



**viu**

**Universidad  
Internacional  
de Valencia**

# Análisis de datos de *single cell RNA-seq* para evaluar la eficacia de la terapia hormonal en el tratamiento del dolor en la endometriosis

Titulación:  
Máster en Bioinformática  
  
Curso académico  
2023 – 2024

Alumna: Dodu, M Alexandra  
  
Director/a de TFM: Nieto Sachicha,  
J

Convocatoria:  
Segunda

# Índice

<b>1.</b>	<b>INTRODUCCIÓN .....</b>	<b>6</b>
1.1.	Introducción general a la endometriosis .....	6
1.1.1.	<i>Definición, localización y prevalencia .....</i>	6
1.1.2.	<i>Origen .....</i>	7
1.1.3.	<i>Clasificación.....</i>	8
1.1.4.	<i>Diagnóstico y tratamiento .....</i>	8
1.2.	Sintomatología de la endometriosis.....	10
1.3.	Biología del dolor .....	11
1.4.	Tratamiento del dolor: enfoque en el tratamiento farmacológico.....	15
1.5.	La tecnología de <i>single cell RNA sequencing</i> .....	17
<b>2.</b>	<b>OBJETIVOS .....</b>	<b>19</b>
<b>3.</b>	<b>METODOLOGÍA.....</b>	<b>20</b>
3.1.	Obtención de datos públicos de scRNA-seq .....	20
3.2.	Preprocesamiento de datos de scRNA-seq.....	20
3.3.	Procesamiento y análisis de datos de scRNA-seq .....	22
3.3.1.	<i>Anotación.....</i>	22
3.3.2.	<i>Identificación de firma del dolor en tipos celulares y evaluación del impacto del tratamiento hormonal .....</i>	22
3.3.3.	<i>Visualización.....</i>	22
<b>4.</b>	<b>RESULTADOS .....</b>	<b>24</b>
4.1.	Descarga de datos públicos de scRNA-seq de pacientes con endometriosis.....	24
4.2.	Preprocesamiento de datos de scRNA-seq.....	24
4.3.	Caracterización de células del endometrio sano y lesiones endometriales .....	25
4.4.	Caracterización de genes del dolor en células estromales, inmunes y fibroblastos...	29
4.5.	Impacto del tratamiento en los genes implicados en la vía del dolor .....	31
<b>5.</b>	<b>DISCUSIÓN.....</b>	<b>33</b>
<b>6.</b>	<b>CONCLUSIONES .....</b>	<b>37</b>
<b>7.</b>	<b>CÓDIGO Y RECURSOS EN GITHUB .....</b>	<b>37</b>
<b>8.</b>	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>38</b>
<b>9.</b>	<b>ANEXOS.....</b>	<b>47</b>

## Listas de abreviaturas

<b>ADAEC</b>	Asociación de Afectadas de Endometriosis Crónica
<b>ADNc</b>	ADN complementario
<b>AINE</b>	Fármacos antiinflamatorios no-esteroideos
<b>COX-2</b>	Ciclooxygenasa-2
<b>DAMP</b>	Patrones moleculares asociados al peligro
<b>DEG</b>	Genes diferencialmente expresados
<b>DPC</b>	Dolor pélvico crónico
<b>E2</b>	Estradiol
<b>EP</b>	Endometriosis profunda
<b>GEO</b>	<i>Gene Expression Omnibus</i>
<b>GnRH</b>	Hormona liberadora de gonadotropinas
<b>NGS</b>	<i>Next Generation Sequencing</i> , secuenciación de próxima generación
<b>PCA</b>	Análisis de Componentes Principales
<b>PGE2</b>	Prostaglandinas E2
<b>RNA-seq</b>	Secuenciación de ARN
<b>ROS</b>	Especies reactivas de oxígeno
<b>scRNA-seq</b>	Secuenciación de ARN de célula única o <i>single cell</i>
<b>SNC</b>	Sistema nervioso central
<b>SNP</b>	Sistema nervioso periférico
<b>UMI</b>	Identificador único molecular

# Índice de Figuras

<b>Figura 1. Representación de (A) la estructura del útero, (B) endometrio (Modificado de Pawlina et al., 2020 [18] y Zhu et al., 2014 [19]) y (C) de la teoría de menstruación retrógrada de Sampson (Elaborada con BioRender.com).</b> .....	7
<b>Figura 2. Tipos de lesiones endometriales y su localización (Elaborado con BioRender.com).</b> .....	8
<b>Figura 3. Clasificación revisada de la endometriosis de la Sociedad Estadounidense de Medicina Reproductiva (rASRM) (Modificado de Zondervan et al., 2018 [25]).</b> .....	9
<b>Figura 4. Efecto hormonal sobre (A) el endometrio sano (Elaborado con BioRender.com) y (B) las lesiones endometriales (Modificado de Laux-Biehlmann et al., 2015 [46]).</b> .....	12
<b>Figura 5. Representación esquemática general de los mecanismos del dolor en las lesiones endometriales ((A, C) Modificado de Laux-Biehlmann et al., 2015 [46]; (B) Adaptado de Saunders et al., 2021 [1] con BioRender.com).</b> .....	13
<b>Figura 6. Representación esquemática local de los mecanismos del dolor en las lesiones endometriales (Modificado y adaptado de Laux-Biehlmann et al., 2015 [46]).</b> .....	14
<b>Figura 7. Tratamientos farmacológicos usados para disminuir el dolor y sus dianas (Modificado de [48] con BioRender.com).</b> .....	16
<b>Figura 8. Flujo de trabajo bioinformático para el análisis de datos públicos de scRNA-seq de pacientes con endometriosis con y sin tratamiento hormonal (Elaborado con BioRender.com).</b> .....	23
<b>Figura 9. Representación de la cantidad de genes (nGene) frente a la cantidad de UMI coloreadas por el porcentaje de lecturas mitocondriales en el dataset A (A) y dataset B (B).</b> .....	24
<b>Figura 10. Elbow plots o gráficos de codo del dataset A (A) y dataset B (B).</b> .....	25
<b>Figura 11. Agrupación de células del endometrio eutópico y ectópico.</b> .....	26



<b>Figura 12. Caracterización de células del endometrio eutópico y ectópico I.</b> .....	27
<b>Figura 13. Caracterización de células del endometrio eutópico y ectópico II.</b> .....	28
<b>Figura 14. Expresión de la firma de dolor en los diferentes tipos celulares.</b> .....	29
<b>Figura 15. Observación en mapas de color de la expresión de los genes diferencialmente expresados en los tipos celulares con mayor firma del dolor de ambos conjuntos de datos.</b> .....	30
<b>Figura 16. Mapas de expresión génica en diferentes tipos celulares y condiciones de endometriosis y control, con y sin tratamiento hormonal, en el (A) dataset A y en el (B) dataset B.</b> .....	32

## Resumen y palabras clave

La endometriosis es un síndrome inflamatorio dependiente de estrógenos que afecta al 10% de las mujeres a nivel mundial. Se caracteriza por la presencia de tejido similar al endometrial fuera de su ubicación normal, denominado endometrio ectópico o lesiones endometriales, que puede causar inflamación y el consecuente dolor crónico y cíclico (dismenorrea) en las pacientes. El tratamiento principal de este síntoma a largo plazo es la terapia hormonal. Si bien los mecanismos que generan el dolor no están completamente comprendidos, se ha identificado una serie de genes involucrados en los procesos de inflamación y nocicepción. En este trabajo, se analizó el impacto del tratamiento hormonal en la expresión de genes asociados al dolor en pacientes con endometriosis mediante el análisis de datos de secuenciación de ARN de célula única o *single cell* (scRNA-seq). El objetivo principal del estudio fue analizar el impacto de la terapia hormonal en la expresión de estos genes en diferentes tipos celulares del endometrio eutópico y ectópico. Para ello, se procesaron y analizaron con Seurat 60 muestras de dos conjuntos de datos públicos de scRNA-seq, uno proveniente de pacientes con tratamiento hormonal y otro compuesto mayoritariamente por muestras sin tratamiento.

Los resultados muestran una mayor expresión de genes asociados al dolor en células estromales, monocitos, macrófagos y fibroblastos estromales, tipos celulares que están implicados en el mecanismo del dolor. El tratamiento hormonal tuvo un impacto significativo en la reducción de la expresión de *SPARC*, *MDK*, *PTN* y *PRRX1* en células inmunes y estromales. Además, la reducción en la expresión del receptor de estrógenos *ESR1* en las células estromales sugiere una disminución de la inflamación mediada por estrógenos, lo que está relacionado con una menor percepción del dolor. También se observó una reducción en la expresión de la mayoría de los genes asociados al dolor en las células estromales durante la fase menstrual, lo que apoya la eficacia del tratamiento hormonal en la disminución de la dismenorrea.

En conclusión, el análisis revela que el tratamiento hormonal modula la expresión de genes relacionados con la percepción del dolor en células estromales, monocitos y macrófagos, lo que podría explicar en parte la mejora de los síntomas en pacientes bajo esta terapia.

**Palabras clave:** Endometriosis, scRNA-seq, terapia hormonal, dolor, dismenorrea, expresión génica, *SPARC*, *MDK*, *PNT*, *PRRX1*, *ESR1*, Seurat.

# 1. Introducción

## 1.1. Introducción general a la endometriosis

### 1.1.1. Definición, localización y prevalencia

La endometriosis es un síndrome [1] inflamatorio complejo [2] y dependiente de estrógenos [3], caracterizado por el crecimiento de tejido funcional similar al endometrial fuera del endometrio, causando las denominadas lesiones [4,5]. Este endometrio ectópico se da principalmente en la cavidad pélvica, como paredes externas del útero y ovarios, aunque en algunos casos raros, puede encontrarse en otras localizaciones como intestino, pulmones, hígado, vejiga, recto [4–6], ojos [7], entre otros.

*En este trabajo se usará el término "mujeres", pero hay que resaltar que la endometriosis puede afectar a cualquier persona con órganos del aparato reproductor femenino, incluidos hombres trans y personas no binarias.*

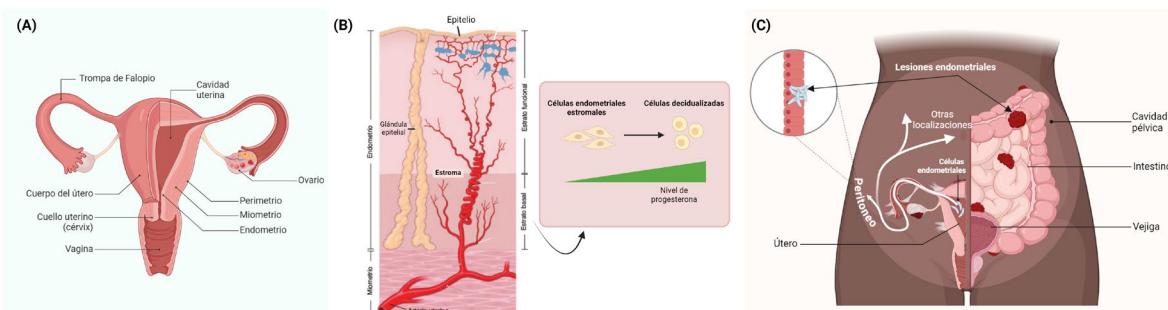
Este síndrome afecta hasta a un **10% de mujeres** en edad reproductiva (15-54 años) a nivel mundial [4,8], es decir, a unas 190 millones. Concretamente en España, se ha reportado una incidencia anual de 16,1 casos por 10.000 mujeres entre los años 2014 y 2017, es decir, unos 5.000 nuevos casos cada año [9].

Es importante destacar que los datos de prevalencia de endometriosis podrían estar subestimados. Actualmente, según un meta-análisis, las mujeres negras tienen menos probabilidades de ser diagnosticadas con endometriosis en comparación con las mujeres blancas, por lo que la prevalencia de la enfermedad podría verse influenciada por la raza o la etnia [10]. Además, el diagnóstico de la endometriosis puede tardar entre cuatro y once años desde la aparición de los primeros síntomas [11]. Esto podría deberse no solo a la complejidad del síndrome, sino también al sesgo de género de los profesionales de la salud hacia la subestimación del dolor en las mujeres [12,13] y a la normalización del mismo en la sociedad, considerándolo algo inevitable de la menstruación [14]. Según una encuesta, la mayoría de las estudiantes de Enfermería en España consideran el dolor menstrual normal y se automedican debido a las experiencias desfavorables con los profesionales de la salud [15]. Además, a pesar de tener síntomas anormales, muchas de las estudiantes perciben sus ciclos menstruales como normales, lo que demuestra un déficit de

conocimiento sobre la salud menstrual [16]. Esto indica que la falta de conocimiento o la incapacidad para autodiagnosticarse puede ser mayor en jóvenes que estudian otras disciplinas no relacionadas con la salud [16].

### 1.1.2. Origen

La endometriosis es un síndrome complejo y multifactorial cuya etiología aún se desconoce. Aunque se han relacionado diversos factores genéticos, endocrinos, inmunológicos y ambientales con el desarrollo de la endometriosis, no se ha establecido un origen concreto [4,5,17].



**Figura 1.** Representación de (A) la estructura del útero, (B) endometrio (Modificado de Pawlina et al., 2020 [18] y Zhu et al., 2014 [19]) y (C) de la teoría de menstruación retrógrada de Sampson (Elaborada con [BioRender.com](#)).

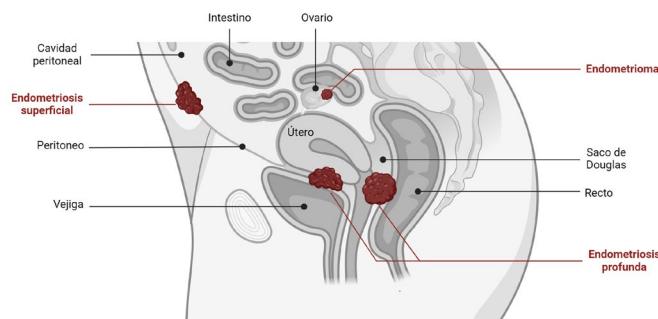
(A) Anatomía del órgano sexual femenino, el útero. Este está compuesto por tres capas: endometrio (capa más interna), miometrio (capa intermedia muscular lisa) y perimetrio (capa externa). (B) El endometrio se halla compuesto por una capa basal y una funcional. Una capa epitelial protege este tejido. Se hallan glándulas epiteliales embebidas en el estroma del endometrio. En el estrato basal, por efecto del aumento de la progesterona, las células estromales endometriales se diferencian a células decidualizadas. (C) Según la menstruación retrógrada, la endometriosis se desarrolla debido al flujo retrógrado de sangre menstrual a través de las trompas de Falopio hacia la cavidad pélvica, donde se depositan en el peritoneo, órganos pélvicos y otras localizaciones, formando las lesiones endometriales. Figura en alta resolución en este [enlace](#).

Entre las diversas teorías propuestas para explicar la patogénesis de la endometriosis ectópica, la teoría de la menstruación retrógrada de Sampson es la más conocida [20]. Esta sugiere que, durante la menstruación, una parte del tejido endometrial (células estromales) fluye hacia la cavidad peritoneal a través de las trompas de Falopio. Allí, se adhiere a los órganos pélvicos y peritoneo, lo que da como resultado las lesiones endometriales (**Figura 1C**). Sin embargo, la menstruación retrógrada ocurre en hasta el 90% de las mujeres que menstrúan, mientras que la endometriosis es mucho menos prevalente [4,17]. El sistema inmunológico suele eliminar estos fragmentos endometriales, evitando su implantación y crecimiento [21], lo que ha llevado a cuestionar esta teoría ampliamente. Por ello, se están

desarrollando nuevas teorías [21,22], aunque la menstruación retrógrada sigue siendo la más aceptada [4].

### 1.1.3. Clasificación

La presentación de la endometriosis es variada y puede incluir lesiones superficiales en el peritoneo (endometriosis superficial), quistes endometriales ováricos (endometriomas) y lesiones profundas en la pared vaginal, vejiga, colon, recto y otros lugares (endometriosis profunda, EP) [22–24] (**Figura 2**). Además, a menudo puede ir acompañada de adherencias y fibrosis [25].



**Figura 2.** Tipos de lesiones endometriales y su localización (Elaborado con [BioRender.com](#)).

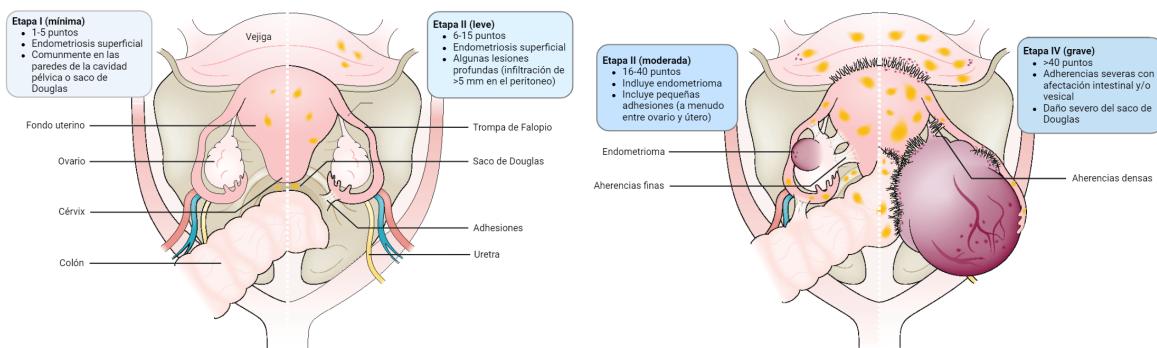
La endometriosis superficial afecta la superficie del peritoneo; el endometrioma se desarrolla en los ovarios; la endometriosis profunda (EP) infiltra tejidos subyacentes como vejiga y recto.

