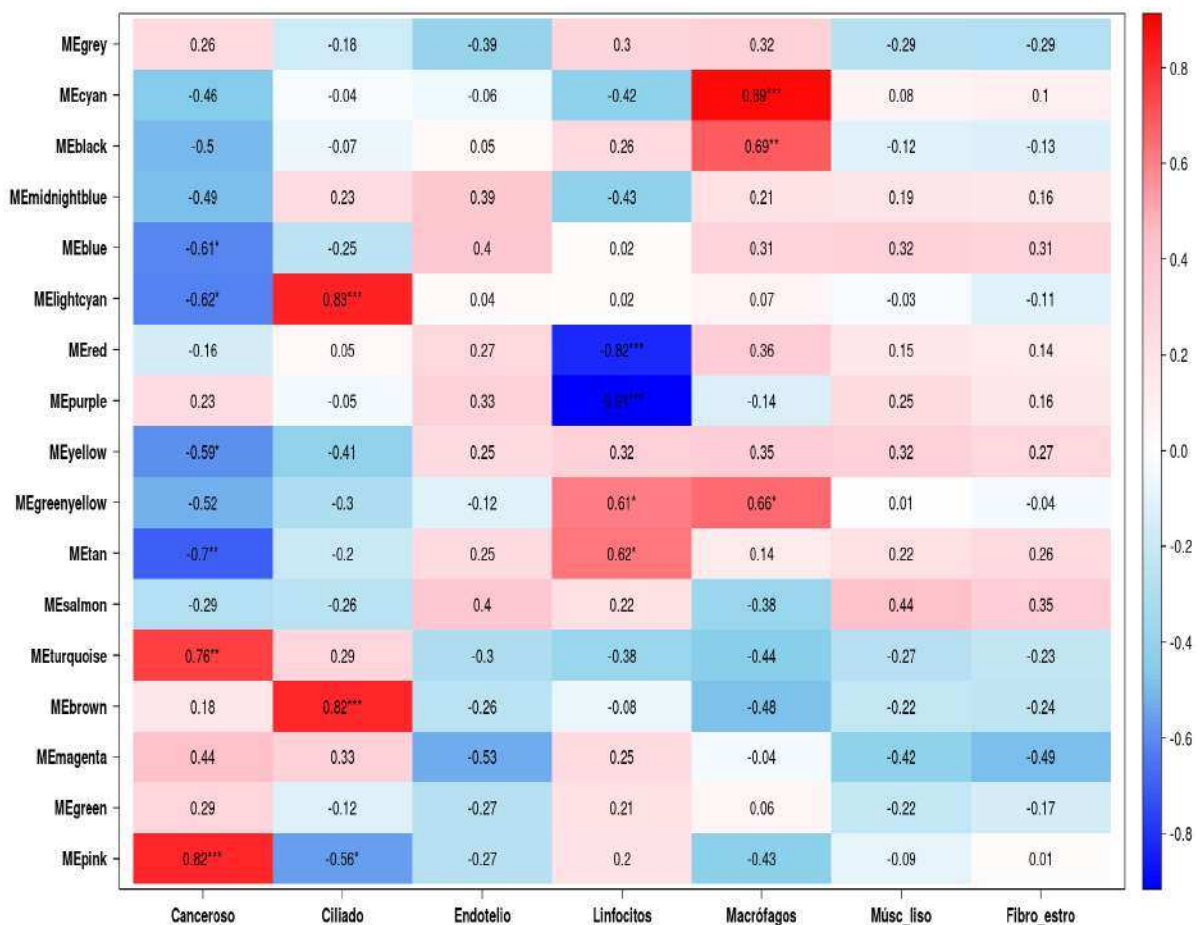


Figura 18. Correlación de módulos y características. Visualización de las correlaciones entre los eigengenes pertenecientes a los módulos y el estado cancerígeno y tipo celular de los datos analizados. El color rojo representa una correlación positiva entre el eigengen del módulo y la característica evaluada, mientras que el color azul apunta a una correlación negativa. Los asteriscos señalan cuáles correlaciones son estadísticamente significativas y el grado de esta significancia: * <0.5 , ** <0.01 , *** <0.001 .

5.10 Enriquecimiento funcional

Al efectuar el análisis de enriquecimiento funcional, se encontraron un total de 545 términos enriquecidos para los genes del clúster rosa y 240 para el clúster cian. De manera particular, se destaca la presencia de términos como GO:002261, GO:0042254, GO:0042274, GO:0006364, GO:0042255, GO:0042273, GO:0022618, GO:0000028, GO:0000027, GO:0000462 y GO:0030490 asociados con los genes del módulo rosa (**Fig. 19A**), pues están relacionados con la biogénesis ribosomal. De igual manera, los términos GO:0002181, GO:0043043, GO:0006412, GO:0010467, GO:0034470 y GO:0006417, igualmente provenientes del análisis efectuado sobre los genes del módulo rosa, resaltan por su

aparente participación en la regulación positiva de procesos como la transcripción y traducción de proteínas. Aunado a estos resultados, la visualización de los términos enriquecidos para los genes del módulo rosa con el plugin ClueGO, muestra un agrupamiento de términos relacionados con la inhibición y regulación de las ubiquitininas ligasas y su actividad como transferasas (**Fig. 20**).

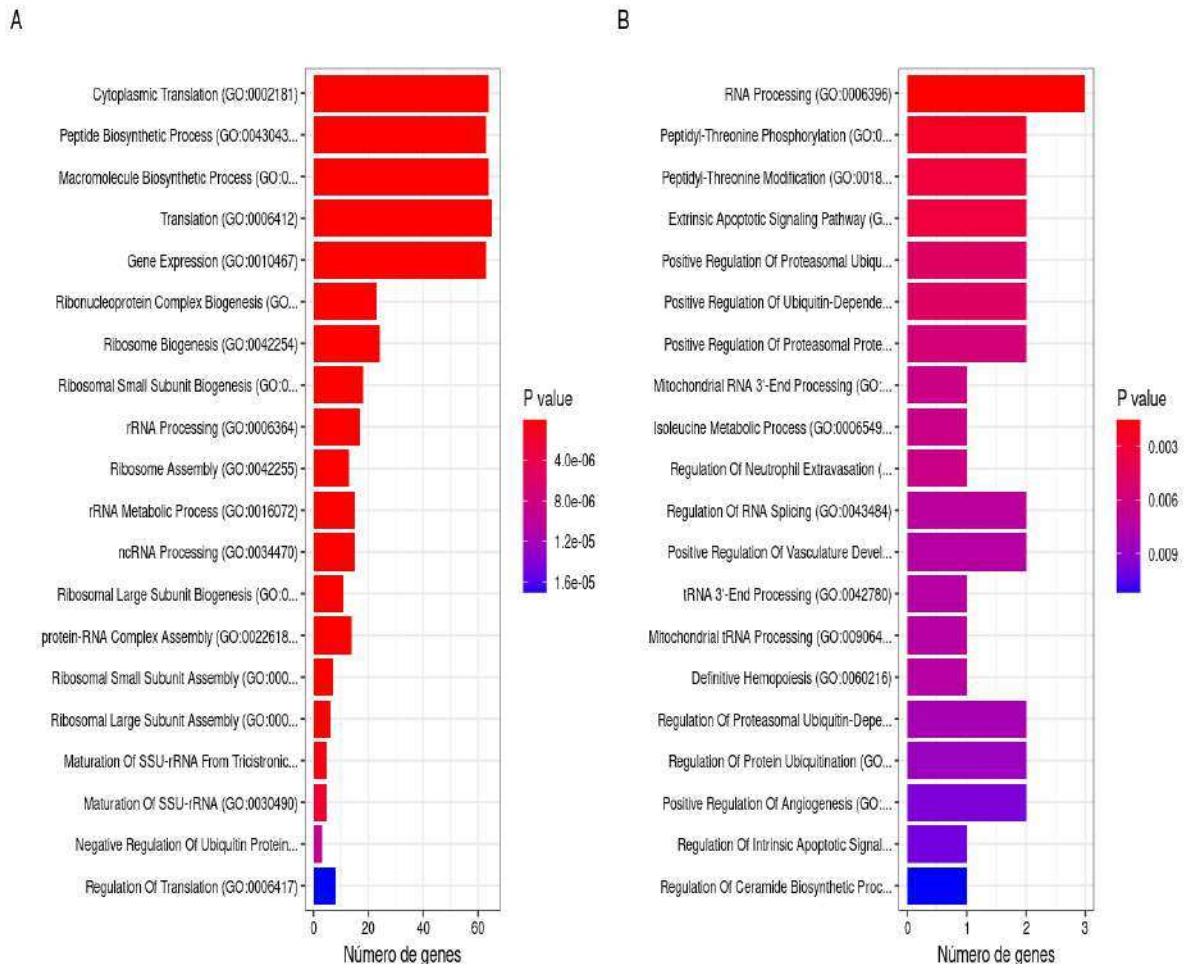


Figura 19. Resultados del análisis de enriquecimiento funcional. A) Términos enriquecidos asociados a las funciones moleculares de los genes pertenecientes al módulo rosa. B) Términos enriquecidos asociados a las funciones moleculares de los genes pertenecientes al módulo cian.

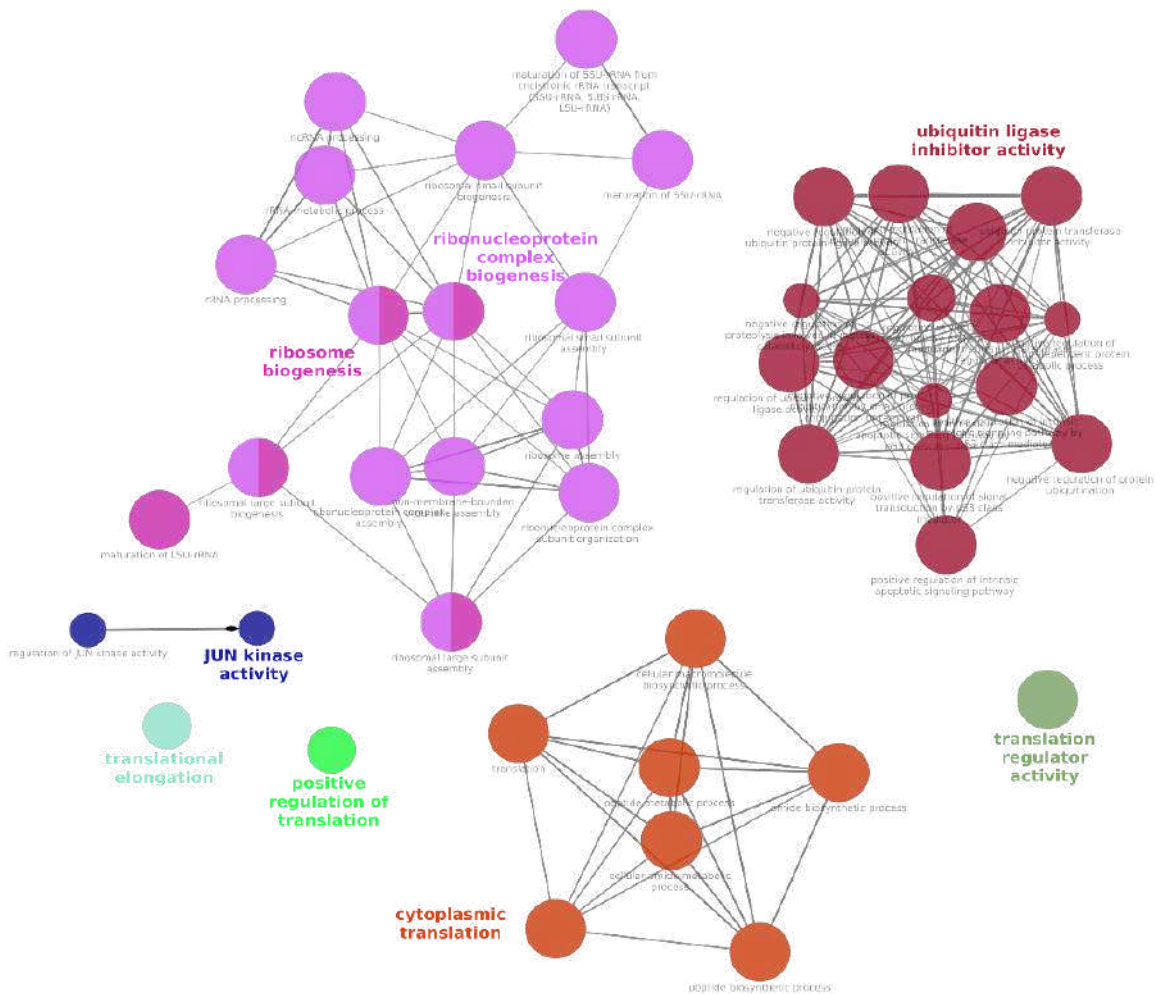


Figura 20. Agrupamiento y visualización de los términos enriquecidos provenientes del análisis de los genes del módulo rosa. Se muestran un total de 7 clústeres formados por los términos funcionales más abundantes asociados con los genes del módulo rosa. Los tres clústeres más extensos se ven relacionados con la biogénesis ribosomal. la actividad inhibitoria sobre las ubiquitinas ligasas y la traducción proteica a nivel citoplasmático.

Por otro lado, los términos GO:1904018, GO:0060216 y GO:0045766, encontrados durante el análisis de los genes del módulo cian (**Fig. 19B**), señalan la posible participación de los macrófagos encontrados en procesos angiogénicos. Asimismo, GO:0060216, GO:1902253, GO:2000389, GO:0097191, entre otros términos, indican una probable participación de los macrófagos presentes en mecanismos asociados con la proliferación celular y el control de eventos apoptóticos.

5.11 Análisis topológico y estructural de las redes

En cuanto a las propiedades topológicas de la red del módulo rosa respecta, los datos obtenidos señalan a los genes codificantes de proteínas ribosomales como componentes importantes para la estructura de dicha red; sin embargo, también demuestran posibles

interacciones de genes no ribosomales que afectan a la misma. En particular, es de notar que, de los 50 nodos con mayor fuerza en la red, únicamente los genes NACA y EEF1B2 no codifican proteínas ribosomales. Por otro lado, la centralidad según el número de veces que cada nodo aparece entre los caminos más cortos de cada pareja de nodos (*betweenness*), señala únicamente a los genes HMGXB3, ALKBH7, GPR107, SLC25A6, OAZ2, COX17, TDG y ZFAS1, como nodos importantes dentro de esta medida de centralidad. Finalmente, al evaluar la cercanía de cada nodo con todas las posibles parejas de elementos restantes (*closeness*), se observó que, de los 50 genes con mayor *closeness*, únicamente 16 de estos nodos codifican para proteínas ribosomales; de igual manera, cabe resaltar que los genes que presentaron valores de *betweenness* también se encontraron dentro de los 50 genes con mayor *closeness*.

Respecto a las características de la red del módulo cian, los genes TOR1AIP2, PCMTD1, ARHGEF10L, POLR2K y CASK representan los únicos nodos con valores de *betweenness* y los primeros cinco genes con un mayor índice de *closeness*. En cuanto a la fuerza de los nodos, los genes anteriormente mencionados no figuran dentro de los 15 nodos con mayor fuerza; mientras tanto, los genes C6orf62, ALAS1, DSE, MFSD12, TCIRG1, IRAK1, PLD3, CNOT7, NQO2 y ABCA1 representan los 10 genes con más fuerza.