Aunque se han publicado más de 22 sistemas para clasificar la endometriosis en pacientes [26], el sistema más conocido y utilizado es el revisado por la Sociedad Americana de Medicina Reproductiva (rASRM) [27]. Este sistema de clasificación se basa en puntuaciones aleatorias y clasifica la endometriosis en cuatro etapas según la gravedad, el tamaño, la profundidad y la localización en el útero de las lesiones, y la presencia de adherencias: etapa I (endometriosis mínima, 1 a 5 puntos), etapa II (leve, 5 a 15 puntos), etapa III (moderada, 16 a 40 puntos) y etapa IV (grave, más de 40 puntos) (**Figura 3**). Sin embargo, a pesar de ser el sistema más empleado, no considera ubicaciones ectópicas de, por ejemplo, la EP [4,28], y tampoco toma en consideración una de las características más relevantes de la endometriosis: el dolor [28] (ver apartado 1.2). A día de hoy, no hay ninguna clasificación que tome en consideración este síntoma, por lo que se espera que se desarrolle un nuevo sistema que complemente los existentes.

### 1.1.4. Diagnóstico y tratamiento

La confirmación definitiva de la endometriosis generalmente requiere una cirugía con biopsia, específicamente mediante laparoscopia, para visualizar y categorizar las lesiones en el útero según el sistema de clasificación rASRM. Para la detección de endometriomas

y EP se pueden emplear enfoques de diagnóstico de imagen, como la ecografía transvaginal y la resonancia magnética [1,29].



**Figura 3. Clasificación revisada de la endometriosis de la Sociedad Estadounidense de Medicina Reproductiva (rASRM)** (Modificado de Zondervan et al., 2018 [25]).

Corte transversal de la pelvis. La etapa I (mínima, 1-5 puntos) generalmente incluye lesiones superficiales mínimas. La etapa II (leve, 6-15 puntos) puede presentar algunas lesiones peritoneales profundas, solas o junto con endometriosis superficial y adherencias ligeras. La etapa III (moderada, 16-40 puntos) con frecuencia presenta un endometrioma solo o en conjunto con endometriosis superficial, EP y/o adherencias más densas. La etapa IV (grave, más de 40 puntos) se distingue por múltiples implantes profundos, endometriomas en ambos ovarios y adherencias densas que pueden provocar una obstrucción parcial o completa de la pelvis. Figura en alta resolución en este [enlace](#).

Sin embargo, el diagnóstico de la endometriosis es complicado porque muchas mujeres pueden ser asintomáticas [4] o tener lesiones pequeñas (milímetros de diámetro) [25]. Dado que algunas mujeres con lesiones mínimas experimentan dolor intenso, la etapa rASRM no está relacionada con la intensidad de los síntomas [25]. Además, estos pueden estar asociados con otras condiciones médicas [4]. El sesgo de género y la normalización del dolor menstrual también contribuyen a la subestimación de los síntomas y los retrasos en la búsqueda de atención médica para su diagnóstico [12,14–16].

En España, en 2013 se publicó una guía nacional con el protocolo a seguir en casos con y de sospecha de endometriosis [30], pero no se ha actualizado y, según la Asociación de Afectadas de Endometriosis Crónica (ADAEC), tampoco se ha implementado [31]. Además, la Comunidad de Andalucía es la única que ha actualizado recientemente su propio guía [29] y que cuenta con unidades multidisciplinares de referencia [31], tal como indican necesarias los protocolos [29,30].

El objetivo del tratamiento de la endometriosis es tratar los síntomas (ver apartado 1.2): aliviar el dolor mediante tratamiento farmacológico y/o quirúrgico (en más detalle en el apartado 1.4) y aumentar la fertilidad mediante la fertilización *in vitro* [5,25].

## 1.2. Sintomatología de la endometriosis

Aunque hay casos de asintomatología [1], además de la **infertilidad**, el **dolor** es uno de los síntomas más frecuentes de la endometriosis [4]. Al principio, el síntoma principal es el dolor menstrual (**dismenorrea**) [32]. Con el tiempo, se desarrollan otros procesos dolorosos tanto cíclicos como crónicos, que hacen que la persona sea más sensible al dolor, resultando en dolor pélvico crónico (**DPC**, dolor no cíclico que dura al menos 6 meses) [32], dolor durante las relaciones sexuales con penetración (**dispareunia**), e incluso dolor al orinar (**disuria**) y al defecar (**disquecia**) [33], **dolor de espalda y migrañas** [32,34]. En concreto, el DPC es uno de los síntomas más comunes, llegándose a dar en un 60-75% de los casos [34,35]. Esto indica un sufrimiento significativo por parte de las mujeres con endometriosis.

Se ha demostrado que estos síntomas y, concretamente el dolor, reducen la calidad de vida de las mujeres, afectando negativamente la mayoría de los ámbitos de su vida, tanto social, profesional, académico y sexual, así como su salud física y mental [32,35–38]. Las mujeres con endometriosis suelen experimentar a causa del dolor niveles elevados de fatiga y problemas de sueño [32,37], lo que limita su capacidad para realizar ejercicio y llevar a cabo actividades diarias y laborales [36]. Además, la dismenorrea y la imprevisibilidad y frecuencia del DPC aumentan la frecuencia de ausencias y disminuyen la productividad y el rendimiento académico. Esto puede conllevar una reducción en el nivel educativo y laboral, y tener un impacto a largo plazo en el desarrollo profesional y personal [32,36,37]. Estos dolores también limitan la participación en actividades sociales y dificultan las relaciones interpersonales, lo que desemboca en sentimientos de soledad y aislamiento [32,36]. Es destacable mencionar que las mujeres con dolor crónico tienen un mayor riesgo de desarrollar depresión y ansiedad [35,37,38], en parte también debido al estigma social y al sesgo de género que minimizan su sufrimiento [12–14]. Finalmente, la dispareunia reduce la capacidad de experimentar orgasmos, empeorando la vida sexual y contribuyendo aún más al deterioro de la salud mental y emocional de las mujeres con endometriosis [37].

Además, estas mujeres presentan un mayor riesgo de desarrollar trastornos metabólicos, como la diabetes mellitus gestacional [39], enfermedades cardiovasculares [40] y ciertos tipos de cáncer, como el cáncer de ovario [41].

## 1.3. Biología del dolor

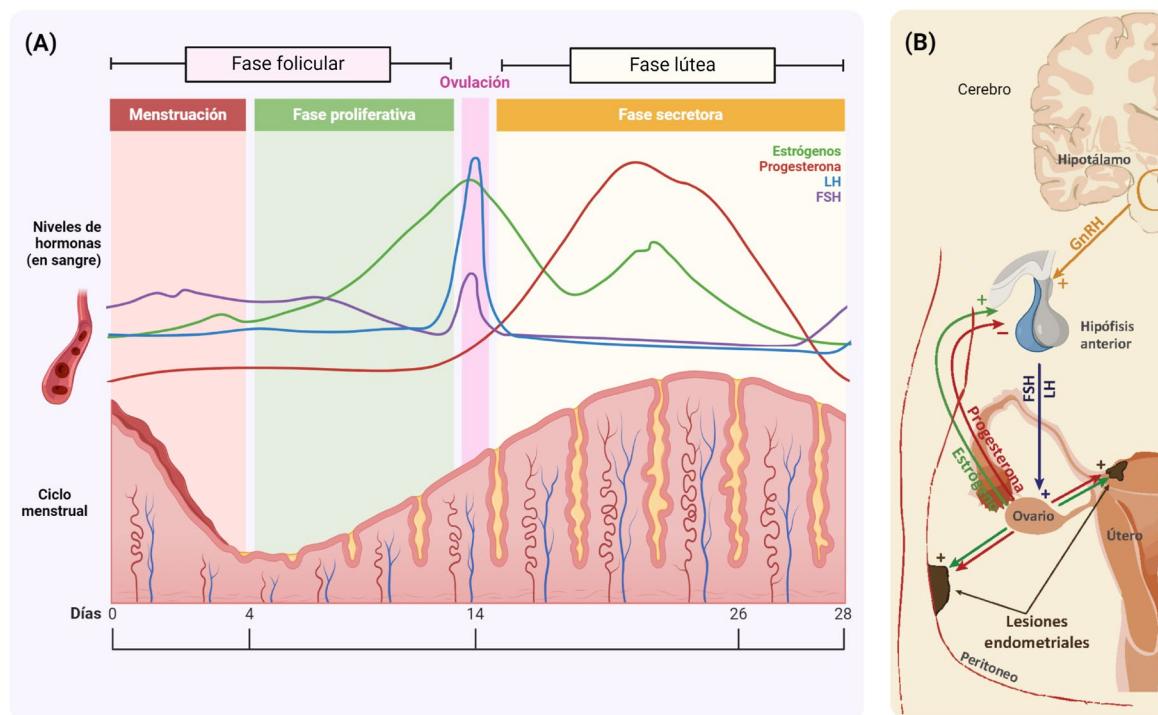
El dolor en la endometriosis es altamente complejo, siendo el resultado de diversos mecanismos que incluyen la inflamación y la percepción del dolor (nocicepción), mediante la interacción entre el sistema nervioso periférico (SNP) y central (SNC) [42,43]. Se ha establecido que la causa del dolor radica en el microambiente inflamatorio de las lesiones endometriales y adherencias, si bien la extirpación y eliminación de estas no garantiza siempre un alivio [4,34,43]. Además, la intensidad del dolor no se relaciona con la puntuación rASRM de las lesiones en el útero, posiblemente debido a la influencia del SNC, el cual puede provocar alteraciones a largo plazo en el procesamiento del dolor, similar a la formación de una "memoria" [34].

La endometriosis, al igual que el endometrio sano, es influenciada por hormonas y experimenta un sangrado cíclico en las lesiones endometriales (**Figura 4**), seguido por una respuesta inflamatoria [42]. En mujeres que no presentan endometriosis, la fase menstrual se caracteriza por la eliminación de los restos del endometrio mediante una respuesta inmune innata, lo que posibilita la regeneración de este tejido. Por el contrario, en mujeres con endometriosis se observa una alteración en esta respuesta inmune, con una disminución en la capacidad fagocítica de los macrófagos [44], que favorece la proliferación del tejido endometrial [34] (**Figura 5**).

La endometriosis es una condición dependiente de estrógenos, particularmente del estradiol (E2), que es esencial para la proliferación y mantenimiento del tejido endometrial tanto eutópico como ectópico [3,45]. En contraste con el endometrio sano, en las lesiones se sobreexpresa la enzima aromatasa, lo que permite la síntesis local de E2 a partir de colesterol. Esto activa la ciclooxygenasa-2 (COX-2) para producir prostaglandinas proinflamatorias E2 (PGE2). Esto conduce a un ciclo de retroalimentación positiva en el que E2 y PGE2 se potencian mutuamente (**Figura 6A**).

Durante la menstruación, se producen y liberan diversos compuestos como resultado de la muerte celular, tales como hierro, especies reactivas de oxígeno (ROS), prostaglandinas y patrones moleculares asociados al peligro (DAMP). En las menstruaciones extra-uterinas, es decir, en las lesiones endometriales, estos se liberan en la cavidad peritoneal junto a los PGE2, que da lugar a acidosis y una respuesta inmune innata exacerbada [34,42]. Por un lado, el hierro, las ROS, la acidosis y las PEG2 de las lesiones estimulan directamente los

nervios sensoriales del SNP [5,42,45,46] (**Figura 6A**). En particular, las ROS pueden alterar el canal iónico del receptor TRPV1 de estos nervios [38] (**Figura 6C**). Por otro lado, los DAMP activan el sistema inmunitario innato [42].

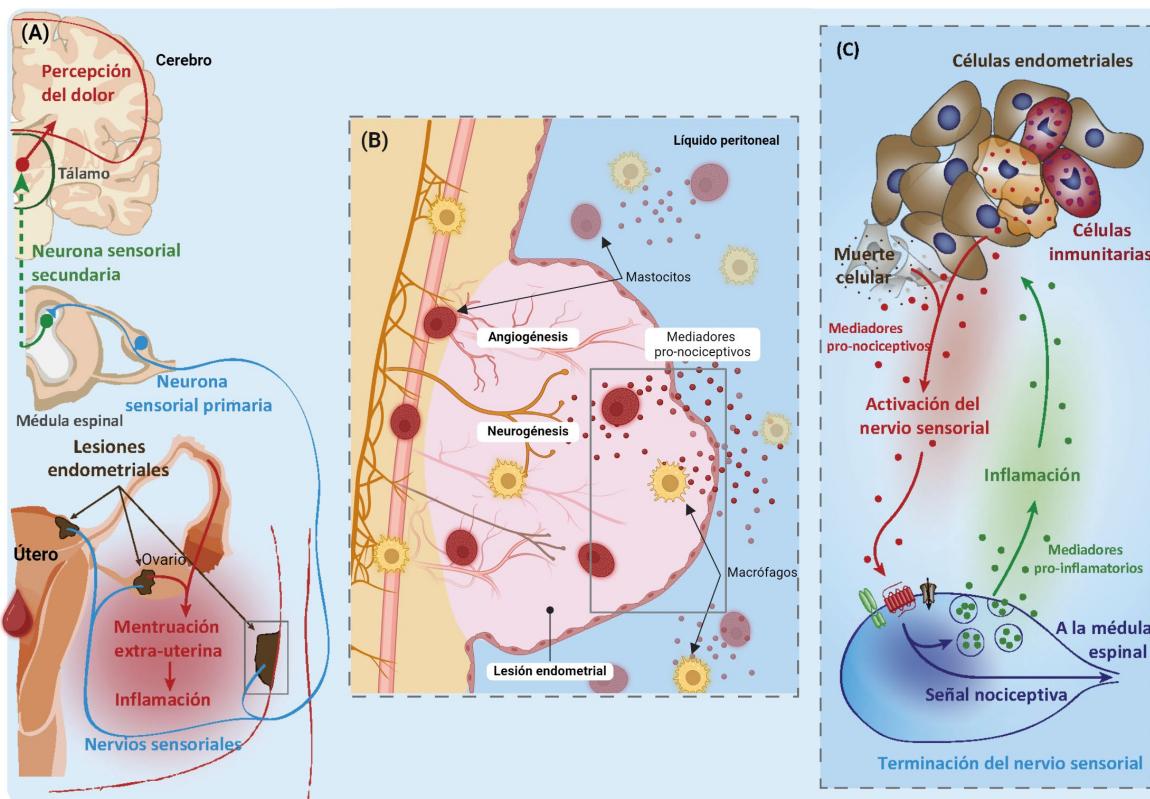


**Figura 4. Efecto hormonal sobre (A) el endometrio sano (Elaborado con BioRender.com) y (B) las lesiones endometriales (Modificado de Laux-Biehlmann et al., 2015 [46]).**

**(A)** Cambios morfológicos que sufre el endometrio sano durante el ciclo menstrual bajo la influencia hormonal. El ciclo menstrual es un proceso cíclico, generalmente de 28 días, que ocurre en la vida fértil de las mujeres y está regulado por las fluctuaciones hormonales. Este ciclo se inicia con la fase menstrual o menstruación de la fase folicular, donde el endometrio, la capa mucosa que recubre el útero, se desprende parcialmente y es expulsado junto con la sangre. Esta fase es seguida por la fase proliferativa, donde los estrógenos inducen el engrosamiento y la regeneración del endometrio. Posteriormente, la fase secretora, también denominada lútea, está dominada por la progesterona, que prepara el endometrio para la posible implantación de un embrión. Si no ocurre la implantación, el ciclo termina con la menstruación. La regulación de estas fases hormonales es controlada por las gonadotropinas (LH y FSH) secretadas por la hipófisis anterior (adenohipófisis), que coordina la secreción de estrógenos y progesterona necesarias para estos cambios cíclicos del endometrio, bajo la influencia de la GnRH del hipotálamo. Durante su vida fértil, una mujer experimenta alrededor de 400 ciclos menstruales [18]. **(B)** Respuesta de las lesiones endometriales a las hormonas del ciclo menstrual, al igual que el endometrio sano, que conlleva a su sangrado. Abreviaciones: LH, hormona luteinizante; FSH, hormona foliculoestimulante; GnRH, hormona liberadora de gonadotropinas.

Las células inmunes innatas, que se pueden infiltrar en el endometrio ectópico, se ven activadas a través de la vía proinflamatoria NF- $\kappa$ B gracias a los DAMP [42]. El complejo NF- $\kappa$ B, un conjunto de factores de transcripción, incrementa la expresión de citoquinas y quimiocinas en estas células. Así, los macrófagos y mastocitos desempeñan un papel

determinante en la generación del dolor al liberar mediadores pro-nociceptivos, como citoquinas proinflamatorias (TNF- $\alpha$ , IL-6, IL-1 $\beta$ , TGF- $\beta$ ), factores de crecimiento nervioso (NGF) y PGE2 [38,46] (**Figura 6B**).