6. DISCUSIÓN

El cáncer endometrial, al igual que los demás tipos de cáncer, es el resultado de una proliferación celular descontrolada con un amplio rango de características moleculares y celulares que contribuyen a la heterogeneidad tumoral (Dagogo-Jack & Shaw, 2017; McGranahan & Swanton, 2017; PDQ® Adult Treatment Editorial Board, 2024; Vitale et al., 2021). Dicha heterogeneidad es capaz de generar complicaciones en el diagnóstico y tratamiento de este padecimiento (Dagogo-Jack & Shaw, 2017; McGranahan & Swanton, 2017; Vitale et al., 2021), destacando la necesidad de comprender a mayor profundidad los mecanismos que impulsan esta neoplasia así como a la heterogeneidad en sí misma.

En este trabajo se realizó el análisis de datos de scRNA-seq provenientes de una muestra de tumor primario de cáncer endometrial, con el objetivo de identificar algunas de las características genéticas asociadas a un fenotipo canceroso. Con este propósito en mente, se llevó a cabo la identificación de distintos tipos celulares y grupos de células cancerosas presentes en el tumor, se elaboraron redes pesadas de co-expresión, se analizaron algunos de los genes asociados al fenotipo de interés y se evaluaron propiedades topológicas de las redes obtenidas. Cabe recalcar que, si bien los resultados obtenidos señalan la presencia e interacción de algunos genes con el estado canceroso de las células, algunos de los

métodos empleados durante este trabajo han sido puntos de contención en la comunidad científica.

En primera instancia, todo trabajo que haga uso de datos de scRNA-seq requiere de un adecuado tratamiento de los datos crudos para asegurar su calidad; sin embargo, dependiendo de la muestra, métodos y técnicas empleadas, e incluso del criterio de los investigadores, este control de calidad puede variar (Hippen et al., 2021; Jiang, 2019; Osorio & Cai, 2020). Si bien, la utilización de MADs en este trabajo permite la obtención de medidas objetivas, mensurables, reproducibles y resistentes a la influencia de datos con la capacidad de sesgar el procesamiento (Leys et al., 2013; Wilcox, 2003), este método no es infalible. La consideración de parámetros tales como el número de MADs empleadas, la identificación bidireccional de outliers, las características genéticas evaluadas, entre otros, pueden tener efectos considerables sobre el número de muestras con las que se trabaja, la detección de artefactos de secuenciación o incluso la remoción de muestras de interés (Cormio et al., 2015; Kunes et al., 2023; Musicco et al., 2018).

Aunado al punto anteriormente mencionado, algunos otros pasos en el procesamiento de datos de scRNA-seq generan debate entre la comunidad científica (Kolodziejczyk et al., 2015; Svensson et al., 2017). Estos puntos de contención abarcan una variedad de temas como la selección de lenguaje de programación, paquetes informáticos, métodos para calcular los genes con mayor variabilidad, identificación de tipos celulares, etc. (Luecken & Theis, 2019; Pullin & McCarthy, 2024); no obstante, uno de los temas más discutidos actualmente es la utilización y selección de técnicas de reducción dimensional (Imoto et al., 2022; Kobak & Linderman, 2021; Shinn, 2023). En este sentido, la implementación de PCAs como una herramienta base para una subsecuente clusterización de las muestras evaluadas y visualización por medio de otras técnicas de reducción dimensional, se ha estandarizado en los flujos de trabajo para datos de scRNA-seq (Kobak & Linderman, 2021; Luecken & Theis, 2019; Tsuyuzaki et al., 2020); sin embargo, diferentes investigaciones han demostrado que, características intrínsecas a los datos biológicos y a los algoritmos utilizados, pueden generar resultados inconsistentes que afectan gravemente resultados como la clusterización de las muestras (Draganov et al., 2023; Imoto et al., 2022; Kiselev et al., 2019; Xiang et al., 2021).

Ahora bien, aún tomando en consideración los puntos anteriormente mencionados que tienen la capacidad de afectar los resultados, es importante reconocer la información obtenida a través de los análisis realizados. Particularmente, los resultados obtenidos de la identificación de tipos celulares habidos en la muestra biológica analizada señalan la presencia de células ciliadas, células endoteliales, linfocitos, macrófagos, células de

músculo liso, fibroblastos y células de epitelio no ciliado (**Fig. 15**); lo cual, al considerar el origen de la muestra, es algo esperado.

En primer lugar, la presencia de linfocitos no es sorprendente debido al rol que estas células llevan a cabo en procesos de control y progresión en los tumores (Brummel et al., 2023; Gooden et al., 2011); sin embargo, cabe recalcar que estas células inmunes también se encuentran de forma natural en esta zona pese a los cambios habidos post-menopausia (Agostinis et al., 2019; Rodriguez-Garcia et al., 2021). De manera muy similar, los macrófagos son otro de tipo de célula inmune comúnmente encontrada dentro de tumores cancerígenos y tejidos endometriales sanos (Agostinis et al., 2019; Pan et al., 2020); no obstante, los cambios en las poblaciones de macrófagos en tejido endometrial de mujeres postmenopáusicas no ha sido estudiada a profundidad (Rodriguez-Garcia et al., 2021). En el caso de las células de músculo liso, su presencia se puede deber simplemente a la propia fisiología del endometrio, pues este tipo de célula se encuentra naturalmente en el miometrio y se encarga de generar los movimientos de contracción durante el parto (Arrowsmith et al., 2018). A continuación, la identificación de células endoteliales y fibroblastos puede deberse a las características fisiológicas intrínsecas del endometrio al igual que las células de músculo liso (Arrowsmith et al., 2018; Iruela-Arispe et al., 1999); sin embargo, al considerar los cambios generados por la edad en estas células y la capacidad tumoral de inducir su propia vascularización y activación de células con habilidades proliferativas (McKinnon et al., 2022; Pan et al., 2020), el origen de estas poblaciones termina siendo un poco ambiguo. Finalmente, la identificación de células ciliadas y no ciliadas dentro de la muestra biológica analizada genera un poco de confusión, pues aunque su presencia puede ser parcialmente explicada por condiciones naturales en el tejido endometrial (Maenhoudt et al., 2022; Tian et al., 2023), la posterior reasignación de estas células a células cancerosas hace hincapié a un origen diferente.

Retomando la presencia de las células ciliadas y no ciliadas en la muestra analizada, es de importancia resaltar el hecho de que estos tipos celulares se presentan en tejidos sanos de forma natural ya que tienen una variedad de funciones secretoras y estructurales, esto sin mencionar que sus poblaciones aumentan conforme la edad (Horree et al., 2007; Loid et al., 2023; Maenhoudt et al., 2022); sin embargo, también es sabido que desregulaciones genéticas asociadas a la generación de metaplasias y cáncer, también tienden a influenciar las poblaciones de estas células (Horree et al., 2007; Klaes et al., 2001; Loid et al., 2023). Tomando esto en cuenta, se optó por utilizar los mismos marcadores descritos en el trabajo de Regner y colaboradores (2021) para identificar células cancerosas (**Fig. 7**); no obstante, pese a la coincidencia en resultados obtenidos, se considera que estos marcadores podrían ser complementados. En particular, al observar los bajos niveles de expresión del gen KIT

en tan solo 63 células (**Fig. 7C**), la idea de que algún clúster haya sido erróneamente asignado como canceroso, toma fuerza; en especial en un caso como el clúster tres, pues únicamente fue designado como canceroso gracias a los niveles de expresión del gen WFDC2. Es por esto, y considerando que el gen WFDC2 se expresa de forma natural en tejido ciliado (Galvano et al., 2006), que los clústeres asignados como células ciliadas, no fueron consideradas como cáncer para los análisis subsecuentes realizados en este trabajo. Para combatir esta problemática, la implementación de otros biomarcadores como los genes p16, p53 o Bmi-1, los cuales se han asociado con cáncer endometrial (Arend et al., 2018; Banno et al., 2012; Klaes et al., 2001; Nam et al., 2014; Xu et al., 2022), podrían ser evaluados junto con los tres genes utilizados en este trabajo para ofrecer una mayor certeza en la identificación de células cancerígenas.

Una vez que todos los tipos celulares fueron designados y se implementó el flujo de trabajo descrito en la documentación del paquete WGCNA (Langfelder & Horvath, 2008), se optó por analizar un módulo asociado con células cancerosas y otro a macrófagos. Esta decisión se tomó al observar que las correlaciones de los eigengenes pertenecientes a cada módulo no mostraron una correlación significativa con células endoteliales, fibroblastos o incluso células de músculo liso; sin embargo, los resultados obtenidos fueron adecuados para seguir el análisis.

El análisis de enriquecimiento funcional de los genes pertenecientes a los módulos rosa y cian, permiten entender un poco acerca del funcionamiento detrás de las células cancerosas comparadas con macrófagos. En particular, es de importancia recalcar algunos de los términos presentes en el análisis efectuado sobre los genes pertenecientes al módulo rosa, pues varios de ellos describen procesos biológicos que participan en el desarrollo y mantenimiento de tumores cancerígenos (**Fig. 19A** y **Fig. 20**). En primera instancia, una cantidad de términos enriquecidos pueden agruparse en dos funciones biológicas relacionadas: biogénesis ribosomal y la biogénesis de complejos ribonucleoproteicos (**Fig. 20**). La presencia de estos términos asociados con los genes del módulo rosa son significativos, pues investigaciones previas han apuntado a las desregulaciones en la biogénesis ribosomal como un factor de importancia para el desarrollo, y posible factor causal, de distintos tipos de cáncer (Elhamamsy et al., 2022; Mourksi et al., 2023; Pelletier et al., 2018; Yu et al., 2020; Zou et al., 2020). De manera puntual, las desregulaciones en la biogénesis ribosomal se han asociado con un incremento en la inestabilidad del genoma, una alza en la tasa de síntesis proteica que permite la proliferación de células cancerígenas, la facilitación de procesos metastásicos y la resistencia a tratamientos, entre otras características (Elhamamsy et al., 2022; Pelletier et al., 2018); lo cual es coincide con la determinación de los genes de este módulo al fenotipo cancerígeno de las células. De

manera similar, el agrupamiento de términos relacionados con la supresión de la actividad de las proteínas ubiquitina ligasas (**Fig. 20**), es un resultado que refuerza la posible participación de los genes del módulo rosa en el desarrollo del cáncer en esta muestra. De manera específica, es necesario recalcar que la ubiquitinación es una modificación postraduccional de las proteínas necesaria para la inducción de respuestas celulares ante estímulos específicos, sean benéficos o perjudiciales, y la homeostasis proteica dentro de las células (Guo et al., 2024; Lee & Gu, 2010), por lo que una desregulación en este mecanismo invariablemente genera repercusiones en las células afectadas. Ahora bien, resulta de gran interés mencionar que los términos asociados con la inhibición de las ubiquitina ligasas, igualmente están relacionados con la inhibición de la ubiquitinación de la proteína p53, la cual, al no ser estable, es incapaz de inducir los mecanismos celulares que conllevan a la apoptosis, de modo que permite el desarrollo de las células cancerígenas. Asimismo, el último conjunto de gran tamaño en los términos enriquecidos agrupados, pone en evidencia un incremento en la tasa de traducción proteica dentro de las células (**Fig. 20**). Este hallazgo supone aún más evidencia a favor de la consideración de los genes del módulo rosa como posibles agentes en el desarrollo del cáncer endometrial, pues la actividad traduccional elevada es un marcador importante para la detección y pronóstico del cáncer como grupo de enfermedades (Robichaud et al., 2019). Aunado a este hecho, si se considera la desregulación en la biogénesis ribosomal previamente mencionada, junto con una actividad traduccional elevada, resulta muy probable considerar que estos dos hallazgos se encuentran estrechamente relacionados entre sí y con el fenotipo cancerígeno en cuestión. Finalmente, los términos relacionados con la regulación de quinasa JUN son explicables en igual medida por el cáncer, esto debido a que las rutas de señalización en las que esta quinasa se ve involucrada regulan la apoptosis y pueden modular positivamente la proliferación celular (Grynberg et al., 2017; Hammouda et al., 2020; Vaiopoulos et al., 2010).

En contraste, los resultados del análisis de enriquecimiento funcional de los genes del módulo cian (**Fig. 19B**), facilitan la identificación de mecanismos llevados a cabo por macrófagos bajo estrés. Por ejemplo, es sabido que los macrófagos asociados a tumores son capaces de limitar el reclutamiento de células T al sitio afectado, lo cual ayuda a la evasión inmune de las células cancerígenas y permite su continuo desarrollo (DeNardo & Ruffell, 2019; Pan et al., 2020); sin embargo, este mecanismo no es suficiente para determinar un secuestro de estas células inmunes por parte de las células cancerígenas. Por otro lado, los términos GO:1904018, GO:0060216 y GO:0045766, los cuales se asocian con eventos angiogénicos, sí ofrecen más de evidencia del secuestro y reprogramación de los macrófagos. Este hecho se debe a que el cáncer es capaz de forzar a los macrófagos a inducir la angiogénesis cerca de los tumores, pues al tener elevadas tasas de replicación,

las células cancerígenas requieren de una gran cantidad de nutrientes que de otra manera no están disponibles en las cantidades que estas células necesitan para seguir proliferando (Fu et al., 2020; Riabov et al., 2014). Finalmente, existen una variedad de términos como GO:0060216, GO:1902253, GO:2000389 y GO:0097191, que abarcan mecanismos indicativos de un posible secuestro de la maquinaria celular de los macrófagos, pues cubren funciones como la regulación de p53 durante la apoptosis, la extravasación de neutrófilos, la inducción de células madre hematopoyéticas, etc.

Finalmente, durante la evaluación de las redes producidas para los genes de los módulos rosa y cian, se obtuvieron resultados congruentes con investigaciones previas y se dilucidaron genes previamente no estudiados bajo el contexto de cáncer endometrial. Por una parte, al analizar la fuerza de los nodos en la red del módulo rosa, resulta evidente que los genes codificantes para proteínas ribosomales son fundamentales para la estructura de la red, pues son los genes cuyos enlaces tienen más fuerza en toda la red. Este resultado podría indicar que los genes codificantes de proteínas ribosomales están dirigiendo parte del desarrollo y proliferación del cáncer; sin embargo, es imposible determinar con los resultados obtenidos si estos genes son agentes causales o parte del desarrollo intrínseco al cáncer mismo. En este mismo sentido, aunque los genes NACA y EEF1B2 resaltan por estar dentro de los 50 genes con mayor fuerza en el módulo, no representan resultados inesperados al investigar más a fondo las funciones que llevan a cabo. Por su parte, NACA codifica para una proteína que se liga al factor de transcripción BTF3, el cual a su vez puede iniciar la transcripción de CHD1L (oncogen capaz de inducir resistencia a diversos tratamientos), o bien, puede actuar como un posible agente inhibitorio de ubiquitina ligasa E3 (Fan et al., 2021; H. Wang et al., 2021). Por otro lado, EEF1B2 es un factor de elongación traduccional que se ha reportado con un posible rol en la tumorigénesis debido a su función en el metabolismo del piruvato, la oxidación fosforilativa y su asociación con los ribosomas (Jia et al., 2019). En cuanto al índice de *closeness* y *betweenness* de cada nodo, los valores elevados para estas medidas en los genes previamente mencionados, pueden indicar la presencia de un mecanismo subsecuente en las distintas cascadas de señalización reguladas por el cáncer. Al considerar que los mismos nodos poseen valores en ambos índices, es natural considerar que estos genes pueden actuar como puentes para propagar la señal a través de la red, incluso a nodos que se encuentran “alejados” de otros nodos que inducen señales más fuertes (aquellos nodos con más fuerza), por lo que un estudio más detallado de las funciones de estos genes, y de las cascadas de señalización en las que participan, podría dilucidar más información.