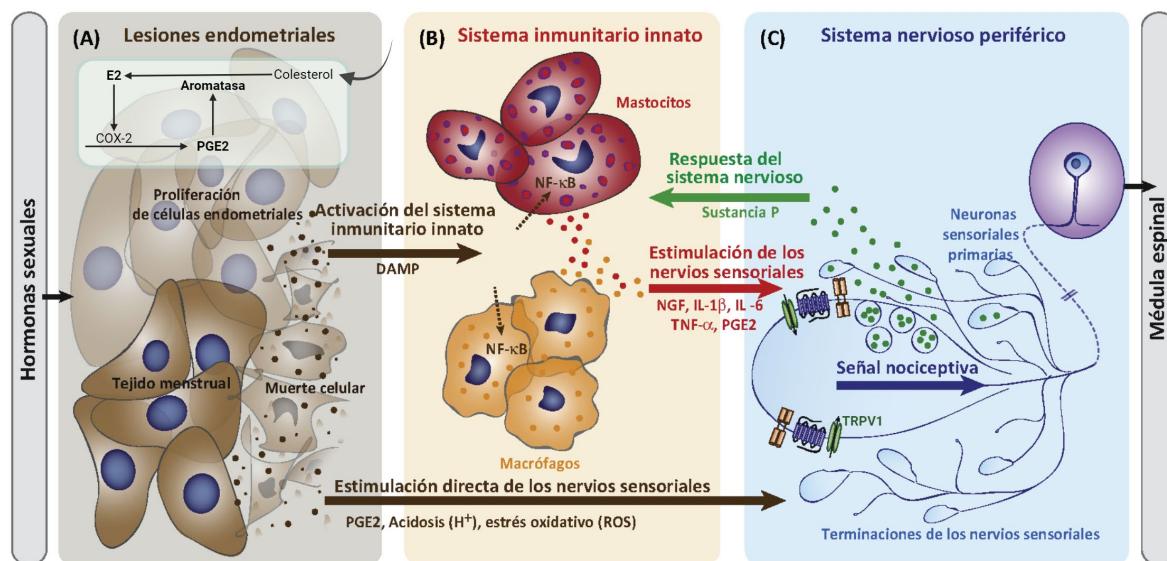


**Figura 5.** Representación esquemática general de los mecanismos del dolor en las lesiones endometriales ((A, C) Modificado de Laux-Biehlmann et al., 2015 [46]; (B) Adaptado de Saunders et al., 2021 [1] con [BioRender.com](#)).

**(A)** Ruta del dolor asociado con lesiones endometriales. El tejido endometrial migra gracias a la menstruación retrógrada hacia la cavidad peritoneal, lo que provoca una menstruación extrauterina e inflamación. Esta inflamación activa las fibras nerviosas sensoriales en la región afectada, generando señales nociceptivas que son transmitidas por las neuronas sensoriales primarias hacia la médula espinal. Posteriormente, estas señales son enviadas al cerebro a través de las neuronas sensoriales secundarias, donde finalmente se percibe la sensación de dolor. **(B)** Ampliación de una lesión endometrial, mostrando la angiogénesis (formación de nuevos vasos sanguíneos) y neurogénesis (crecimiento de nuevos nervios) que se da en este endometrio ectópico. Los mastocitos y macrófagos, presentes en las lesiones y en el líquido peritoneal, liberan mediadores pro-nociceptivos que intensifican la respuesta al dolor y la inflamación. **(C)** Detalle de una parte de una lesión endometrial, ilustrando la interacción entre células endometriales y células inmunitarias. El tejido endometrial inflamado libera mediadores pro-nociceptivos e inflamatorios, que activan los nervios sensoriales locales. Esta activación envía señales nociceptivas hacia la médula espinal, contribuyendo a la percepción del dolor crónico asociado con la endometriosis.

Estos mediadores pro-nociceptivos también estimulan los nervios sensoriales del SNP, al igual que los productos de la muerte celular de células endometriales. Al unirse a los

nociceptores, generan la señal nociceptiva que es transmitida al SNC. La señal nociceptiva se integra en la médula espinal y es transmitida por neuronas sensoriales secundarias al cerebro. Como respuesta, los nervios sensoriales intensifican la inflamación mediante la secreción de mediadores proinflamatorios, como la sustancia P [42,46], lo que desencadena en un aumento en la inflamación y en la nocicepción (**Figura 5C, Figura 6C**).



**Figura 6. Representación esquemática local de los mecanismos del dolor en las lesiones endometriales** (Modificado y adaptado de Laux-Biehlmann et al., 2015 [46]).

**(A)** Las hormonas sexuales, particularmente el E2, promueven la proliferación de células endometriales gracias a una sobreexpresión de la enzima aromatasa por parte de la PGE2. La aromatasa convierte el colesterol en E2, que estimula la COX-2 para seguir produciendo PGE2 y así aumentar la cantidad de E2. Estas prostaglandinas, junto a factores como acidosis y estrés oxidativo (ROS), estimulan directamente los nervios sensoriales del SNP, contribuyendo así a la sensación de dolor. Con la muerte celular en este tejido, se liberan DAMP, que activan el sistema inmunitario innato. **(B)** Los mastocitos y macrófagos son activados y liberan mediadores pro-nociceptivos como NGF, IL-1 $\beta$ , IL-6, TNF- $\alpha$  y PGE2, que estimulan los nervios sensoriales. La respuesta del sistema nervioso incluye la liberación de sustancia P, que amplifica aún más la inflamación y, por tanto, la percepción del dolor. **(C)** Los mediadores pro-nociceptivos estimulan los nervios sensoriales del SNP, al unirse a receptores como TRPV1, que induce señales nociceptivas. Estas son transmitidas a través de las neuronas sensoriales primarias hasta la médula espinal. Abreviaciones: E2, estradiol; PGE2, prostaglandina proinflamatoria E2; COX-2, ciclooxygenasa-2; SNP, sistema nervioso periférico; DAMP, patrones moleculares asociados al peligro.

Además, una vez realizada la implementación de la lesión por menstruación retrógrada, se llevan a cabo, de manera coordinada, los procesos de neurogénesis (formación de nuevas fibras nerviosas) y angiogénesis (formación de nuevos vasos sanguíneos) [24] (**Figura 5B**). Por un lado, el factor de crecimiento endotelial vascular (VEGF) es responsable de la angiogénesis en el tejido ectópico [4], al ser sobreexpresado debido al aumento local de E2

y a los mediadores pro-nociceptivos y proinflamatorios [42]. Por otro lado, la sobreexpresión de genes como el NGF y la IL-1 $\beta$  promueve la neurogénesis para poder transmitir el dolor en la cavidad peritoneal, dado que las lesiones no tienen un suministro nervioso propio [38].

Como resultado, esto conduce a una retroalimentación positiva en que los mediadores pro-nociceptivos y pro-inflamatorios se potencian mutuamente, desencadenando en un ciclo inflamatorio persistente que fomenta la formación de las lesiones endometriales. Se crea de esta manera un microambiente inflamatorio crónico que aumenta la nocicepción, desembocando en dolor crónico [38]. Es importante mencionar que este entorno impide el embarazo y causa la infertilidad en las mujeres con endometriosis [44].

Diversos estudios han observado experimentalmente, por un lado, que las altas concentraciones de PGE2 se asocian con dismenorrea grave, dispareunia y DPC; y niveles altos de TGF- $\beta$ , TNF- $\alpha$  y NGF, con la dismenorrea y dolor intenso [38,42]. Por otro lado, se han encontrado concentraciones elevadas de citocina proinflamatoria IL-1 $\beta$  y sustancia P en las lesiones y en el líquido peritoneal, y de mastocitos en la EP. También se ha determinado una sobreexpresión del canal TRPV1 en endometriomas y lesiones endometriales [34,38,42]. Sin embargo, la mitad de las mujeres con endometriosis no han localizado el dolor en las lesiones, lo que indica una sensibilización general y dolor independiente de la localización de estas [34].

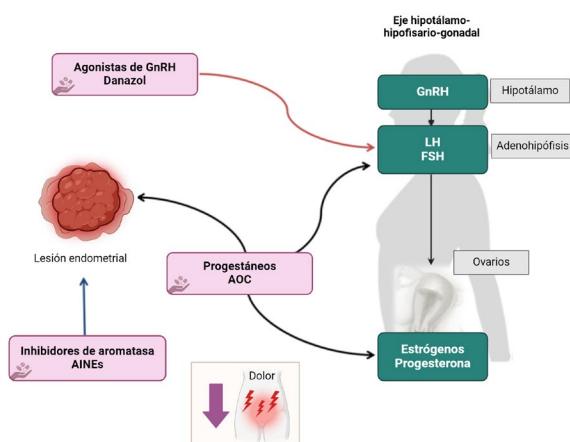
## 1.4. Tratamiento del dolor: enfoque en el tratamiento farmacológico

El tratamiento del dolor en la endometriosis puede variar según su gravedad, la extensión de las lesiones y las preferencias de la paciente. En general, se aconseja iniciar con un tratamiento farmacológico. Para casos graves como EP y endometriomas y que no respondan a esta terapia, es recomendable llevar a cabo un tratamiento quirúrgico, principalmente mediante laparoscopia, para su extirpación [43]. Sin embargo, el dolor crónico puede persistir [34] y las lesiones pueden recurrir en aproximadamente un 50% de los casos al cabo de 5 años [1,4]. Por ello, se recomienda combinar la cirugía con terapia farmacológica, para proporcionar un alivio sintomático del dolor más duradero que la cirugía por sí sola [43].

Como tratamiento farmacológico para la endometriosis y, por consiguiente, del dolor, se consideran terapias hormonales, inhibidores de la producción de aromatasa y fármacos antiinflamatorios no-esteroideos (AINE) [5,8,24,25,29,42,43,45,47,48] (**Figura 7**). Por un lado, los inhibidores de la producción de aromatasa, como Anastrozol y Letrozol [5,42], se utilizan para inhibir la síntesis de prostaglandinas como PGE2 [45]. Estos medicamentos, aunque comúnmente empleados en el tratamiento del cáncer de mama [49], no están aún aprobados para el tratamiento de la endometriosis en España [29,30]. Por otro lado, los AINE, como el Ibuprofeno y el Naproxeno, actúan inhibiendo las ciclooxygenasas (como la COX-2), que son necesarias para la producción de prostaglandinas [24,42]. Estos fármacos son efectivos para disminuir los síntomas de dolor [24], principalmente la dismenorrea [8], proporcionando un alivio significativo a muchas pacientes.

Por último, las terapias hormonales tienen como finalidad impedir la proliferación del tejido endometrial al bloquear la menstruación y, como consecuencia, disminuir el dolor. Este tratamiento puede incluir anticonceptivos orales combinados (AOC), progestágenos (también denominados progestinas), Danazol, o agonistas de la hormona liberadora de gonadotropinas (GnRH), entre otros [47,48] (**Figura 7**). Por una parte, estos pueden inhibir la producción de hormonas gonadotrópicas (como Danazol) o desensibilizar los receptores de la adenohipófisis para la GnRH (como los agonistas de la GnRH) para reducir la producción de las hormonas sexuales femeninas [5,25,47,48]. Por otra parte, pueden inhibir directamente la producción de estrógeno (como los AOC y progestágenos) o progesterona (como los progestágenos), lo que resulta en una disminución de la producción de prostaglandinas y, por consiguiente, de la inflamación y dolor [8,25,43,47,48].

No obstante, la mayoría de las veces, a pesar de la aplicación de una variedad de tratamientos farmacológicos, solo se logra reducir el dolor en el 40-70% de las pacientes [34]. Aun así, el tratamiento



**Figura 7.** Tratamientos farmacológicos usados para disminuir el dolor y sus dianas (Modificado de [48] con [BioRender.com](#)). Abreviaciones: AINEs, fármacos antiinflamatorios no-esteroideos; AOC, anticonceptivos orales combinados; LH, hormona luteinizante; FSH, hormona foliculoestimulante; GnRH, hormona liberadora de gonadotropinas.

principal para la endometriosis y los síntomas de dolor a largo plazo, independientemente del tipo de endometriosis (superficial, endometrioma o EP) [47], son las terapias hormonales, especialmente los progestágenos [47,48].

Por lo tanto, el impacto del tratamiento hormonal no se limitaría únicamente a la reducción de los síntomas, sino que puede influir en la expresión de genes implicados en la nocicepción [48]. Se han identificado varios genes que destacan en la percepción del dolor en la endometriosis, tales como *TRPV1*, *NGF*, *IL1 $\beta$* , *BDNF*, *NTRK1*, entre otros [38]. Estos genes están involucrados en la modulación de la inflamación y la transmisión del dolor en diversos tipos celulares, incluyendo macrófagos, mastocitos y células estromales, como se ha comentado anteriormente en el apartado 1.3. Estos genes, que forman parte de una red más amplia de genes validados en el contexto del dolor, como recoge la revisión de Wistrom *et al.* [50], podrían verse afectados por este tratamiento farmacológico.

## 1.5. La tecnología de *single cell RNA sequencing*

El dogma central de la expresión génica establece un flujo unidireccional de información genética, donde el ADN es transcritto a ARN mensajero (ARNm) y posteriormente traducido a proteínas, las cuales desempeñan las funciones biológicas en las células [51]. La transcripción puede generar diversos tipos de trascritos además del ARNm. El conjunto de todos los trascritos de una célula en unas condiciones y momento determinados recibe el nombre de transcriptoma [52]. Este, a diferencia del genoma, es un componente dinámico que puede variar en función de las condiciones internas y externas de la célula. Esta característica ha impulsado al desarrollo de nuevas tecnologías para su estudio [53]. De esta manera, nace la transcriptómica que permite una comprensión indirecta de la expresión génica y sus efectos en la función celular. Este campo es fundamental para comprender las enfermedades humanas [53,54].

En la década de los 90 se comenzaron a utilizar los *microarrays*, que permiten cuantificar miles de genes simultáneamente gracias a la hibridación por similitud [55]. No obstante, su capacidad de detección es limitada y únicamente posibilita el análisis de genes previamente identificados [52]. Es por ello que, a día de hoy, gracias a la secuenciación de próxima generación (*Next Generation Sequencing*, NGS), la tecnología más empleada es la secuenciación de ARN (RNA-seq) [52,56]. A diferencia de los *microarrays*, esta técnica

demandó una cantidad reducida de ARN, presenta una sensibilidad elevada y una baja tasa de error, evita las complicaciones derivadas de la hibridación cruzada, identifica incluso genes con bajo umbral de expresión y exhibe una alta reproducibilidad técnica [52–54].

No obstante, la técnica convencional de RNA-seq, también denominada secuenciación *bulk*, únicamente puede calcular la expresión génica promedio de un conjunto de células. La heterogeneidad celular y la expresión específica de subconjuntos de células no son tenidas en cuenta [57]. Desde 2009, la técnica de secuenciación de ARN de célula única o *single cell* (scRNA-seq) se encarga de enfrentar esta dificultad al posibilitar la evaluación del transcriptoma a escala individual de las células, tanto en estado de salud como enfermedad [57,58].

El procedimiento de scRNA-seq incluye primero un flujo de trabajo en el laboratorio y después, un flujo de trabajo computacional para el análisis de los datos obtenidos [59]. Aunque hay varios métodos de scRNA-seq, la opción que más se está empleando, debido a su alto rendimiento y bajo costo, es el método 10x Genomics [57]. El trabajo de laboratorio consiste principalmente en la construcción de librerías *single cell* [59]. Para realizar la captura celular, primero se marca el transcriptoma (paso de multiplexado), que consiste en un código de barra celular y un identificador único molecular (UMI) por cada transcripto. Posteriormente, se lleva a cabo una transcripción inversa para convertir el ARN en ADN complementario (ADNc) y, en última instancia, amplificarlo [58–60]. Los datos crudos de la secuenciación se obtienen en formato FASTQ y contienen las secuencias leídas (conocidas como lecturas o *reads*, en inglés).

Para el flujo de trabajo bioinformático, primero es necesario preprocesar estos archivos hasta crear una matriz de conteo, que contiene el recuento de trascritos por célula [59]. Este proceso se puede realizar empleando el programa Cell Ranger para los datos obtenidos en la plataforma 10x Genomics [59]. El preprocesamiento incluye el alineamiento de las lecturas crudas contra un genoma de referencia, agrupadas por genes y asignadas a su código de barras celular correspondiente [58,59]. Después, es imprescindible realizar el análisis global y más específico de los datos de scRNA-seq a partir de la matriz de conteo con el fin de obtener la información sobre la expresión génica [58,59,61,62]. Estos pasos se desarrollan en el apartado de **Metodología**. En la actualidad, debido a que es una tecnología reciente, no existe un protocolo estándar establecido para este fin. Sin embargo, generalmente se utilizan protocolos de buenas prácticas [61,62].

La técnica de scRNA-seq está siendo utilizada en la investigación de la endometriosis para identificar las características moleculares y celulares particulares del síndrome [63,64]. Esto podría facilitar el desarrollo de un tratamiento efectivo que prevenga la recurrencia del tejido endometrial después de la extirpación y alivie el dolor asociado. En cuanto al genoma, se ha realizado una revisión que ha identificado 242 genes relacionados con la sensación de dolor en general [50]. Entre estos genes se encuentran genes como *NGF* y *TRPV1* que participan en la percepción de este síntoma en la endometriosis [38] (apartado 1.3). Recientemente, se han identificado cinco genes más asociados a la percepción y el mantenimiento del dolor específicamente en este síndrome: *SRP14/BMF*, *GDAP1*, *MLLT10*, *BSN* y *FSHB* [65]. El hecho de que estos genes no solo estén relacionados con la percepción del dolor en general, sino que también sean modulados en condiciones inflamatorias, sugiere que el tratamiento hormonal podría influir en su expresión y, por ende, en la intensidad del dolor experimentado por las pacientes con endometriosis. Se plantea la hipótesis de que la terapia hormonal, que es el tratamiento inicial para el síntoma [47,48], puede disminuir la expresión de genes relacionados con el dolor. Esta suposición podría ser confirmada mediante el análisis de datos scRNA-seq.

## 2. Objetivos

### Objetivo general:

- Identificar el impacto de la terapia hormonal en la expresión de genes asociados al dolor en la endometriosis mediante el análisis de datos de scRNA-seq.

### Objetivos específicos:

- Procesamiento y análisis comparativo de dos conjuntos de datos públicos de scRNA-seq, uno con tratamiento hormonal y otro compuesto mayoritariamente por muestras sin tratamiento, para visualizar la expresión de genes asociados al dolor en la endometriosis.
- Análisis de la expresión de genes relacionados con el dolor en datos de scRNA-seq en pacientes con endometriosis con y sin tratamiento hormonal para identificar el impacto de la terapia en las células que componen el endometrio eutópico y ectópico.

# 3. Metodología

## 3.1. Obtención de datos públicos de scRNA-seq

Se obtuvieron datos públicos de scRNA-seq del repositorio *Gene Expression Omnibus* (GEO), correspondientes a pacientes con endometriosis, con y sin tratamiento hormonal. Se generaron dos conjuntos de datos: *dataset A* ([GSE179640](#)), publicado por Tan *et. al* [63], que incluye muestras de pacientes con tratamientos hormonales, y *dataset B* ([GSE213216](#)), publicado por Fonseca *et. al* [64], que contiene sobre todo muestras de pacientes sin tratamiento hormonal. Ambos estudios realizaron la secuenciación en la plataforma 10X Genomics, en Illumina Novaseq 6000.

Se descargaron los archivos con las matrices de conteo filtradas con Cell Ranger (versión 3.1.0) en formato H5 utilizando un equipo de 16 GB de RAM. Para el *dataset A*, se descargaron 17 matrices de conteo de muestras de endometrio, ovarios y peritoneo de 13 pacientes (14 muestras de pacientes con endometriosis (1 sin y 13 con tratamiento hormonal) y 3 controles (1 sin y 2 con tratamiento hormonal)) (**ANEXO I**). Para el *dataset B*, se descargaron 43 archivos correspondientes a muestras de endometrio, ovario, recto y otras localizaciones de 19 pacientes (40 de pacientes con endometriosis (34 sin y 6 con tratamiento hormonal) y 3 controles (1 sin y 2 con tratamiento hormonal)) (**ANEXO II**, **Figura 8**).

Los metadatos se obtuvieron a partir de los archivos correspondientes subidos a *SRA Run Selector* de cada proyecto y de la información específica de cada paciente suplementada en cada artículo científico (**ANEXO I**, **ANEXO II**).

## 3.2. Preprocesamiento de datos de scRNA-seq

En este trabajo se ha empleado el lenguaje de programación R (versión 4.4.1) para llevar al cabo el flujo de trabajo bioinformático para analizar los datos de scRNA-seq de pacientes con endometriosis (**Figura 8**).

Se importaron las matrices de conteo de cada conjunto de datos en R utilizando la función `Read10X_h5()` del paquete `hdf5r` (versión 1.3.11). El paquete `Seurat` (versión 5.1.0) [66,67] se utilizó para explorar los datos scRNA-seq. Se crearon objetos `Seurat` para cada *dataset* con la función `CreateSeuratObject()`, donde se establecieron los *barcodes* de las células en las columnas y los genes en las filas. Posteriormente, se calcularon los porcentajes de genes mitocondriales y ribosomales, y se añadieron los metadatos (**ANEXO III**).

Para garantizar la viabilidad de las células, se filtraron las que presentaban características indicativas de daño o apoptosis, como un alto porcentaje de genes mitocondriales y un bajo recuento génico total [68], según parámetros implementados por cada equipo de investigación.

En el *dataset A*, se eliminaron las células con menos de 500 genes y más del 25% de contenido mitocondrial, y se incluyeron aquellas entre 1.000 y 100.000 UMIs [63]. En el *dataset B*, se descartaron células con menos de 200 genes y más del 20% de contenido mitocondrial [64]. Adicionalmente, se eliminaron los dobletes, generados cuando dos células son encapsuladas en la misma gota durante la preparación de la librería [68], utilizando `DoubletFinder` [69] (versión 2.0.4) (**ANEXO IV**).

Se normalizaron los datos de expresión génica con `NormalizeData()` de `Seurat` para mitigar efectos técnicos y mejorar la comparabilidad entre células. La fase celular de cada célula (G1, S o G2M) fue asignada con `CellCycleScoring()`, para controlar variaciones en la expresión génica relacionadas con la fase del ciclo celular. Después, se identificaron genes con mayor variabilidad mediante `FindVariableFeatures()`, seleccionando los 2.000 genes más variables. Posteriormente, se realizó el escalado de los datos con `ScaleData()` para normalizar la expresión génica a puntuaciones Z (**ANEXO V**).

Se aplicó una reducción de dimensionalidad utilizando Análisis de Componentes Principales (PCA) [68] con `RunPCA()` con las características más variables seleccionadas previamente. El número de componentes principales se determinó con `ElbowPlot()` [70]. Finalmente, se realizó el agrupamiento de células mediante `FindNeighbors()` y `FindClusters()` (con resoluciones de 0.4, 0.6, 0.8, 1 y 1.4) de `Seurat`, seguido de la reducción dimensional con `RunUMAP()` para visualizar los clústeres celulares resultantes (**ANEXO V**).

## 3.3. Procesamiento y análisis de datos de scRNA-seq

### 3.3.1. Anotación

Se identificaron los genes diferencialmente expresados (DEG) en cada clúster utilizando la función `wilcoxauc()` del programa Presto (versión 1.0.0) por su rapidez. Posteriormente, se seleccionaron los 10 genes principales mediante `top_markers()` y, empleando estudios previos de marcadores en células del endometrio sano y lesiones endometriales [63,64,71], se caracterizaron los tipos celulares del endometrio eutópico y ectópico (**ANEXO VI**). Se eliminaron aquellos clústeres de mala calidad (dobletes o células de mala calidad).

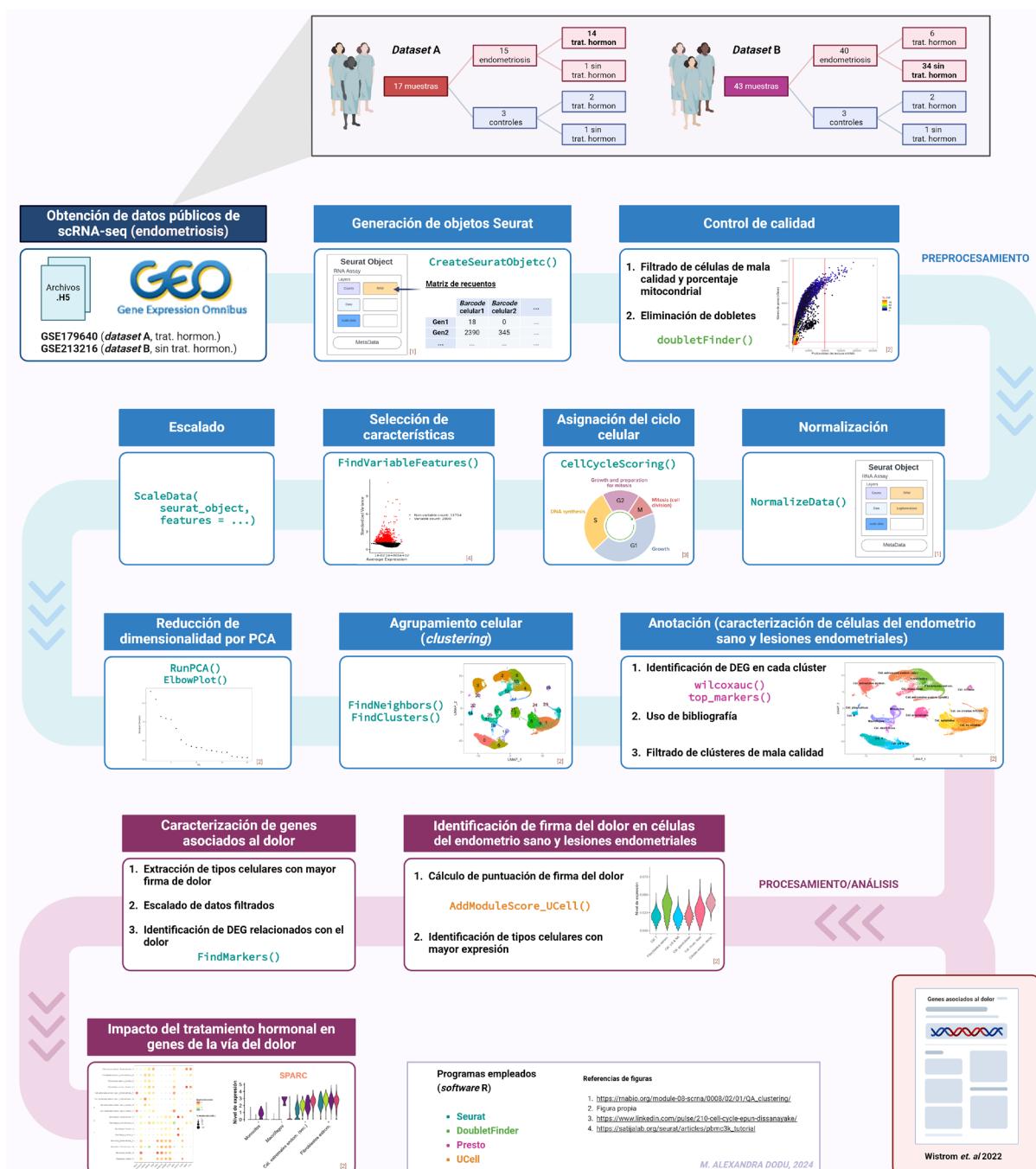
### 3.3.2. Identificación de firma del dolor en tipos celulares y evaluación del impacto del tratamiento hormonal

Se procedió a generar una firma de dolor utilizando genes previamente asociados al síntoma de dolor en endometriosis, basados en la revisión de Wistrom *et al.* [50] y estudios adicionales [65]. Para calcular el puntaje de expresión de esta firma en los diferentes tipos celulares, se empleó la función `AddModuleScore_UCell()` del programa UCell (versión 2.8.0) (**ANEXO VIII**).

Posteriormente, se identificaron los tipos celulares con mayor expresión de los genes asociados al dolor para extraerlos. A partir de esta información, se escaló nuevamente y se determinaron los 5 DEG principales en cada clúster después del empleo de `FindMarkers()` de Seurat. Después, se volvió a escalar la expresión génica, aunque solo de estos genes (**ANEXO VIII**). Finalmente, los genes asociados al dolor de cada *dataset* se visualizaron mediante mapas de calor, mapas de puntos, y violín *plots*, permitiendo evaluar tanto su expresión como el impacto del tratamiento hormonal en dicha expresión.

### 3.3.3. Visualización

Para la visualización de los resultados se empleó `ggplot2` (versión 3.5.1.) y funciones específicas de Seurat como `ElbowPlot()` para los gráficos de codo, `DimPlot()` para los gráficos UMAP, `DimHeatmap()` para los mapas de calor y `DotPlot()` para gráficos de puntos (**ANEXO VII, ANEXO VIII**).



**Figura 8.** Flujo de trabajo bioinformático para el análisis de datos públicos de scRNA-seq de pacientes con endometriosis con y sin tratamiento hormonal (Elaborado con [BioRender.com](#)). Este diagrama ilustra los pasos clave seguidos en el procesamiento y análisis de datos de scRNA-seq de 60 muestras obtenidas de pacientes con endometriosis, diferenciando entre aquellos con (dataset A) y sin tratamiento hormonal (dataset B). El flujo de trabajo abarca las fases de preprocessamiento (obtención de datos públicos, generación de objetos Seurat, control de calidad, normalización, selección de características variables, asignación del ciclo celular, reducción de dimensionalidad por Análisis de Componentes Principales (PCA), agrupamiento celular (o clustering) y anotación celular) y procesamiento y análisis de los datos (caracterización de tipos celulares y genes asociados al dolor e Impacto del tratamiento hormonal). Figura en alta resolución en este [enlace](#).

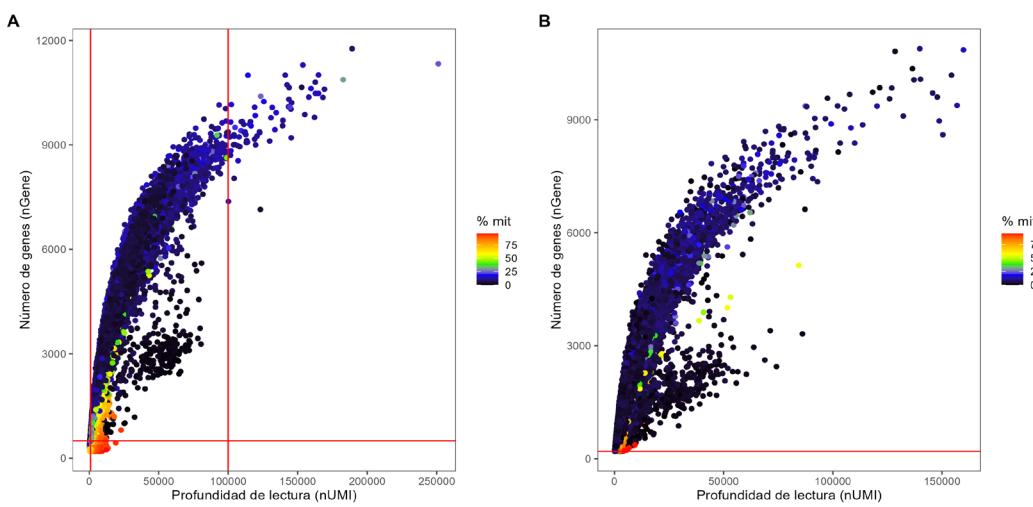
# 4. Resultados

## 4.1. Descarga de datos públicos de scRNA-seq de pacientes con endometriosis

Se descargaron directamente desde GEO 60 archivos con las matrices de conteo filtradas por Cell Ranger por cada equipo de investigación en formato H5, habiendo descartado previamente seis matrices de conteo para el *dataset* B.

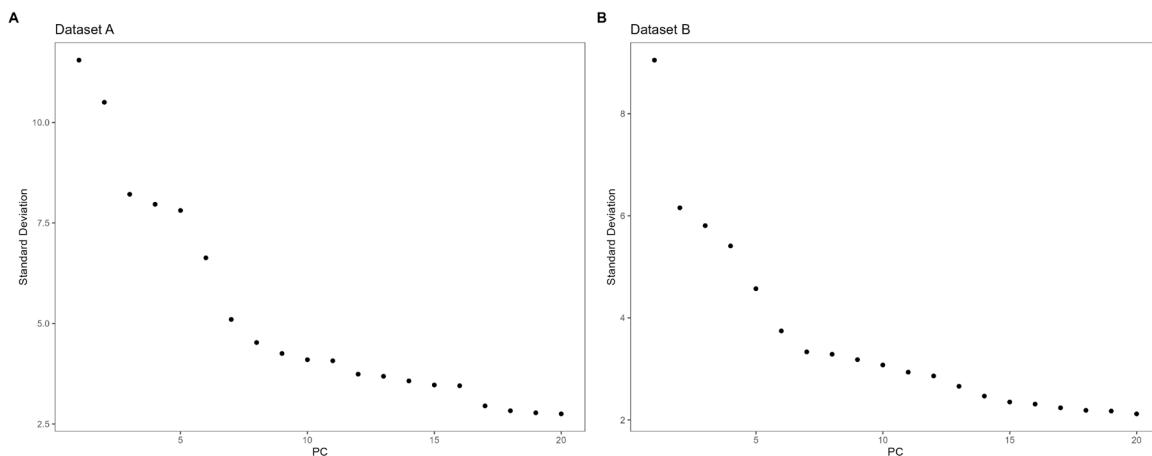
## 4.2. Preprocesamiento de datos de scRNA-seq

En ambos conjuntos de datos de scRNA-seq provenientes de pacientes con endometriosis, se llevó a cabo el proceso de filtrado para excluir aquellas células que presentaban un alto porcentaje de genes mitocondriales y un bajo recuento génico total (**Figura 8**, ver **Metodología** para los parámetros empleados). Además, se utilizó la herramienta DoubletFinder en ambos conjuntos para detectar y eliminar cualquier doblete remanente. Como resultado de este filtrado, en el *dataset A* se eliminaron 18.733 células, dejando un total de 59.890 células de alta calidad. En el *dataset B*, se descartaron 52.883 células, lo que resultó en 311.058 células finales para el análisis.



**Figura 9.** Representación de la cantidad de genes (nGene) frente a la cantidad de UMI coloreadas por el porcentaje de lecturas mitocondriales en el *dataset A* (A) y *dataset B* (B). Los puntos con colores cálidos indican un alto porcentaje de genes mitocondriales. Las líneas rojas indican los umbrales utilizados para el filtrado, eliminando las células con alto porcentaje de genes mitocondriales o baja y alta profundidad de lectura.