Para concluir, el análisis de la red generada con los genes pertenecientes al módulo cian, muestra a los genes TOR1AIP2, PCMTD1, ARHGEF10L, POLR2K y CASK como posibles

genes responsables de la propagación de señales celulares inducidas por células cancerígenas. En particular, los genes CASK y ARHGEF10L se han asociado con proliferación celular y tumorigénesis (Qu et al., 2021; D.-W. Wang et al., 2020), de modo que un estudio más detallado de su rol en los macrófagos bajo un estrés cancerígeno podría ser de utilidad. Finalmente, los nodos con mayor fuerza en la red posiblemente tengan roles asociados a la supervivencia de las células cancerígenas, ya sea por la inhibición de mecanismos capaces de dañarlas, o bien, al inducir mecanismos benéficos para el desarrollo del cáncer.

7. CONCLUSIONES

En este trabajo, se analizaron datos de secuenciación provenientes de la aplicación de scRNA-seq en una muestra de tumor primario de cáncer endometrial, con el fin de identificar patrones de expresión génica asociados a esta enfermedad a través de la interpretación de los resultados obtenidos de un análisis de redes pesadas de co-expresión. El análisis, el cual implicó la identificación de tipos celulares, la construcción de las redes pesadas de co-expresión, el análisis de enriquecimiento funcional de genes específicos y el análisis de redes asociadas a los fenotipos de interés, dio como resultado información que respalda una gran variedad de investigaciones efectuadas hasta el momento.

Específicamente, los resultados obtenidos del análisis de enriquecimiento funcional efectuado sobre los genes pertenecientes a los módulos rosa y cian (**Fig. 19**), indican la presencia de genes involucrados en procesos biológicos anteriormente relacionados con mecanismos y respuestas propias del organismo al cáncer endometrial. En particular, los procesos biológicos señalados en el análisis de enriquecimiento funcional de los genes del módulo rosa, muestran concordancia con investigaciones precedentes en las que se detalla la habilidad de las células cancerosas para manipular dichos procesos, los cuales incluyen el secuestro de la maquinaria metabólica, la estabilidad del material genético, la inducción de eventos proliferativos y el desarrollo y manutención general de tumores sólidos en cáncer. Por otro lado, el mismo análisis efectuado sobre los genes asociados con macrófagos, también son acordes a la literatura existente. Más concretamente, los resultados obtenidos confirman la existencia y acción de algunos mecanismos presentes en los macrófagos, los cuales son inducidos y manipulados por estímulos propios del cáncer.

Finalmente, el análisis realizado sobre la red del módulo rosa, ofrece información que refuerza conocimientos previos y apunta a posibles investigaciones. Particularmente, una variedad de genes previamente asociados con este cáncer fueron identificados y agrupados en la red, lo cual es de esperarse; sin embargo, también se presentan genes cuya función

ha sido menos estudiada bajo el contexto de cascadas de señalización generadas por estímulos propios del cáncer. Es por esto que un análisis más profundo en el que se analicen una mayor cantidad de genes asociados a este tipo de cáncer, pudiera ofrecer aún más información respecto a esta enfermedad y los mecanismos por los cuales sobrevive y prolifera.

8. PERSPECTIVAS

El trabajo realizado engloba distintos análisis para detectar la posible influencia de algunos genes sobre el cáncer endometrial; sin embargo, es un trabajo con una escala relativamente pequeña que podría beneficiarse al aumentar el número de muestras y profundizar en el análisis de otros tipos celulares. Concretamente hablando, los datos empleados en este proyecto abarcan tan solo una de cinco posibles muestras de cáncer endometrial disponibles en el mismo repositorio, por lo que la inclusión de otras muestras podría reforzar, refutar y/o complementar los resultados aquí obtenidos. Otro punto a considerar es que el análisis se enfocó únicamente en dos de los múltiples módulos asociados con las células de interés, de modo que es posible que algunos genes de importancia hayan sido omitidos durante el procesamiento. Finalmente, es necesario recalcar que el análisis de las redes, aunque práctico, representa solo un fragmento de una estructura mucho más grande de la cual es posible extraer información posiblemente diferente y más completa que la presentada. De modo contundente, si se desea seguir expandiendo la investigación en esta dirección, será necesaria la implementación de más muestras, la comparación entre varios tipos celulares y las células cancerosas, y un análisis más detallado acerca de distintas propiedades de las redes, esto sin mencionar el uso de nuevas herramientas que se desarrollen y tengan una mayor resolución y fiabilidad.

9. REFERENCIAS

Agostinis, C., Mangogna, A., Bossi, F., Ricci, G., Kishore, U., & Bulla, R. (2019). Uterine Immunity and Microbiota: A Shifting Paradigm. *Frontiers in Immunology*, *10*, 2387. <https://doi.org/10.3389/fimmu.2019.02387>

American Cancer Society. (1930). *Cancer Facts & Figures 2024* (pp. 32–33). American Cancer Society. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2024/2024-cancer-facts-and-figures-acf.pdf>

- Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R. P., Wolters, P. J., Abate, A. R., Butte, A. J., & Bhattacharya, M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*, *20*(2), 163–172. <https://doi.org/10.1038/s41590-018-0276-y>
- Arend, R. C., Jones, B. A., Martinez, A., & Goodfellow, P. (2018). Endometrial cancer: Molecular markers and management of advanced stage disease. *Gynecologic Oncology*, *150*(3), 569–580. <https://doi.org/10.1016/j.ygyno.2018.05.015>
- Arrowsmith, S., Keov, P., Muttenthaler, M., & Gruber, C. W. (2018). Contractility Measurements of Human Uterine Smooth Muscle to Aid Drug Development. *Journal of Visualized Experiments : JoVE*, *131*, 56639. <https://doi.org/10.3791/56639>
- Bahena-González, A., Isla-Ortiz, D., Trejo-Durán, E., Arango-Bravo, E., Cano-Blanco, C., Morales-Vásquez, F., Gutiérrez-Torres, S., Flores-Balcázar, C., Flores-Castro, M., Sepúlveda-Rivera, C. M., Vidal-Torres, D., Moreno-Flores, G., & Barquet-Muñoz, S. A. (2023). Oncoguía de cáncer de endometrio 2023. *Latin american journal of clinical sciences and medical technology*, *5*(1), 194–205. <https://doi.org/10.34141/LJCS5711966>
- Banno, K., Kisu, I., Yanokura, M., Tsuji, K., Masuda, K., Ueki, A., Kobayashi, Y., Yamagami, W., Nomura, H., Tominaga, E., Susumu, N., & Aoki, D. (2012). Biomarkers in endometrial cancer: Possible clinical applications (Review). *Oncology Letters*, *3*(6), 1175–1180. <https://doi.org/10.3892/ol.2012.654>
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., & Galon, J. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091–1093. <https://doi.org/10.1093/bioinformatics/btp101>
- Blighe, K. (2024). *Kevinblighe/CorLevelPlot* [R]. <https://github.com/kevinblighe/CorLevelPlot>
- Brown, J. S., Amend, S. R., Austin, R. H., Gatenby, R. A., Hammarlund, E. U., & Pienta, K. J. (2023). Updating the Definition of Cancer. *Molecular Cancer Research*, *21*(11), 1142–1147. <https://doi.org/10.1158/1541-7786.MCR-23-0411>

- Brummel, K., Eerkens, A. L., de Bruyn, M., & Nijman, H. W. (2023). Tumour-infiltrating lymphocytes: From prognosis to treatment selection. *British Journal of Cancer*, *128*(3), 451–458. <https://doi.org/10.1038/s41416-022-02119-4>
- Camps, J., Noël, F., Liechti, R., Massenet-Regad, L., Rigade, S., Götz, L., Hoffmann, C., Amblard, E., Saichi, M., Ibrahim, M. M., Pollard, J., Medvedovic, J., Roeder, H. G., & Soumelis, V. (2023). Meta-Analysis of Human Cancer Single-Cell RNA-Seq Datasets Using the IMMUCan Database. *Cancer Research*, *83*(3), 363–373. <https://doi.org/10.1158/0008-5472.CAN-22-0074>
- Carbone, A. (2020). Cancer Classification at the Crossroads. *Cancers*, *12*(4), 980. <https://doi.org/10.3390/cancers12040980>
- Centers for Disease Control and Prevention. (2021). *Smoking and Cancer*. Smoking & Tobacco Use. https://www.cdc.gov/tobacco/basic_information/health_effects/cancer/index.htm
- Chowdhury, H. A., Bhattacharyya, D. K., & Kalita, J. K. (2019). (Differential) Co-Expression Analysis of Gene Expression: A Survey of Best Practices. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1–1. <https://doi.org/10.1109/TCBB.2019.2893170>
- Cormio, A., Cormio, G., Musicco, C., Sardanelli, A. M., Gasparre, G., & Gadaleta, M. N. (2015). Mitochondrial changes in endometrial carcinoma: Possible role in tumor diagnosis and prognosis (Review). *Oncology Reports*, *33*(3), 1011–1018. <https://doi.org/10.3892/or.2014.3690>
- Csárdi, G., Nepusz, T., Müller, K., Horvát, S., Traag, V., Zanini, F., & Noom, D. (2024). *igraph for R: R interface of the igraph library for graph theory and network analysis* (v2.0.2) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.10681749>
- Dagogo-Jack, I., & Shaw, A. T. (2017). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, *15*. <https://doi.org/10.1038/nrclinonc.2017.166>
- DeNardo, D. G., & Ruffell, B. (2019). Macrophages as regulators of tumor immunity and

- immunotherapy. *Nature Reviews. Immunology*, 19(6), 369.
<https://doi.org/10.1038/s41577-019-0127-6>
- DePasquale, E. A. K., Schnell, D. J., Camp, P.-J. V., Valiente-Alandí, Í., Blaxall, B. C., Grimes, H. L., Singh, H., & Salomonis, N. (2019). DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. *Cell Reports*, 29(6), 1718-1727.e8.
<https://doi.org/10.1016/j.celrep.2019.09.082>
- Draganov, A., Jørgensen, J. R., Nellesmann, K. S., Mottin, D., Assent, I., Berry, T., & Aslay, C. (2023). *ActUp: Analyzing and Consolidating tSNE and UMAP* (arXiv:2305.07320). arXiv. <http://arxiv.org/abs/2305.07320>
- Elhamamsy, A. R., Metge, B. J., Alsheikh, H. A., Shevde, L. A., & Samant, R. S. (2022). Ribosome Biogenesis: A Central Player in Cancer Metastasis and Therapeutic Resistance. *Cancer Research*, 82(13), 2344–2353.
<https://doi.org/10.1158/0008-5472.CAN-21-4087>
- Fan, G.-T., Ling, Z.-H., He, Z.-W., Wu, S.-J., & Zhou, G.-X. (2021). Suppressing CHD1L reduces the proliferation and chemoresistance in osteosarcoma. *Biochemical and Biophysical Research Communications*, 554, 214–221.
<https://doi.org/10.1016/j.bbrc.2020.12.109>
- Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 149(4), 778–789. <https://doi.org/10.1002/ijc.33588>
- Franzén, O., Gan, L.-M., & Björkegren, J. L. M. (2019). PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, 2019, baz046. <https://doi.org/10.1093/database/baz046>
- Fu, L.-Q., Du, W.-L., Cai, M.-H., Yao, J.-Y., Zhao, Y.-Y., & Mou, X.-Z. (2020). The roles of tumor-associated macrophages in tumor angiogenesis and metastasis. *Cellular Immunology*, 353, 104119. <https://doi.org/10.1016/j.cellimm.2020.104119>
- Galgano, M. T., Hampton, G. M., & Frierson, H. F. (2006). Comprehensive analysis of HE4 expression in normal and malignant human tissues. *Modern Pathology*, 19(6),

- 847–853. <https://doi.org/10.1038/modpathol.3800612>
- Gonzalez, H., Hagerling, C., & Werb, Z. (2018). Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes & Development*, *32*(19–20), 1267–1284. <https://doi.org/10.1101/gad.314617.118>
- Gooden, M. J. M., de Bock, G. H., Leffers, N., Daemen, T., & Nijman, H. W. (2011). The prognostic influence of tumour-infiltrating lymphocytes in cancer: A systematic review with meta-analysis. *British Journal of Cancer*, *105*(1), 93–103. <https://doi.org/10.1038/bjc.2011.189>
- Grynberg, K., Ma, F. Y., & Nikolic-Paterson, D. J. (2017). The JNK Signaling Pathway in Renal Fibrosis. *Frontiers in Physiology*, *8*. <https://doi.org/10.3389/fphys.2017.00829>
- Guo, H. J., Rahimi, N., & Tadi, P. (2024). Biochemistry, Ubiquitination. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK556052/>
- Hammouda, M. B., Ford, A. E., Liu, Y., & Zhang, J. Y. (2020). The JNK Signaling Pathway in Inflammatory Skin Disorders and Cancer. *Cells*, *9*(4), Article 4. <https://doi.org/10.3390/cells9040857>
- Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
- Hippen, A. A., Falco, M. M., Weber, L. M., Erkan, E. P., Zhang, K., Doherty, J. A., Vähärautio, A., Greene, C. S., & Hicks, S. C. (2021). miQC: An adaptive probabilistic framework for quality control of single-cell RNA-sequencing data. *PLOS Computational Biology*, *17*(8), e1009290. <https://doi.org/10.1371/journal.pcbi.1009290>
- Horree, N., Heintz, A. P. M., Sie-Go, D. M. D. S., & van Diest, P. J. (2007). P16 is consistently expressed in endometrial tubal metaplasia. *Cellular Oncology: The Official Journal of*

- the International Society for Cellular Oncology*, 29(1), 37–45.
<https://doi.org/10.1155/2007/868952>
- Huang, Q., Liu, Y., Du, Y., & Garmire, L. X. (2021). Evaluation of Cell Type Annotation R Packages on Single-cell RNA-seq Data. *Genomics, Proteomics & Bioinformatics*, 19(2), 267–281. <https://doi.org/10.1016/j.gpb.2020.07.004>
- Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, 50(8), Article 8. <https://doi.org/10.1038/s12276-018-0071-8>
- Imoto, Y., Nakamura, T., Escolar, E. G., Yoshiwaki, M., Kojima, Y., Yabuta, Y., Katou, Y., Yamamoto, T., Hiraoka, Y., & Saitou, M. (2022). Resolution of the curse of dimensionality in single-cell RNA sequencing data analysis. *Life Science Alliance*, 5(12), e202201591. <https://doi.org/10.26508/lsa.202201591>
- Iruela-Arispe, M. L., Rodriguez-Manzaneque, J. C., & Abu-Jawdeh, G. (1999). Endometrial Endothelial Cells Express Estrogen and Progesterone Receptors and Exhibit a Tissue Specific Response to Angiogenic Growth Factors. *Microcirculation*, 6(2), 127–140. <https://doi.org/10.1111/j.1549-8719.1999.tb00095.x>
- Jawaid, W. (2023). *enrichR-package: enrichR package in enrichR: Provides an R Interface to 'Enrichr'* [Computer software]. <https://rdrr.io/cran/enrichR/man/enrichR-package.html>
- Jia, L., Yang, T., Gu, X., Zhao, W., Tang, Q., Wang, X., Zhu, J., & Feng, Z. (2019). Translation elongation factor eEF1B α is identified as a novel prognostic marker of gastric cancer. *International Journal of Biological Macromolecules*, 126, 345–351. <https://doi.org/10.1016/j.ijbiomac.2018.12.126>
- Jiang, P. (2019). Quality Control of Single-Cell RNA-seq. In G.-C. Yuan (Ed.), *Computational Methods for Single-Cell Data Analysis* (pp. 1–9). Springer. https://doi.org/10.1007/978-1-4939-9057-3_1
- Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., & Luo, Y. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine*, 12(3), e694. <https://doi.org/10.1002/ctm2.694>