Después de la normalización de los datos, se escaló la expresión de los 2.000 genes altamente variables en el *dataset B*, mientras que en el *dataset A* se escaló la de todos los genes. Posteriormente, se redujo la dimensionalidad de los conjuntos de datos mediante PCA. A partir de los *elbow plots*, se determinaron de manera subjetiva los primeros 10 PCs, de un total de 20 dimensiones, como suficientes para capturar la mayor parte de la variabilidad presente en ambos conjuntos de datos (**Figura 10**).



**Figura 10.** *Elbow plots* o gráficos de codo del *dataset A* (A) y *dataset B* (B).

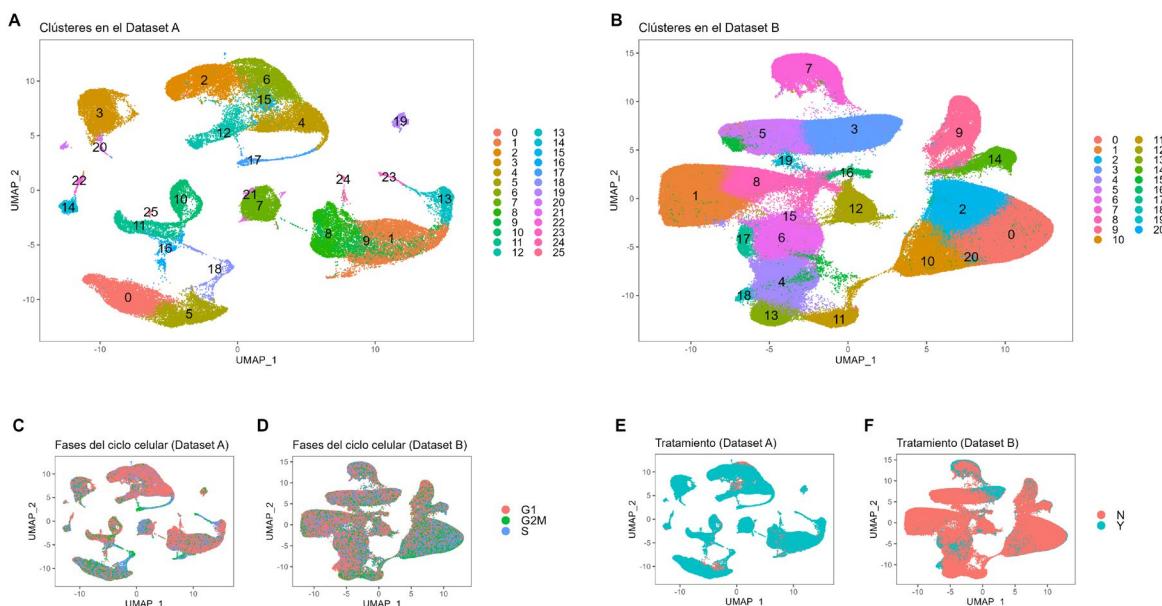
Se visualiza la desviación estándar de cada componente principal (PC). El lugar donde aparece el codo suele ser el umbral para identificar la mayor parte de la variación, en este caso, a partir de 7-8 PC. Cada PC captura una porción de la desviación estándar total de los datos, que es una medida de la variabilidad. El PC1, o primer componente principal, explica la mayor desviación estándar posible y representa la mayor parte de la variabilidad en los datos, dicho de otra forma, representa la mayor diferencia en la expresión génica entre las muestras. Los componentes principales se ordenan de manera que los primeros PCs capturan la mayor variabilidad, mientras que los PCs de menor rango explican menos de esta variabilidad [70]. Al incluir PCs más allá del punto de codo, el riesgo es añadir ruido en lugar de información relevante.

## 4.3. Caracterización de células del endometrio sano y lesiones endometriales

Se identificaron 25 clústeres en el *dataset A* y 20 en el *dataset B* (**Figura 11 A, B**) al emplear una resolución de 0.4. Esta resolución fue suficiente para identificar los tipos celulares que intervienen en la vía del dolor en pacientes con endometriosis. La anotación se realizó teniendo en cuenta los 10 DEG principales de cada grupo (**ANEXO IX, ANEXO X**). Se descartaron clústeres de mala calidad: clúster 18 del *dataset A* y clúster 15 del *dataset B*.

Con la anotación, se identificaron 18 subtipos celulares en el *dataset A* y 15 en el *dataset B* (**Figura 12 A, B**), los cuales correspondieron a seis tipos celulares generales del endometrio

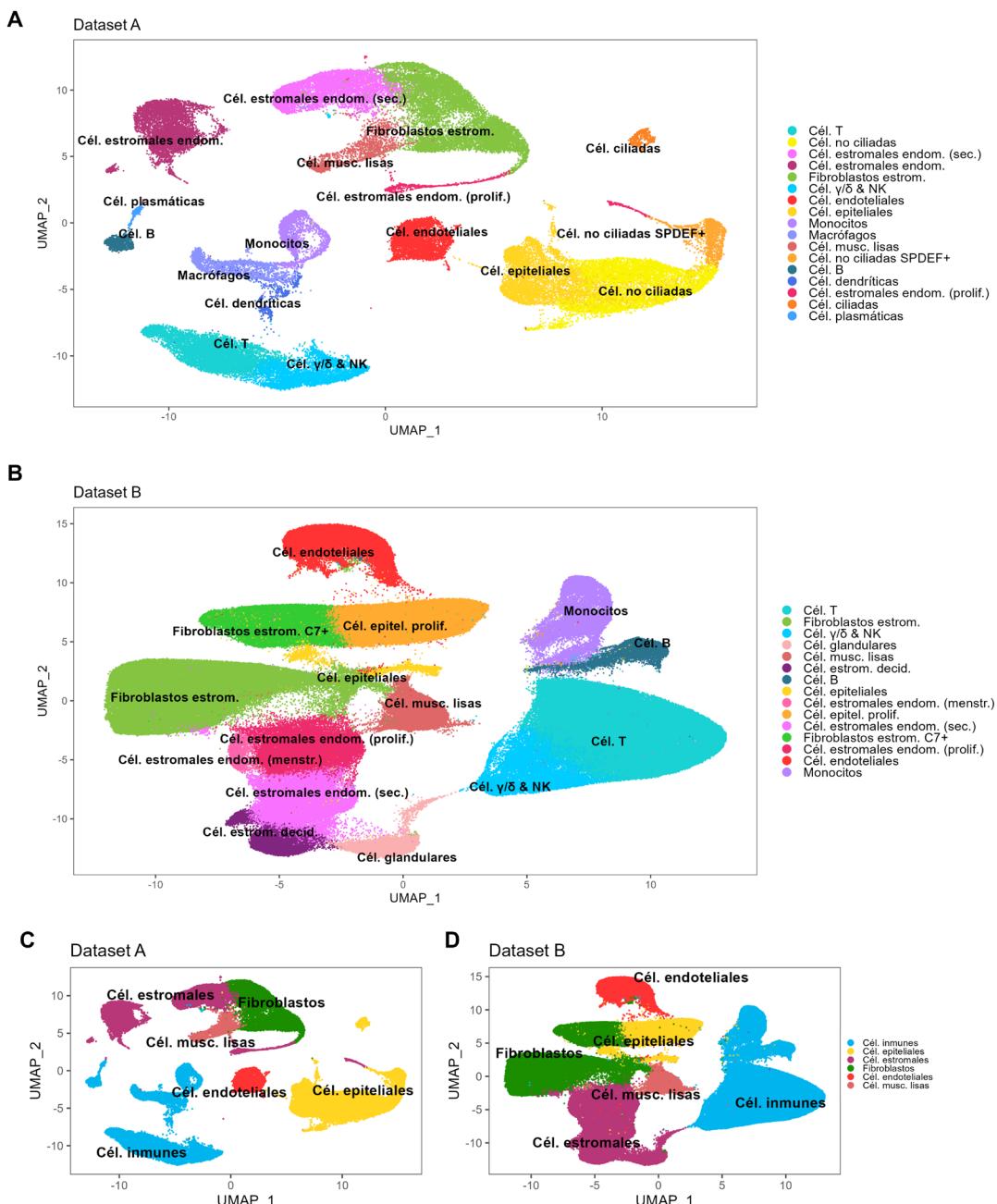
y lesiones endometriales: células inmunes, estromales, epiteliales, musculares lisas, endoteliales y fibroblastos (**Figura 12 C, D**).



**Figura 11.** Agrupación de células del endometrio eutópico y ectópico.

**A, B.** Representación UMAP de los clústeres identificados en los conjuntos de datos A y B. Los números indican los clústeres específicos y los colores corresponden a diferentes grupos celulares. **C, D.** Mapas UMAP que muestran la distribución de las células según las fases del ciclo celular (G1, S, G2M) en los conjuntos de datos A y B. **E, F.** Mapas UMAP que representan las células tratadas (Y: Yes) y no tratadas (N: No) en los conjuntos de datos A y B. Se puede observar cómo en el dataset A hay más células con tratamiento hormonal, mientras que, en el B, lo contrario. Figura en alta resolución en este [enlace](#).

En cuanto a los subtipos celulares, por un lado, en el *dataset* A, los fibroblastos estromales, identificados por la expresión de *COL1A1*, *COL3A1* y *LUM*, fueron el subtipo celular con mayor representación (10.007 células, 16,88% de las 59.352 células restantes después de la eliminación del clúster 18). Las células T (*CD3D* y *CD3E*) fueron el segundo tipo celular más prevalente, con 6.691 células (11,29% de las células). Las células epiteliales (*EPCAM*) y no ciliadas (*GABRP*) representaron cada una más de un 10,20% de las células (>6.050 células/cada tipo celular). Entre los subtipos celulares menos comunes se hallaron células plasmáticas, identificadas por la expresión de *CD79A* (280 células, 0,47% de las células), y células epiteliales ciliadas (*CAPS*, *LRRIQ1* y *C5ORF49*) (483 células, 0,82% de las células) (**Figura 13A**). Por otro lado, en el *dataset* B, el subtipo celular más prevalente fueron las células T (70.574 células, 23,02% del total de 276.060 células restantes después del filtrado de los clústeres 3 y 15), seguido de los fibroblastos estromales (54.509 células, 17,78% de las células). Entre los subtipos celulares menos comunes se hallaron las células estromales

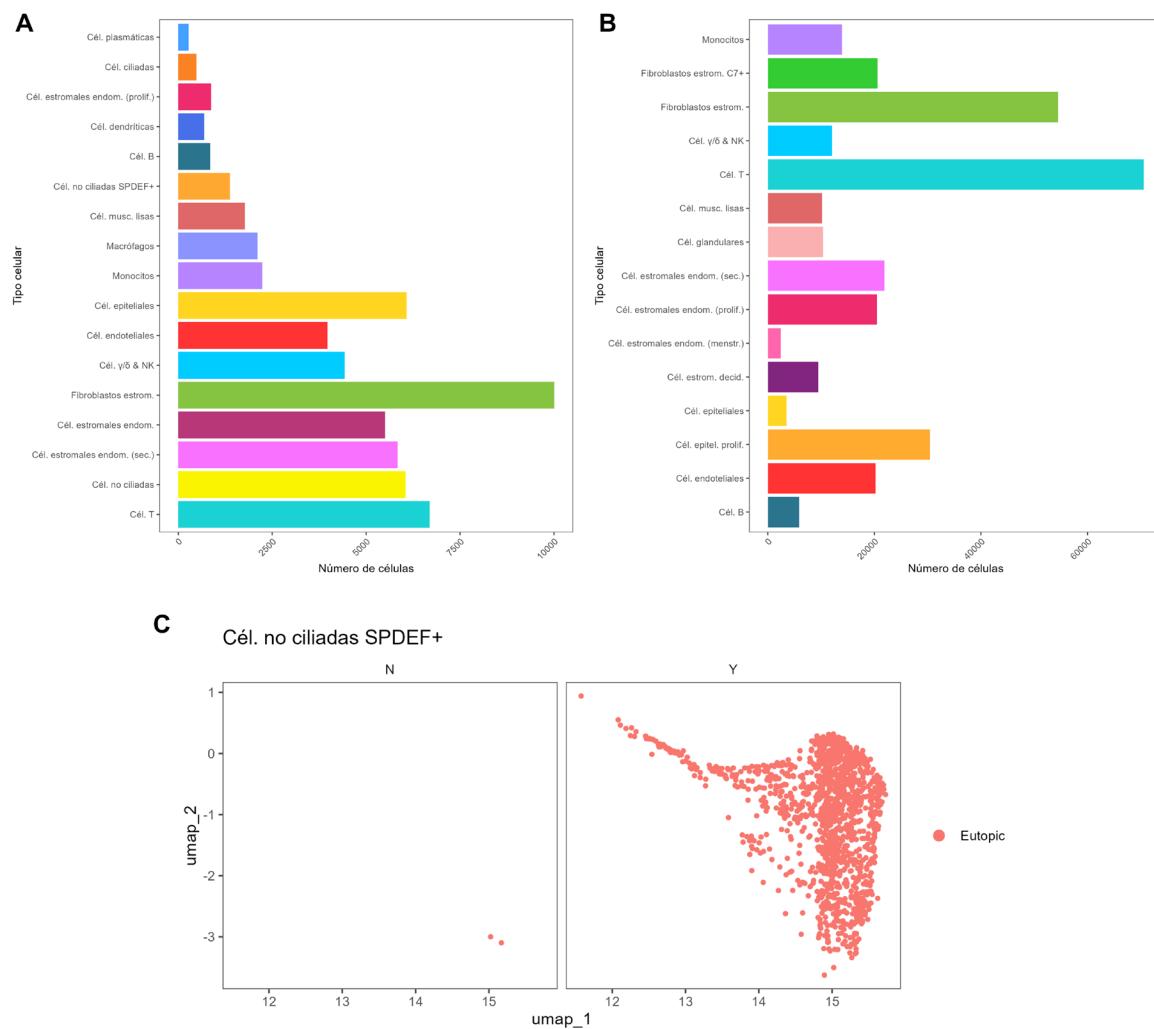


**Figura 12. Caracterización de células del endometrio eutópico y ectópico I.**

**A, B.** Análisis de UMAP de la anotación de clústeres celulares en los dos datasets. **A.** En el Dataset A, se identificaron 18 tipos celulares: células T (Cél. T), células no ciliadas (Cél. no ciliadas), células estromales endometriales (Cél. estromales endom.), que se subdividen en células secretoras (sec.) y células proliferativas (prolif.). Además, se detectaron fibroblastos estromales (Fibroblastos estrom.), células musculares lisas (Cél. musc. lisas), células epiteliales, células plasmáticas (Cél. plasmáticas), células B, macrófagos, monocitos, células dendríticas (Cél. dendríticas), células endoteliales, células  $\gamma/\delta$  y Natural Killers (Cél.  $\gamma/\delta$  & NK) y células no ciliadas SPDEF+, que son aquellas con alta expresión del factor de transcripción SPDEF. **B.** En el Dataset B, se identificaron 15 tipos celulares: células T, fibroblastos estromales (con una subpoblación marcada con C7+), células estromales deciduales (Células estrom. decid.), células glandulares, células epiteliales proliferativas (Cél. epitel. prolif.), células estromales endometriales en diferentes estados (menstruantes y secretoras), células epiteliales, células musculares lisas, células  $\gamma/\delta$  y NK, monocitos y células B. En estos mapas, cada punto representa una célula individual, y la proximidad entre puntos sugiere similitudes en las características de expresión génica entre las células. **C, D.** Representación de los seis tipos generales de células en ambos conjuntos de datos: células inmunes, estromales, epiteliales, endoteliales, musculares lisas y fibroblastos. Figura en alta resolución en este [enlace](#).

endometriales en la fase de menstruación (2.400 células, 0,78% de las células) y células epiteliales (*S100A11*) (3.506 células, 1,14% de las células) (**Figura 13B**).

Se observó también la presencia únicamente en el tejido eutópico con tratamiento hormonal del *dataset A* de la subpoblación de células epiteliales no ciliadas con expresión de *SPDEF* (**Figura 13C**). En el *dataset B*, se identificaron las subpoblaciones de fibroblastos con expresión de *C7*, células decidualizadas y células endometriales en fase de menstruación, no observadas en el *dataset A* (**Figura 12**).