- Kiselev, V. Y., Andrews, T. S., & Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nature Reviews Genetics*, *20*(5), 273–282.
<https://doi.org/10.1038/s41576-018-0088-9>
- Klaes, R., Friedrich, T., Spitkovsky, D., Ridder, R., Rudy, W., Petry, U., Dallenbach-Hellweg, G., Schmidt, D., & Von Knebel Doeberitz, M. (2001). Overexpression of p16INK4A as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. *International Journal of Cancer*, *92*(2), 276–284. <https://doi.org/10.1002/ijc.1174>
- Kobak, D., & Linderman, G. C. (2021). Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nature Biotechnology*, *39*(2), 156–157.
<https://doi.org/10.1038/s41587-020-00809-z>
- Koboldt, D. C. (2020). Best practices for variant calling in clinical sequencing. *Genome Medicine*, *12*(1), 91. <https://doi.org/10.1186/s13073-020-00791-w>
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, *58*(4), 610–620. <https://doi.org/10.1016/j.molcel.2015.04.005>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(Web Server issue), W90–W97. <https://doi.org/10.1093/nar/gkw377>
- Kunes, R. Z., Walle, T., Land, M., Nawy, T., & Pe'er, D. (2023). Supervised discovery of interpretable gene programs from single-cell data. *Nature Biotechnology*, 1–12.
<https://doi.org/10.1038/s41587-023-01940-3>
- Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559.
<https://doi.org/10.1186/1471-2105-9-559>
- Lee, J. T., & Gu, W. (2010). The multiple levels of regulation by p53 ubiquitination. *Cell Death & Differentiation*, *17*(1), 86–92. <https://doi.org/10.1038/cdd.2009.77>

- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766.
<https://doi.org/10.1016/j.jesp.2013.03.013>
- Loid, M., Obukhova, D., Kask, K., Meltsov, A., Derks, K., Altmäe, S., Saare, M., Peters, M., Minajeva, A., Adler, P., Krjutškov, K., Esteki, M. Z., & Salumets, A. (2023). *Aging affects ciliated cells development in the human endometrial epithelium*.
<https://doi.org/10.1101/2023.05.22.23290333>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
<https://doi.org/10.1186/s13059-014-0550-8>
- Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: A tutorial. *Molecular Systems Biology*, 15(6), e8746.
<https://doi.org/10.15252/msb.20188746>
- Maenhoudt, N., De Moor, A., & Vankelecom, H. (2022). Modeling Endometrium Biology and Disease. *Journal of Personalized Medicine*, 12(7), 1048.
<https://doi.org/10.3390/jpm12071048>
- Makker, V., MacKay, H., Ray-Coquard, I., Levine, D. A., Westin, S. N., Aoki, D., & Oaknin, A. (2021). Endometrial Cancer. *Nature Reviews. Disease Primers*, 7(1), 88.
<https://doi.org/10.1038/s41572-021-00324-8>
- McCarthy, D. J., Campbell, K. R., Lun, A. T. L., & Wills, Q. F. (2017). Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, 33(8), 1179–1186. <https://doi.org/10.1093/bioinformatics/btw777>
- McGinnis, C. S., Murrow, L. M., & Gartner, Z. J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems*, 8(4), 329–337.e4. <https://doi.org/10.1016/j.cels.2019.03.003>
- McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, 168(4), 613–628.

<https://doi.org/10.1016/j.cell.2017.01.018>

McKinnon, B. D., Lukowski, S. W., Mortlock, S., Crawford, J., Atluri, S., Subramaniam, S., Johnston, R. L., Nirgianakis, K., Tanaka, K., Amoako, A., Mueller, M. D., & Montgomery, G. W. (2022). Altered differentiation of endometrial mesenchymal stromal fibroblasts is associated with endometriosis susceptibility. *Communications Biology*, 5(1), 1–14. <https://doi.org/10.1038/s42003-022-03541-3>

Montenegro, J. D. (2022). Gene Co-expression Network Analysis. In *Plant Bioinformatics Methods and Protocols* (3rd ed., pp. 387–404). Humana New York, NY. <https://www.scribd.com/document/665640859/Methods-in-Molecular-Biology-2443-David-Edwards-Plant-Bioinformatics-Methods-and-Protocols-2022-Humana-10-1007-978-1-0716-2067-0-Libgen-l>

Mourksi, N.-E.-H., Dalban, C., Colombe-Vermorel, A., Odeyer, L., Simioni, V., Frenel, J.-S., Fabbro, M., Bazan, F., Abadie-Lacourtoisie, S., Coquan, E., Martinez, S., Garin, G., Tabone-Eglinger, S., Treilleux, I., Chabaud, S., Pérol, D., Ray-Coquard, I., Heudel, P.-E., Diaz, J.-J., & Marcel, V. (2023). Ribosome biogenesis-based predictive biomarkers in endocrine therapy (Anastrozole) combined with mTOR inhibitor (Vistusertib) in endometrial cancer: Translational study from the VICTORIA trial in collaboration with the GINECO group. *Molecular Oncology*, 17(1), 27–36. <https://doi.org/10.1002/1878-0261.13340>

Müllner, D. (2013). fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software*, 53, 1–18. <https://doi.org/10.18637/jss.v053.i09>

Musicco, C., Cormio, G., Pesce, V., Loizzi, V., Cicinelli, E., Resta, L., Ranieri, G., & Cormio, A. (2018). Mitochondrial Dysfunctions in Type I Endometrial Carcinoma: Exploring Their Role in Oncogenesis and Tumor Progression. *International Journal of Molecular Sciences*, 19(7), 2076. <https://doi.org/10.3390/ijms19072076>

Nam, J.-H., Park, J.-Y., Baek, M.-H., Han, J.-H., Han, W.-S., & Ko, A.-R. (2014). Immunohistochemistry as Biomarker Platform in Endometrial Cancer. In V. R. Preedy