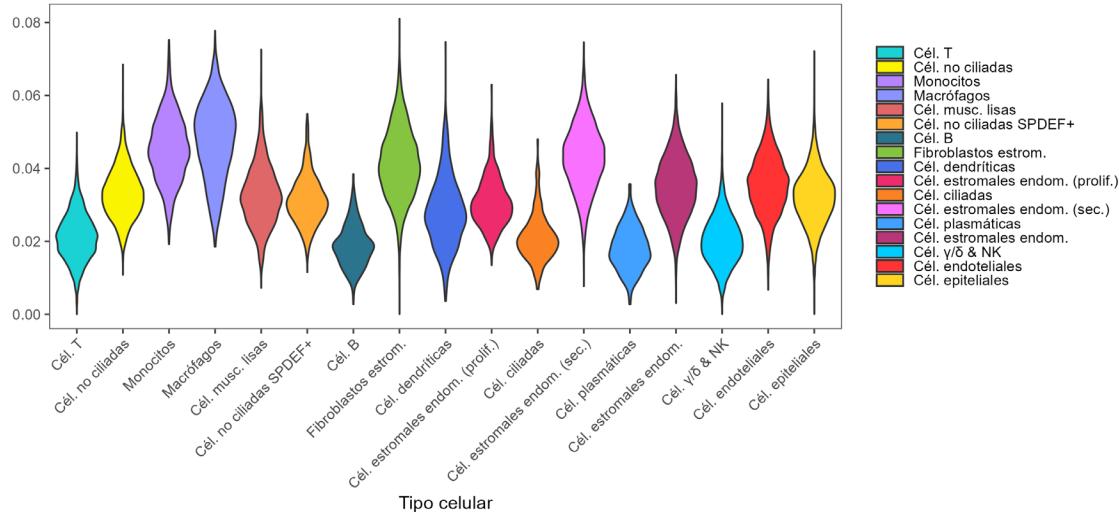


**Figura 13. Caracterización de células del endometrio eutópico y ectópico II.**

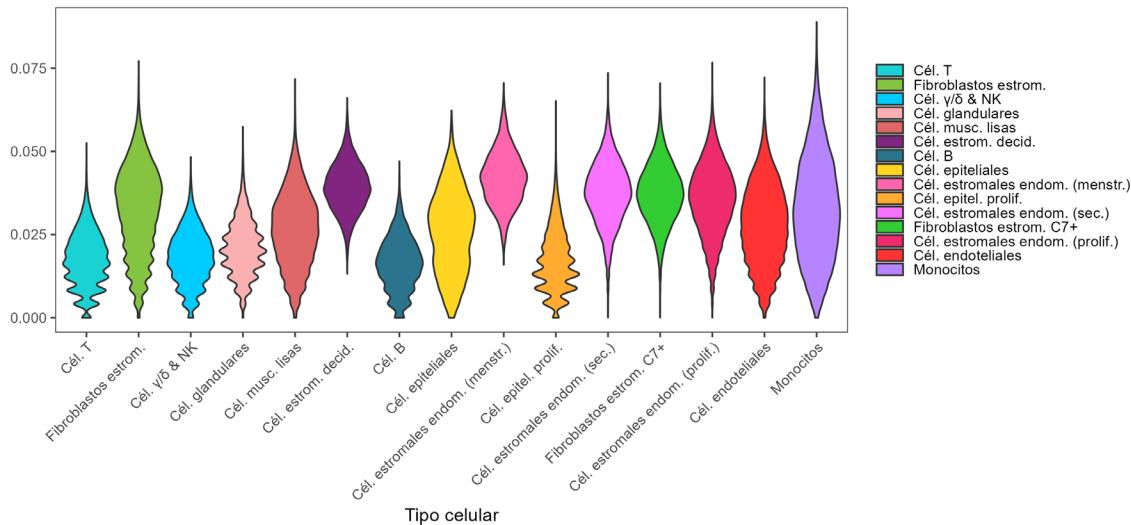
Número de células por tipo celular del endometrio eutópico y ectópico del **(A)** conjunto A y **(B)** B, y **(C)** caracterización del subtipo celular de células epiteliales no ciliadas con expresión de *SPDEF* en el *dataset A*. Este grupo se halla en el endometrio eutópico con tratamiento hormonal. Figura en alta resolución en este [enlace](#).

**A**

Expresión de la firma de genes del dolor por tipo celular (Dataset A)

**B**

Expresión de la firma de genes del dolor por tipo celular (Dataset B)



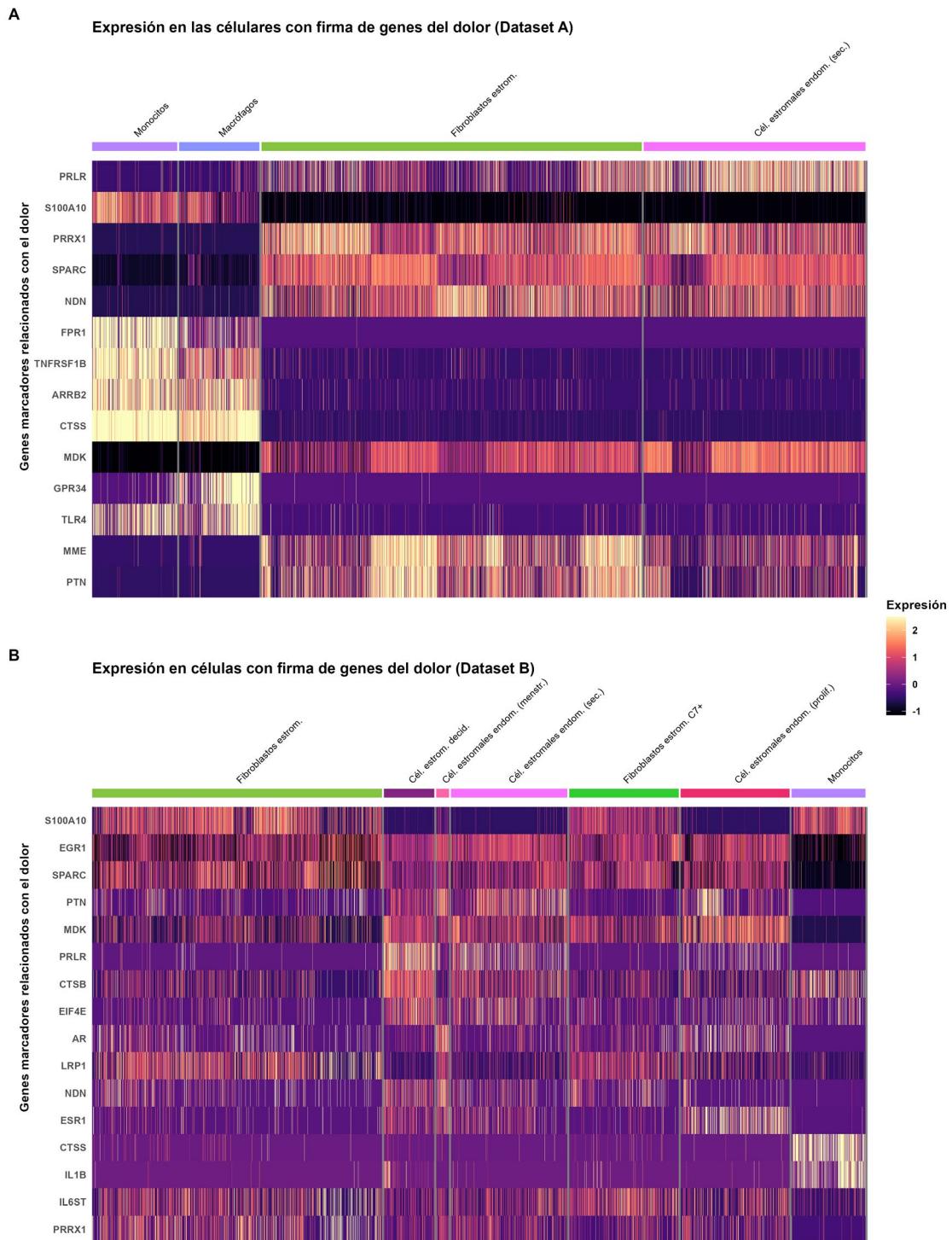
#### Figura 14. Expresión de la firma de dolor en los diferentes tipos celulares.

Representación de la expresión de la firma de genes del dolor en cada tipo celular (**A**) del conjunto de datos A y (**B**) B, para identificación de poblaciones con mayor expresión.

## 4.4. Caracterización de genes del dolor en células estromales, inmunes y fibroblastos

Se observó una mayor expresión de genes asociados al dolor en células estromales, inmunes y fibroblastos. En el caso del conjunto de datos A, se dio un mayor puntaje de la firma de dolor en células estromales endometriales en fase secretora, fibroblastos estromales, monocitos y macrófagos (Figura 14A). En el dataset B, hubo una mayor

expresión también en estas poblaciones, además de en fibroblastos estromales C7+ y células estromales endometriales en fase proliferativa y en la menstruación (**Figura 14B**).



**Figura 15.** Observación en mapas de color de la expresión de los genes diferencialmente expresados en los tipos celulares con mayor firma del dolor de ambos conjuntos de datos. Figura en alta resolución en este [enlace](#).

En cuanto a los DEG asociados al dolor, por un lado, en el *dataset A*, se determinaron 14 genes (**Figura 15A**). Comparando la expresión entre los distintos tipos celulares, los monocitos y los macrófagos tuvieron una mayor expresión de *S100A10*, *FPR1*, *TNFRSF1B*, *ARRB2*, *CTSS*, *GPR34* y *TLR4*, donde además se observó poca expresión de *S100A10* en células estromales en fase secretora y fibroblastos estromales. Estos tipos celulares, por su lado, tuvieron una mayor expresión de *PRLR*, *PRRX1*, *SPARC*, *NDN*, *MDK*, *MME* y *PTN*. Se observó menos expresión de *SPARC* y *MDK* en monocitos y fibroblastos. Por otro lado, en el *dataset B*, se determinaron 16 genes asociados al dolor en los tipos celulares con firma de dolor (**Figura 15B**). En general, se observó una expresión media-alta de ellos en todos los tipos celulares, a excepción de *EGR1*, *SPARC* y *MDK*, que tuvieron poca en monocitos. En este tipo celular, se observó además la mayor expresión de *CTSS* e *IL1B*. Se observó también una expresión alta de *S100A10* en fibroblastos estromales y monocitos.

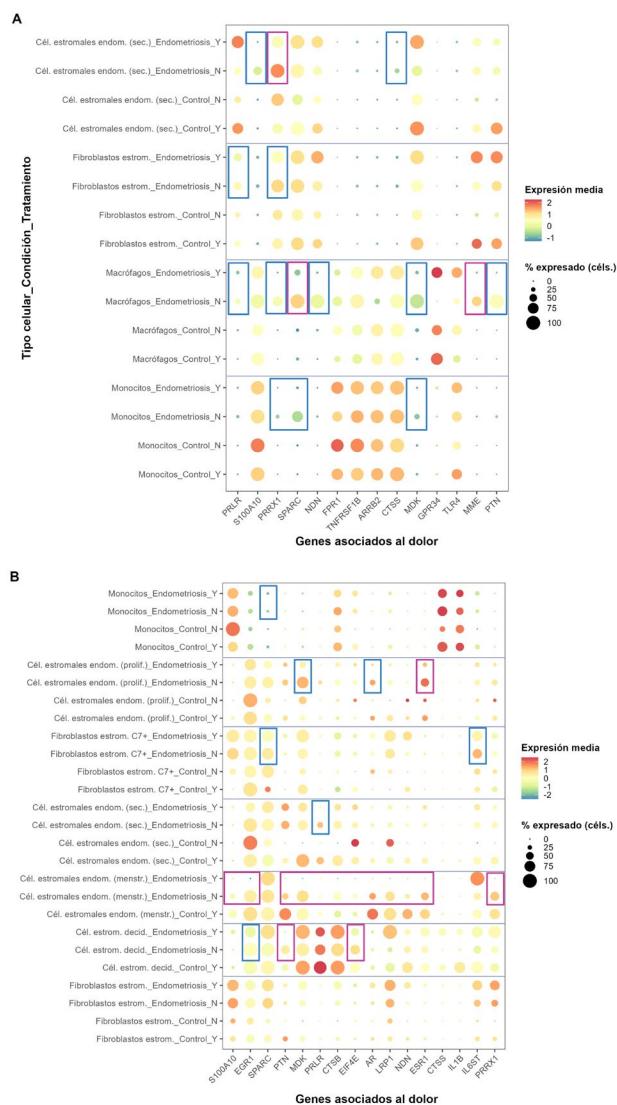
En conjunto, se determinaron 22 genes asociados al dolor, habiendo ocho en común entre ambos conjuntos de datos: *CTSS*, *MDK*, *NDN*, *PRLR*, *PRRX1*, *PTN*, *S100A10* y *SPARC*.

## 4.5. Impacto del tratamiento en los genes implicados en la vía del dolor

Se comparó la expresión de los DEG asociados al dolor entre las condiciones de tratamiento de las muestras. En el *dataset A*, se observó que *SPARC* disminuyó considerablemente su expresión con el tratamiento hormonal en macrófagos y algo menos en monocitos (**Figura 16A**). Además, cabe destacar que no se observó expresión de estos genes en los controles, solo en los casos de endometriosis. También se determinó una disminución importante de la expresión de *MME* en macrófagos y de *PRRX1* en las células estromales en fase de secretora, que también disminuye en monocitos, macrófagos y fibroblastos. Entre los genes que se expresaron algo menos con el tratamiento hormonal se hallaron *PRLR* en macrófagos y fibroblastos, *S100A10* y *CTSS* en células estromales, *NDN* y *PTN* en macrófagos, y *MDK* en las células inmunes.

En el *dataset B*, el efecto del tratamiento hormonal tuvo un mayor impacto en la expresión de los DEG en las células estromales durante la menstruación, llegándose a disminuir la expresión de la mayoría los genes (**Figura 16B**). Entre ellos se hallaron *EGR1*, *PTN* y *EIF4E*, cuyas expresiones también disminuyeron en células decidualizadas; *MDK*, *AR* y *ESR1*, que también se vieron afectadas en células estromales en fase proliferativa; y *PRLR*,

cuya reducción también se apreció en células estromales en fase secretora. Cabe mencionar que *SPARC* disminuyó también su expresión en monocitos, al igual que en dataset A, y también en fibroblastos estromales C7+. El gen *IL6ST* también se vio afectado por la terapia hormonal en esta última población. En los fibroblastos estromales no se observó la reducción en la expresión de ningún gen.



**Figura 16.** Mapas de expresión génica en diferentes tipos celulares y condiciones de endometriosis y control, con y sin tratamiento hormonal, en el (A) dataset A y en el (B) dataset B.

Los recuadros en rosa destacan genes asociados al dolor cuya expresión es considerablemente más alta en las condiciones de endometriosis, pero que disminuyen significativamente con el tratamiento hormonal. Los recuadros en azul señalan genes que muestran una disminución menos pronunciada tras el tratamiento. El tamaño de los círculos indica el porcentaje de células que expresan cada gen, y la intensidad de color refleja la expresión media. Figura en alta resolución en este [enlace](#).

En general, el análisis del impacto del tratamiento hormonal en estos genes asociados al dolor reveló una disminución de la expresión de *SPARC*, *MDK*, *PTN* y *PRRX1* en ambos conjuntos de datos en células inmunes (*SPARC*, *MDK* y *PTN*) y células estromales (*PRRX1*, *MDK* y *PTN*).

## 5. Discusión

Dada la falta de requerimiento computacional (mínimo 64GB de RAM [72]), no se pudieron preprocesar los archivos crudos con Cell Ranger, ni tampoco con otros programas que permiten obtener una matriz de conteo con datos de 10X Genomics como Alevin [73] o [STARsolo](#). Por ello, siendo posible trabajar con las matrices de conteo ya filtradas por Cell Ranger [74], estas se descargaron directamente desde GEO. La falta de procesamiento personalizado de los datos crudos implica una dependencia directa del preprocesamiento realizado por los autores de los estudios originales. Esto puede introducir variabilidad en los resultados, dado que no se puede asegurar uniformidad en los criterios aplicados para el preprocesamiento en los dos *datasets* analizados [75]. Además, dado que los archivos del *dataset* A solo fueron subidos en formato H5, se descargaron en el mismo formato los del *dataset* B para facilitar el flujo de trabajo. Para el *dataset* B, de las 49 muestras totales que mencionan haber secuenciado los investigadores [64], se descartaron es uso de 6 (4 matrices no fueron subidas y 2, no en formato H5). Por tanto, estos factores podrían afectar la confianza y la generalización de las conclusiones, ya que las posibles variaciones técnicas en el preprocesamiento de datos crudos y la pérdida de muestras podrían influir en la expresión observada de los genes asociados al dolor y en el impacto del tratamiento hormonal.

Además, después de la normalización de los datos, es recomendable proceder con la integración de los *datasets* [68], un proceso que agrupa los datos según condiciones experimentales, lotes y otras variables técnicas, con el fin de facilitar comparaciones directas y mejorar la detección de subpoblaciones celulares [76]. Sin embargo, en este trabajo, la integración de los *datasets* no se realizó debido a las limitaciones en los recursos computacionales. La capacidad de memoria RAM disponible no fue suficiente para manejar la integración a gran escala, ya que se estimó el requerimiento de al menos 128 GB de RAM

para ello. Por tanto, los análisis y conclusiones se realizaron por separado para cada conjunto de datos, lo que limitó las comparaciones entre ellos. Asimismo, debido a estas limitaciones de memoria, sólo se pudo escalar la expresión de los 2.000 genes altamente variables en el *dataset* B, aunque esto no afecta a los análisis posteriores, sólo evita que los genes altamente variables dominen [67].

En cuanto a la caracterización de los tipos celulares generales presentes en el endometrio eutópico y ectópico, estos corresponden a clases de células ya identificadas en los estudios de scRNA-seq originales [63,64] y otros [71,77]. Las células epiteliales y estromales son los dos tipos principales de células estructurales tanto en el endometrio sano como en las lesiones asociadas con la endometriosis [64]. El miometrio es la capa muscular del útero que se encuentra justo debajo del endometrio eutópico [18] (**Figura 1A**), lo que puede justificar la presencia de una población de células musculares lisas. Asimismo, en el caso de la EP, las lesiones endometriales pueden infiltrar el tejido muscular de los diferentes órganos [22] de los que se ha obtenido muestras, como el peritoneo, el recto o los ovarios. La presencia de células inmunes era esperable también dada la característica inflamatoria de la endometriosis [42,43]. La vascularización del tejido endometrial y de las lesiones ectópicas, esencial para la sostenibilidad de estas lesiones fuera del entorno uterino [24], explicarían también la presencia de células endoteliales. Asimismo, la proliferación anormal de los fibroblastos del estroma endometrial está asociada con la presencia de endometriosis [78,79], lo que justifica su alta representación en ambos conjuntos de datos.

Las diferencias entre las subpoblaciones celulares observadas entre los conjuntos de datos pueden deberse a una combinación de factores hormonales y muestrales. Por un lado, el *dataset* A incluye pacientes que recibieron tratamiento hormonal (progestágenos/estrógenos) (**Figura 11E**). La menor proporción de células estromales puede deberse a la influencia de los progestágenos, que pueden alterar la composición celular del tejido endometrial, reduciendo la proliferación de este tipo celular [1]. En cuanto a la presencia de la subpoblación de células epiteliales *SPDEF*<sup>+</sup> en tejido ectópico, al igual que en el estudio original [64], se ha determinado que la expresión de este gen se restringe a este tipo celular [80] en órganos también dependientes de hormonas, como los ovarios [81]. La ausencia de esta subpoblación en el conjunto de datos B podría sugerir una influencia del tratamiento hormonal combinado de progestágenos/estrógenos.

Por otro lado, el *dataset* B, incluye sobre todo pacientes que no recibieron tratamiento hormonal, lo que podría justificar la mayor presencia de células T [82] (**Figura 11F**). Las lesiones ectópicas en la endometriosis generan un microambiente inflamatorio [44] y, en ausencia de una terapia hormonal [82], se podría dar una respuesta inmune más intensa. Además, en este conjunto de datos, las muestras fueron recolectadas en más localizaciones anatómicas (**ANEXO II**), lo que podría haber influido en la composición celular. La presencia de la subpoblación de fibroblastos C7+, se ha caracterizado al igual que en el artículo original [63]. Se ha determinado que la expresión de C7 aumenta en los fibroblastos en condiciones de endometriosis, participando en la inflamación y en el crecimiento de lesiones [79]. La presencia de células estromales en fase de menstruación, células estromales decidualizadas y células glandulares podría estar explicada por la fase del ciclo menstrual de las pacientes durante la cual fueron extraídas las muestras. El *dataset* B contiene muestras en fases luteal/secretora y folicular (**ANEXO II**). Durante la fase luteal, la decidualización es más pronunciada y, debido al incremento de progesterona, es el periodo donde las células epiteliales glandulares secretan productos promotores de la implantación y las células estromales se diferencian a células decidualizadas [19] (**Figura 1B**).

Los tipos celulares del endometrio eutópico y ectópico determinados en la vía del dolor, las células estromales (en diferentes fases del ciclo), inmunes (monocitos y macrófagos) y fibroblastos estromales, son coherentes con la literatura sobre los mecanismos celulares implicados en la patogénesis de la endometriosis y su relación con el dolor (apartado 1.3.) [5,21,22,83]. Los fibroblastos estromales contribuyen también al microambiente inflamatorio y, por ende, a la nocicepción, secretando citoquinas proinflamatorias [84] y llevando a cabo el endurecimiento del tejido endometrial (fibrosis) [83,85,86]. En el *dataset* B, es la subpoblación de fibroblastos C7+ que tiene una mayor firma de dolor, lo que sugiere una posible especialización funcional de estas células en la mediación del dolor, aunque el gen C7 no esté asociado a este síntoma [50]. En cuanto a la expresión de genes asociados al dolor en células estromales en las distintas fases del ciclo menstrual, la acción de la progesterona en la fase secretora y el aumento de estrógenos en la fase proliferativa pueden fomentar un entorno inflamatorio mediante la liberación de citoquinas y otros mediadores inflamatorios [42,87]. Durante la menstruación, la degradación tisular y la respuesta inmune son comunes [44], conllevado a un aumento de prostaglandinas, lo que contribuye a la dismenorrea [38,42,88].

Se caracterizaron ocho genes asociados al dolor en estos tres tipos celulares de la vía del dolor en ambos conjuntos de datos, de un total de 22. En concreto, la mayor expresión de *S100A10* en fibroblastos, monocitos y macrófagos coincide con otros estudios [79,89,90]. La expresión de cuatro de estos genes se vio afectada por el tratamiento hormonal en células inmunes y estromales: *SPARC*, *MDK*, *PTN* y *PRRX1*. La disminución en su expresión podría estar vinculada con una reducción del dolor en pacientes con endometriosis. Por un lado, *SPARC* está involucrado en la síntesis de la matriz extracelular, pudiendo contribuir a la proliferación de nuevas lesiones [91–93] y fibrosis [94]. Teniendo en cuenta que estos procesos están asociados al dolor [34,42,86], el tratamiento hormonal estaría disminuyendo su progresión y, por ende, también la nocicepción. Por otro lado, *PTN* (pleiotrofina) y *MDK* (*midkine*) están involucrados en la angiogénesis y tienen un posible papel en la implantación del endometrio ectópico [95,96]. *PRRX1*, a su vez, está relacionado con la proliferación de células tumorales [97]. La disminución en la expresión de estos genes en células estromales debido al tratamiento hormonal podría estar relacionada con una menor proliferación celular y, por tanto, una menor activación inflamatoria [34,42], lo que contribuiría a una disminución del dolor percibido.

Además, en el conjunto de datos B, se observó que el tratamiento hormonal tuvo un impacto en la mayoría de los genes asociados al dolor en las células estromales durante la menstruación, que podría estar contribuyendo a un alivio en el dolor menstrual, un efecto esperado del tratamiento [88]. Concretamente, la reducción en la expresión de *ESR1* (receptor de estrógeno 1) también en células estromales, aunque de la fase proliferativa, sugiere que el tratamiento disminuye la inflamación y, por tanto, el dolor asociado, puesto que la endometriosis es una enfermedad inflamatoria dependiente de estrógenos [21,38].

En general, la reducción de la expresión genética de los genes asociados al dolor tanto en células inmunes, como estromales, en diferentes fases del ciclo menstrual, y algo menos en fibroblastos, sugiere que el tratamiento hormonal impacta en los mecanismos que regulan la percepción del dolor. Sin embargo, las limitaciones computacionales pueden haber influido en las conclusiones sobre la expresión diferencial de genes asociados al dolor. Para obtener resultados más robustos y fiables, sería necesario abordar estas limitaciones técnicas en futuros análisis. Además, se requiere la realización de más estudios para validar estos hallazgos y explorar en mayor profundidad el papel de las subpoblaciones

celulares con firma de dolor en la regulación de este síntoma en pacientes con endometriosis.

## 6. Conclusiones

- Gracias al análisis de dos conjuntos de datos de scRNA-seq (A y B) se logró identificar y caracterizar los principales tipos celulares del endometrio eutópico y ectópico implicados en la vía del dolor en pacientes con endometriosis: las células estromales, inmunes (monocitos y macrófagos) y fibroblastos estromales.
- Los resultados mostraron diferencias significativas en la expresión de varios genes asociados al dolor entre las pacientes con endometriosis que han recibido tratamiento hormonal y aquellas que no lo han recibido, en células inmunes y estromales, con menor impacto en fibroblastos estromales.
- Genes como *SPARC*, *MDK*, *PRRX1* y *PTN* mostraron una expresión reducida en células inmunes y estromales en pacientes bajo tratamiento hormonal en ambos conjuntos de datos analizados, lo que sugiere que esta terapia tiene un impacto en los mecanismos responsables de la percepción del dolor, al reducir la proliferación celular.
- La disminución de *ESR1* (receptor de estrógenos) en células estromales en el conjunto de datos B refuerza la relación entre la reducción del dolor y la modulación hormonal, lo cual es coherente con el carácter inflamatorio y dependiente de estrógenos de la endometriosis.
- La mayor parte de los genes estudiados en el conjunto de datos B mostró una reducción en su expresión en células estromales durante la fase menstrual del ciclo, lo que respalda el efecto positivo del tratamiento hormonal en la disminución del dolor menstrual (dismenorrea).

## 7. Código y recursos en GitHub

Los *scripts* utilizados y todo este trabajo se hallan subidos a [GitHub](#).

## 8. Referencias bibliográficas

1. Saunders PTK, Horne AW. Endometriosis: Etiology, pathobiology, and therapeutic prospects. *Cell.* 2021 May;184(11):2807–24.
2. Koga K. Immunology of Endometriosis: Pathogenesis and Management. Kuga K, editor. Vol. 2. Elsevier; 2022. 1–327 p.
3. Mori T, Kitawaki J. Role of estrogen and estrogen-related factors in endometriosis. In: Kuga K, editor. *Immunology of Endometriosis: Pathogenesis and Management.* Academic Press; 2022. p. 105–19.
4. Pašalić E, Tambuwala MM, Hromić-Jahjefendić A. Endometriosis: Classification, pathophysiology, and treatment options. *Pathol Res Pract.* 2023 Nov;251:154847.
5. Bulun SE, Yilmaz BD, Sison C, Miyazaki K, Bernardi L, Liu S, et al. Endometriosis. *Endocr Rev.* 2019 May;40(4):1048.
6. Piccolo CL, Cea L, Sbarra M, De Nicola AM, De Cicco Nardone C, Faiella E, et al. Magnetic Resonance Roadmap in Detecting and Staging Endometriosis: Usual and Unusual Localizations. *Applied Sciences.* 2023 Sep;13(18):10509.
7. Rahman S, Youssef Y, Maroun G, Inaty D, Kheil MH, Moawad G. Eyes, Menstruation and Endometriosis. *Facts Views Vis Obgyn.* 2023 Jun;15(2):107.
8. Elbasueny B, Geerts M, Yang EC, Allaire C, Yong PJ, Bedaiwy MA. Medical treatments of endometriosis: A review. *Reproductive and Developmental Medicine.* 2023 Sep;7(3):166–79.
9. Ávalos Marfil A, Barranco Castillo E, Martos García R, Ladrón de Guevara NM, Mazheika M. Epidemiology of Endometriosis in Spain and Its Autonomous Communities: A Large, Nationwide Study. *Int J Environ Res Public Health.* 2021 Aug;18(15):7861.
10. Bougie O, Yap MI, Sikora L, Flaxman T, Singh S. Influence of race/ethnicity on prevalence and presentation of endometriosis: a systematic review and meta-analysis. *BJOG.* 2019 Aug;126(9):1104–15.
11. Agarwal SK, Chapron C, Giudice LC, Laufer MR, Leyland N, Missmer SA, et al. Clinical diagnosis of endometriosis: a call to action. *Am J Obstet Gynecol.* 2019 Apr;220(4):354.e1-354.e12.

12. Zhang L, Losin EAR, Ashar YK, Koban L, Wager TD. Gender Biases in Estimation of Others' Pain. *J Pain*. 2021 Sep;22(9):1048–59.
13. Margueritte F, Fritel X, Zins M, Goldberg M, Panjo H, Fauconnier A, et al. The underestimated prevalence of neglected chronic pelvic pain in women, a nationwide cross-sectional study in France. *J Clin Med*. 2021 Jun;10(11):10.
14. Wiggleton-Little J. "Just" a painful period: A philosophical perspective review of the dismissal of menstrual pain. *Women's Health*. 2024 Jan;20(17455057241255646).
15. Ramos-Pichardo JD, Ortega-Galán ÁM, Iglesias-López MT, Abreu-Sánchez A, Fernández-Martínez E. Why Do Some Spanish Nursing Students with Menstrual Pain Fail to Consult Healthcare Professionals? *Int J Environ Res Public Health*. 2020 Nov;17(21):1–14.
16. Abreu-Sánchez A, Parra-Fernández ML, Onieva-Zafra MD, Fernández-Martínez E. Perception of Menstrual Normality and Abnormality in Spanish Female Nursing Students. *Int J Environ Res Public Health*. 2020 Sep;17(17):1–12.
17. Signorile PG, Baldi A, Viceconte R, Ronchi A, Montella M. Pathogenesis of Endometriosis: Focus on Adenogenesis-related Factors. *In Vivo (Brooklyn)*. 2023 Sep;37(5):1922–30.
18. Pawlina W, Ross M. Sistema genital femenino. In: Ross: Histología Texto y Atlas. 8th ed. Wolters Kluwer; 2020. p. 897–863.
19. Zhu H, Hou CC, Luo LF, Hu YJ, Yang WX. Endometrial stromal cells and decidualized stromal cells: Origins, transformation and functions. *Gene*. 2014 Nov;551(1):1–14.
20. Sampson JA. Metastatic or Embolic Endometriosis, due to the Menstrual Dissemination of Endometrial Tissue into the Venous Circulation. *Am J Pathol*. 1927 Mar;3(2):93–110.
21. Ochoa Bernal MA, Fazleabas AT. The Known, the Unknown and the Future of the Pathophysiology of Endometriosis. *Int J Mol Sci*. 2024 Jun;25(11):5815.
22. Wang Y, Nicholes K, Shih IM. The Origin and Pathogenesis of Endometriosis. *Annu Rev Pathol*. 2020 Jan;15:71–95.
23. Tomassetti C, Johnson NP, Petrozza J, Abrao MS, Einarsson JI, Horne AW, et al. An international terminology for endometriosis, 2021. *Hum Reprod Open*. 2021;13(4):295–304.

24. García-Gómez E, Vázquez-Martínez ER, Reyes-Mayoral C, Cruz-Orozco OP, Camacho-Arroyo I, Cerbón M. Regulation of Inflammation Pathways and Inflammasome by Sex Steroid Hormones in Endometriosis. *Front Endocrinol (Lausanne)*. 2020 Jan;10:935.
25. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Viganamp P. Endometriosis. *Nat Rev Dis Primers*. 2018;4(1):9.
26. Vermeulen N, Abrao MS, Einarsson JI, Horne AW, Johnson NP, Lee TTM, et al. Endometriosis classification, staging and reporting systems: a review on the road to a universally accepted endometriosis classification. *Hum Reprod Open*. 2021;2021(4):1–28.
27. Canis M, Donnez JG, Guzick DS, Halme JK, Rock JA, Schenken RS, et al. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertil Steril*. 1997;67(5):817–21.
28. Andres MP, Borrelli GM, Abrão MS. Endometriosis classification according to pain symptoms: can the ASRM classification be improved? *Best Pract Res Clin Obstet Gynaecol*. 2018 Aug;51:111–8.
29. Servicio Andaluz de Salud. Guía de atención a mujeres con endometriosis en el Sistema Sanitario Público de Andalucía. 2017.
30. Lebovic DI, Falcone T. Guía de atención a las mujeres con endometriosis en el Sistema Nacional de Salud (SNS). INFORMES, ESTUDIOS E INVESTIGACIÓN. Madrid (España); 2013.
31. Asociación de Afectadas de Endometriosis Crónica (ADAEC) [Internet]. [cited 2024 Aug 6]. Available from: <https://adaec.es/>
32. Missmer SA, Tu FF, Agarwal SK, Chapron C, Soliman AM, Chiuve S, et al. Impact of Endometriosis on Life-Course Potential: A Narrative Review. *Int J Gen Med*. 2021;14:9–25.
33. Mechsnner S. Endometriosis, an Ongoing Pain—Step-by-Step Treatment. *J Clin Med*. 2022 Jan;11(2):467.
34. Maddern J, Grundy L, Castro J, Brierley SM. Pain in Endometriosis. *Front Cell Neurosci*. 2020 Oct;14:590823.

35. Koller D, Pathak GA, Wendt FR, Tylee DS, Levey DF, Overstreet C, et al. Epidemiologic and Genetic Associations of Endometriosis With Depression, Anxiety, and Eating Disorders. *JAMA Netw Open*. 2023 Jan;6(1):e2251214–e2251214.
36. Missmer SA, Tu F, Soliman AM, Chiuve S, Cross S, Eichner S, et al. Original research: Impact of endometriosis on women's life decisions and goal attainment: a cross-sectional survey of members of an online patient community. *BMJ Open*. 2022 Apr;12(4):52765.
37. As-Sanie S, Shafrir AL, Halvorson L, Chawla R, Hughes R, Merz M. The Burden of Pelvic Pain Associated with Endometriosis among Women in Selected European Countries and the United States: A Restricted Systematic Review. *J Minim Invasive Gynecol*. 2024 May;31(8):653–66.
38. Song SY, Jung YW, Shin WK, Park M, Lee GW, Jeong S, et al. Endometriosis-Related Chronic Pelvic Pain. *Biomedicines* 2023, Vol 11, Page 2868. 2023 Oct;11(10):2868.
39. Salmeri N, Li Piani L, Cavoretto PI, Somigliana E, Viganò P, Candiani M. Endometriosis increases the risk of gestational diabetes: a meta-analysis stratified by mode of conception, disease localization and severity. *Sci Rep*. 2023 Dec;13(1):8099.
40. Okoli U, Charoenngam N, Ponvilawan B, Jaroenlapnopparat A, Mettler S, Obiejesie O. Endometriosis and Risk of Cardiovascular Disease: Systematic Review and Meta-Analysis. *J Womens Health (Larchmt)*. 2023 Dec;32(12):1328–39.
41. Kvaskoff M, Mahamat-Sale Y, Farland L V., Shigesi N, Terry KL, Harris HR, et al. Endometriosis and cancer: a systematic review and meta-analysis. *Hum Reprod Update*. 2021 Mar;27(2):393–420.
42. Barra F, Evangelisti G, Ferrero S, Scala C. Pain. In: *Immunology of Endometriosis: Pathogenesis and Management*. Academic Press; 2022. p. 155–68.
43. Carlyle D, Khader T, Lam D, Vadivelu N, Shiwlchan D, Yonghee C. Endometriosis Pain Management: a Review. *Curr Pain Headache Rep*. 2020 Sep;24(9):1–9.
44. Vallvé-Juanico J, Houshdaran S, Giudice LC. The endometrial immune environment of women with endometriosis. *Hum Reprod Update*. 2019 Sep;25(5):565–92.
45. Chantalat E, Valera MC, Vaysse C, Noirrit E, Rusidze M, Weyl A, et al. Estrogen Receptors and Endometriosis. *International Journal of Molecular Sciences* 2020, Vol 21, Page 2815. 2020 Apr;21(8):2815.