- & V. B. Patel (Eds.), *Biomarkers in Cancer* (pp. 1–19). Springer Netherlands.
https://doi.org/10.1007/978-94-007-7744-6_42-1
- National Cancer Institute. (2021). *What Is Cancer?* (nciglobal,ncienterprise) [cgvArticle].
<https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
- Osorio, D., & Cai, J. J. (2020). Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics*, 37(7), 963–967. <https://doi.org/10.1093/bioinformatics/btaa751>
- Pan, Y., Yu, Y., Wang, X., & Zhang, T. (2020). Tumor-Associated Macrophages in Tumor Immunity. *Frontiers in Immunology*, 11. <https://doi.org/10.3389/fimmu.2020.583084>
- PDQ® Adult Treatment Editorial Board. (2024, February 9). *PDQ Endometrial Cancer Treatment* (nciglobal,ncienterprise) [pdqCancerInfoSummary].
<https://www.cancer.gov/types/uterine/hp/endometrial-treatment-pdq>
- Pelletier, J., Thomas, G., & Volarević, S. (2018). Ribosome biogenesis in cancer: New players and therapeutic avenues. *Nature Reviews Cancer*, 18(1), 51–63.
<https://doi.org/10.1038/nrc.2017.104>
- Pullin, J. M., & McCarthy, D. J. (2024). A comparison of marker gene selection methods for single-cell RNA sequencing data. *Genome Biology*, 25(1), 56.
<https://doi.org/10.1186/s13059-024-03183-0>
- Qu, J., Zhou, Y., Li, Y., Yu, J., & Wang, W. (2021). CASK regulates Notch pathway and functions as a tumor promoter in pancreatic cancer. *Archives of Biochemistry and Biophysics*, 701, 108789. <https://doi.org/10.1016/j.abb.2021.108789>
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. (4.3.3) [Computer software].
- Ramón y Cajal, S., Sesé, M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S. J., Hernández-Losa, J., & Castellví, J. (2020). Clinical implications of intratumor heterogeneity: Challenges and opportunities. *Journal of Molecular Medicine (Berlin, Germany)*, 98(2), 161–177. <https://doi.org/10.1007/s00109-020-01874-2>
- Regner, M. J., Wisniewska, K., Garcia-Recio, S., Thennavan, A., Mendez-Giraldez, R.,

- Malladi, V. S., Hawkins, G., Parker, J. S., Perou, C. M., Bae-Jump, V. L., & Franco, H. L. (2021). A multi-omic single-cell landscape of human gynecologic malignancies. *Molecular Cell*, S1097-2765(21)00842-X. <https://doi.org/10.1016/j.molcel.2021.10.013>
- Riabov, V., Gudima, A., Wang, N., Mickley, A., Orekhov, A., & Kzhyshkowska, J. (2014). Role of tumor associated macrophages in tumor angiogenesis and lymphangiogenesis. *Frontiers in Physiology*, 5, 75. <https://doi.org/10.3389/fphys.2014.00075>
- Robichaud, N., Sonenberg, N., Ruggero, D., & Schneider, R. J. (2019). Translational Control in Cancer. *Cold Spring Harbor Perspectives in Biology*, 11(7), a032896. <https://doi.org/10.1101/cshperspect.a032896>
- Rodriguez-Garcia, M., Patel, M. V., Shen, Z., & Wira, C. R. (2021). The impact of aging on innate and adaptive immunity in the human female genital tract. *Aging Cell*, 20(5), e13361. <https://doi.org/10.1111/accel.13361>
- Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R. E., Neild, A., Farbehi, N., Bartonicek, N., & Gallego-Ortega, D. (2019). Droplet-based single cell RNAseq tools: A practical guide. *Lab on a Chip*, 19(10), 1706–1727. <https://doi.org/10.1039/C8LC01239C>
- Sarmah, T., & Bhattacharyya, D. K. (2021). A study of tools for differential co-expression analysis for RNA-Seq data. *Informatics in Medicine Unlocked*, 26, 100740. <https://doi.org/10.1016/j.imu.2021.100740>
- Shinn, M. (2023). Phantom oscillations in principal component analysis. *PNAS*, 120(48). <https://doi.org/10.1073/pnas.2311420120>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Svensson, V., Natarajan, K. N., Ly, L.-H., Miragaia, R. J., Labalette, C., Macaulay, I. C., Cvejic, A., & Teichmann, S. A. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nature Methods*, 14(4), 381–387. <https://doi.org/10.1038/nmeth.4220>

- The American Cancer Society medical and editorial content team. (2019, March 27). *What Is Endometrial Cancer? What Is Endometrial Cancer?*
<https://www.cancer.org/cancer/types/endometrial-cancer/about/what-is-endometrial-cancer.html>
- Tian, J., Yang, J., Chen, T., Yin, Y., Li, N., Li, Y., Luo, X., Dong, E., Tan, H., Ma, Y., & Li, T. (2023). Generation of Human Endometrial Assembloids with a Luminal Epithelium using Air–Liquid Interface Culture Methods. *Advanced Science*, *10*(30), 2301868.
<https://doi.org/10.1002/advs.202301868>
- Tsuyuzaki, K., Sato, H., Sato, K., & Nikaido, I. (2020). Benchmarking principal component analysis for large-scale single-cell RNA-sequencing. *Genome Biology*, *21*(1), 9.
<https://doi.org/10.1186/s13059-019-1900-3>
- Vaiopoulos, A. G., Papachroni, K. K., & Papavassiliou, A. G. (2010). Colon carcinogenesis: Learning from NF- κ B and AP-1. *The International Journal of Biochemistry & Cell Biology*, *42*(7), 1061–1065. <https://doi.org/10.1016/j.biocel.2010.03.018>
- Vitale, I., Shema, E., Loi, S., & Galluzzi, L. (2021). Intratumoral heterogeneity in cancer progression and response to immunotherapy. *Nature Medicine*, *27*(2), 212–224.
<https://doi.org/10.1038/s41591-021-01233-9>
- Wang, D.-W., Tang, J.-Y., Zhang, G.-Q., & Chang, X.-T. (2020). ARHGEF10L expression regulates cell proliferation and migration in gastric tumorigenesis. *Bioscience, Biotechnology, and Biochemistry*, *84*(7), 1362–1372.
<https://doi.org/10.1080/09168451.2020.1737503>
- Wang, H., Xing, J., Wang, W., Lv, G., He, H., Lu, Y., Sun, M., Chen, H., & Li, X. (2021). Molecular Characterization of the Oncogene BTF3 and Its Targets in Colorectal Cancer. *Frontiers in Cell and Developmental Biology*, *8*.
<https://doi.org/10.3389/fcell.2020.601502>
- Wang, W., Vilella, F., Alama, P., Moreno, I., Mignardi, M., Isakova, A., Pan, W., Simon, C., & Quake, S. R. (2020). Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle. *Nature Medicine*, *26*(10), 1644–1653.

<https://doi.org/10.1038/s41591-020-1040-z>

- Wilcox, R. R. (2003). SUMMARIZING DATA. In *Applying Contemporary Statistical Techniques* (pp. 55–91). Elsevier. <https://doi.org/10.1016/B978-012751541-0/50024-9>
- Xiang, R., Wang, W., Yang, L., Wang, S., Xu, C., & Chen, X. (2021). A Comparison for Dimensionality Reduction Methods of Single-Cell RNA-seq Data. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.646936>
- Xu, J., Li, L., Shi, P., Cui, H., & Yang, L. (2022). The Crucial Roles of Bmi-1 in Cancer: Implications in Pathogenesis, Metastasis, Drug Resistance, and Targeted Therapies. *International Journal of Molecular Sciences*, *23*(15), 8231. <https://doi.org/10.3390/ijms23158231>
- Yu, H., Lin, L., Zhang, Z., Zhang, H., & Hu, H. (2020). Targeting NF- κ B pathway for the therapy of diseases: Mechanism and clinical study. *Signal Transduction and Targeted Therapy*, *5*(1), 1–23. <https://doi.org/10.1038/s41392-020-00312-6>
- Zhao, E. Y., Jones, M., & Jones, S. J. M. (2019). Whole-Genome Sequencing in Cancer. *Cold Spring Harbor Perspectives in Medicine*, *9*(3), a034579. <https://doi.org/10.1101/cshperspect.a034579>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*, 14049. <https://doi.org/10.1038/ncomms14049>
- Zou, Z., Tao, T., Li, H., & Zhu, X. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: Progress and challenges. *Cell & Bioscience*, *10*(1), 31. <https://doi.org/10.1186/s13578-020-00396-1>