46. Laux-Biehlmann A, D'hooghe T, Zollner TM. Menstruation pulls the trigger for inflammation and pain in endometriosis. *Trends Pharmacol Sci.* 2015 May;36(5):270–6.
47. Capezzuoli T, Rossi M, La Torre F, Vannuccini S, Petraglia F. Hormonal drugs for the treatment of endometriosis. *Curr Opin Pharmacol.* 2022 Dec;67:102311.
48. Vannuccini S, Clemenza S, Rossi M, Petraglia F. Hormonal treatments for endometriosis: The endocrine background. *Rev Endocr Metab Disord.* 2022 Jun;23(3):333.
49. Nabholz JM, Mouret-Reynier MA, Durando X, Van Praagh I, Al-Sukhun S, Ferriere JP, et al. Comparative review of anastrozole, letrozole and exemestane in the management of early breast cancer. *Expert Opin Pharmacother.* 2009 Jun;10(9):1435–47.
50. Wistrom E, Chase R, Smith PR, Campbell ZT. A compendium of validated pain genes. *Wires Mechanisms of Disease.* 2022 Nov;14(6):e1570.
51. Crick F. Central Dogma of Molecular Biology. *Nature* 1970 227:5258. 1970;227(5258):561–3.
52. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009 Jan;10(1):57.
53. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol.* 2017 May;13(5):e1005457.
54. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 2008 Sep;18(9):1509.
55. Schena M, Shalon D, Davis RW, Brown PO. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science (1979).* 1995 Oct;270(5235):467–70.
56. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet.* 2019 Jul;20(11):631–56.
57. Wang S, Sun ST, Zhang XY, Ding HR, Yuan Y, He JJ, et al. The Evolution of Single-Cell RNA Sequencing Technology and Application: Progress and Perspectives. *Int J Mol Sci.* 2023 Feb;24(3):2943.

58. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med.* 2022 Mar;12(3):e694.
59. Slovin S, Carissimo A, Panariello F, Grimaldi A, Bouché V, Gambardella G, et al. Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. In: *Methods in Molecular Biology RNA Bioinformatics*. 2nd ed. Humana, New York, NY; 2021. p. 343–65.
60. Ke M, Elshenawy B, Sheldon H, Arora A, Buffa FM. Single cell RNA-sequencing: A powerful yet still challenging technology to study cellular heterogeneity. *BioEssays.* 2022 Nov;44(11):2200084.
61. Andrews TS, Kiselev VY, McCarthy D, Hemberg M. Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nature Protocols* 2020 16:1. 2020 Dec;16(1):1–9.
62. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol.* 2019 Jun;15(6):8746.
63. Tan Y, Flynn WF, Sivajothi S, Luo D, Bozal SB, Davé M, et al. Single cell analysis of endometriosis reveals a coordinated transcriptional program driving immunotolerance and angiogenesis across eutopic and ectopic tissues. *Nat Cell Biol.* 2022 Aug;24(8):1306.
64. Fonseca MAS, Haro M, Wright KN, Lin X, Abbasi F, Sun J, et al. Single-cell transcriptomic analysis of endometriosis. *Nat Genet.* 2023 Feb;55(2):255.
65. Rahmioglu N, Mortlock S, Ghiasi M, Møller PL, Stefansdottir L, Galarneau G, et al. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nat Genet.* 2023 Mar;55(3):423–36.
66. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* 2023 42:2. 2023 May;42(2):293–304.
67. Satija R, Satija Lab and Collaborators. Tools for Single Cell Genomics: Seurat [Internet]. 2024 [cited 2024 Aug 28]. Available from: <https://satijalab.org/seurat/>
68. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol.* 2019 Jun;15(6):8746.

69. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* 2019 Apr;8(4):329-337.e4.
70. AlJanahi AA, Danielsen M, Dunbar CE. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. *Mol Ther Methods Clin Dev.* 2018 Sep;10:189–96.
71. Marečková M, Garcia-Alonso L, Mouillet M, Lorenzi V, Petryszak R, Sancho-Serra C, et al. An integrated single-cell reference atlas of the human endometrium. *Nature Genetics* 2024. 2024 Aug;1–13.
72. Genomics 10x. System Requirements - Official 10x Genomics Support [Internet]. 2024 [cited 2024 Aug 26]. Available from: <https://www.10xgenomics.com/support/software/cell-ranger/downloads/cr-system-requirements>
73. Srivastava A, Malik L, Smith T, Sudbery I, Patro R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol.* 2019 Mar;20(1).
74. Amezquita [aut] R, Lun [aut, cre] A, Hicks [aut] S, Gottardo [aut] R. Advanced Single-Cell Analysis with Bioconductor. 2022 [cited 2024 Sep 15]. Chapter 7: Droplet processing. Available from: <https://bioconductor.org/books/3.15/OSCA.advanced/droplet-processing.html>
75. Official 10x Genomics Support. 10X Genomics Pipelines. 2020 [cited 2024 Sep 15]. Single-Library Analysis with Cell Ranger. Available from: <https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/3.1/using/count>
76. Harvard Chan Bioinformatics Core (HBC). Single-cell RNA-seq data analysis workshop [Internet]. 2024 [cited 2024 Sep 4]. Available from: [https://github.com/hbctraining/scRNA-seq\\_online](https://github.com/hbctraining/scRNA-seq_online)
77. Song Y, Wang L, Zhang Y. Identification of central genes for endometriosis through integration of single-cell RNA sequencing and bulk RNA sequencing analysis. *Medicine*. 2023 Dec;102(50):E36707.
78. McKinnon BD, Lukowski SW, Mortlock S, Crawford J, Atluri S, Subramaniam S, et al. Altered differentiation of endometrial mesenchymal stromal fibroblasts is associated with endometriosis susceptibility. *Communications Biology* 2022 5:1. 2022 Jun;5(1):1–14.

79. Ma J, Zhang L, Zhan H, Mo Y, Ren Z, Shao A, et al. Single-cell transcriptomic analysis of endometriosis provides insights into fibroblast fates and immune cell heterogeneity. *Cell Biosci.* 2021 Dec;11(1):125.
80. Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. *Gene.* 2003 Jan;303(1–2):11–34.
81. Ye T, Li J, Feng J, Guo J, Wan X, Xie D, et al. The subtype-specific molecular function of SPDEF in breast cancer and insights into prognostic significance. *J Cell Mol Med.* 2021 Aug;25(15):7307.
82. Hall OJ, Klein SL. Progesterone-based compounds affect immune responses and susceptibility to infections at diverse mucosal sites. *Mucosal Immunology* 2017 10:5. 2017 Apr;10(5):1097–107.
83. Shinotsuka N, Denk F. Fibroblasts: the neglected cell type in peripheral sensitisation and chronic pain? A review based on a systematic search of the literature. *BMJ Open Science.* 2022 Jan;6(1):100235.
84. Landén NX, Li D, Stähle M. Transition from inflammation to proliferation: a critical step during wound healing. *Cell Mol Life Sci.* 2016 Oct;73(20):3861.
85. Zhu S, Wang A, Xu W, Hu L, Sun J, Wang X. The heterogeneity of fibrosis and angiogenesis in endometriosis revealed by single-cell RNA-sequencing. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease.* 2023 Feb;1869(2):166602.
86. Vissers G, Giacomozzi M, Verdurmen W, Peek R, Nap A. The role of fibrosis in endometriosis: a systematic review. *Hum Reprod Update.* 2024 Jul;
87. Vallvé-Juanico J, Giudice LC. Immune phenotypes and mediators affecting endometrial function in women with endometriosis. In: Koga K, editor. *Immunology of Endometriosis: Pathogenesis and Management.* Academic Press; 2022. p. 169–91.
88. Itani R, Soubra L, Karout S, Rahme D, Karout L, Khojah HMJ. Primary Dysmenorrhea: Pathophysiology, Diagnosis, and Treatment Updates. *Korean J Fam Med.* 2022;43(2):101.
89. Neuditschko B, Leibetseder M, Brunmair J, Hagn G, Skos L, Gerner MC, et al. Epithelial cell line derived from endometriotic lesion mimics macrophage nervous mechanism of pain generation on proteome and metabolome levels. *Biomolecules.* 2021 Aug;11(8):1230.

90. Okura GC, Bharadwaj AG, Waisman DM. Recent Advances in Molecular and Cellular Functions of S100A10. *Biomolecules* 2023, Vol 13, Page 1450 [Internet]. 2023 Sep [cited 2024 Sep 15];13(10):1450. Available from: <https://www.mdpi.com/2218-273X/13/10/1450/htm>
91. Meola J, Rosa e Silva JC, Dentillo DB, da Silva WA, Veiga-Castelli LC, de Souza Bernardes LA, et al. Differentially expressed genes in eutopic and ectopic endometrium of women with endometriosis. *Fertil Steril*. 2010 Apr;93(6):1750–73.
92. Fung JN, Mortlock S, Girling JE, Holdsworth-Carson SJ, Teh WT, Zhu Z, et al. Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. *Scientific Reports* 2018 8:1. 2018 Jul;8(1):1–19.
93. Bradshaw AD. Diverse Biological Functions of the SPARC Family of Proteins. *Int J Biochem Cell Biol*. 2012 Mar;44(3):480.
94. Trombetta-eSilva J, Bradshaw AD. The Function of SPARC as a Mediator of Fibrosis. *Open Rheumatol J*. 2012 Jun;6(1):146–55.
95. Chung HW, Wen Y, Choi EA, Hao-Li, Moon HS, Yu HK, et al. Pleiotrophin (PTN) and midkine (MK) mRNA expression in eutopic and ectopic endometrium in advanced stage endometriosis. *Mol Hum Reprod* [Internet]. 2002 Apr [cited 2024 Sep 15];8(4):350–5. Available from: <https://dx.doi.org/10.1093/molehr/8.4.350>
96. Hirota Y, Osuga Y, Koga K, Yoshino O, Hirata T, Harada M, et al. Possible implication of midkine in the development of endometriosis. *Human Reproduction*. 2005 Apr;20(4):1084–9.
97. Meng Z, Chen Y, Wu W, Yan B, Zhang L, Chen H, et al. PRRX1 Is a Novel Prognostic Biomarker and Facilitates Tumor Progression Through Epithelial–Mesenchymal Transition in Uveal Melanoma. *Front Immunol*. 2022 Feb;13:754645.

# 9. Anexos

**ANEXO I. Metadata correspondiente al dataset A** (muestras de pacientes con endometriosis con tratamiento hormonal). Abreviaturas: Y, con tratamiento hormonal; N, sin tratamiento hormonal.

id	oldID	Dataset	condition	age	ethnicity	menstrualcycle	sample_location	tissue	RASMR_stage	TreatmentState	OralContraceptiveTreatment	OtherDrugTreatment	Assay_Type	AvgSpotLen	Beta	SRA_Study	BioProject	BioSample	Batch
A.E01	C01	A	Control	36	asian	menstrual	Eubiotic	Endometrium	4	Y	nonethindroneE2/120	levonorgestrel	RNA-Seq	88	627012528	SRR1545235	PRJNA74463	SAM2069420	134491515
A.E02	C02	A	Control	22	white	proliferativa	Eubiotic	Endometrium	NA	Y	nonethindroneE2/120	levonorg	RNA-Seq	88	62451747318	SRR1545235	PRJNA74463	SAM2069421	1279562008
A.C03	C03	A	Control	42	white	proliferativa	Eubiotic	Endometrium	NA	N	NA	ibuprofen	RNA-Seq	88	53116929777	SRR1545235	PRJNA74463	SAM2069422	1240670028
A.E01	E01	A	Endometriosis	40	white	menstrual	Eubiotic	Endometrium	3	N	NA	irragulide, lisinopril, januria	RNA-Seq	88	72079819539	SRR1545235	PRJNA74463	SAM2069425	15431302393
A.E02	E02	A	Endometriosis	40	hispanic	proliferativa	Eubiotic	Endometrium	2	Y	nonethindroneE2/120	NA	RNA-Seq	88	72079819540	SRR1545235	PRJNA74463	SAM2069426	15431302394
A.E03	E03	A	Endometriosis	33	hispanic	proliferativa	Eubiotic	Endometrium	4	Y	nonethindroneE2/121	dostine, clonazepam, lamictal, serquel,sumatriptan	RNA-Seq	88	75886627874	SRR1545235	PRJNA74463	SAM2069425	14814502211
A.E04	E04	A	Endometriosis	27	white	NA	Eubiotic	Endometrium	4	Y	nonethindroneE2/122	mirena, lexapro, trazadone,lorazepam	RNA-Seq	88	72362691953	SRR1545235	PRJNA74463	SAM2069428	12460876535
A.E05	E05	A	Endometriosis	33	white	proliferativa	Eubiotic	Endometrium	4	Y	nonethindroneE2/123	NA	RNA-Seq	88	62951250000	SRR1545235	PRJNA74463	SAM2069429	148091407
A.E06	E06	A	Endometriosis	30	white	NA	Eubiotic	Peritoneum	4	Y	nonethindroneE2/124	NA	RNA-Seq	88	63049881891	SRR1545235	PRJNA74463	SAM2069411	12987221567
A.E06.1	E06	A	Endometriosis	30	white	NA	Eubiotic	Endometrium	4	Y	nonethindroneE2/125	NA	RNA-Seq	88	33798523326	SRR1545235	PRJNA74463	SAM2069429	7459634367
A.E07	E07	A	Endometriosis	34	white	NA	Ectopic Ovary	Ovary	4	Y	nonethindroneE2/126	NA	RNA-Seq	88	52027953886	SRR1545235	PRJNA74463	SAM2069413	9965377508
A.E07.1	E07	A	Endometriosis	34	white	NA	Ectopic	Peritoneum	4	Y	nonethindroneE2/127	NA	RNA-Seq	88	63620279509	SRR1545235	PRJNA74463	SAM2069420	117422000
A.E07.3	E07	A	Endometriosis	34	white	NA	Eubiotic	Endometrium	4	Y	nonethindroneE2/128	NA	RNA-Seq	88	59115177094	SRR1545235	PRJNA74463	SAM2069415	12466857905
A.E08	E08	A	Endometriosis	40	white	NA	Eubiotic	Endometrium	4	Y	nonethindroneE2/129	NA	RNA-Seq	88	65410737919	SRR1545235	PRJNA74463	SAM2069416	11510110459
A.E09.1	E09	A	Endometriosis	45	hispanic	secretora	Ectopic Ovary	Ovary	4	Y	nonethindroneE2/130	NA	RNA-Seq	88	65032000000	SRR1545235	PRJNA74463	SAM2069421	620000000
A.E09.2	E09	A	Endometriosis	45	hispanic	proliferativa	Eubiotic	Endometrium	4	Y	nonethindroneE2/131	NA	RNA-Seq	88	72173409975	SRR1545235	PRJNA74463	SAM2069421	14074814039
A.E10	E11	A	Endometriosis	35	hispanic	secretora	Ectopic Ovary	Ovary	4	Y	nonethindrone	NA	RNA-Seq	89	59561974306	SRR1545235	PRJNA74463	SAM2069423	13631472318
id	Consent	Dataset	Condition	DataType	DataProvider	Experiment	Instrument	Library	LibType	LibraryLayout	LibraryLocation	LibrarySource	LibrarySystem	Organism	Platform	PlatformRun	PlatformSeries	Sample	AssayType
A.C01	GEO	run29,TenX	ncbi_gs3	ncbi_gs3	ncbi_gs3	SRX1545228	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T08:15:00Z	GSM6102532	GRCH38	
A.C02	GEO	public	run29,TenX	ncbi_gs3	ncbi_gs3	SRX1545227	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T12:21:00Z	GSM6102530	GRCH38	
A.C03	GEO	public	run29,TenX	ncbi_gs3	ncbi_gs3	SRX1545234	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T10:10:00Z	GSM6102534	GRCH38	
A.E01	GEO	public	run29,TenX	ncbi_gs3	ncbi_gs3	SRX1545231	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T10:00:00Z	GSM6102531	GRCH38	
A.E02	GEO	public	TenX,run,zq	ncbi_gs3	ncbi_gs3	SRX1545234	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T10:59:00Z	GSM6102540	GRCH38	
A.E03	GEO	public	TenX,run,zq,era	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545253	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T05:29:00Z	GSM6102543	GRCH38	
A.E04	GEO	public	run29,TenX	ncbi_gs3	gs_ncbi_gs3	SRX1545254	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T05:29:00Z	GSM6102544	GRCH38	
A.E05	GEO	public	run29,TenX	ncbi_gs3	gs_ncbi_gs3	SRX1545250	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T06:32:00Z	GSM6102549	GRCH38	
A.E06.1	GEO	public	run29,TenX	ncbi_gs3	gs_ncbi_gs3	SRX1545260	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T18:28:00Z	GSM6102562	GCA 000001405_29	
A.E06.2	GEO	public	run29,TenX	ncbi_gs3	ncbi_gs3	SRX1545261	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T07:14:00Z	GSM6102561	GRCH38	
A.E07.1	GEO	public	run29,TenX	ncbi_gs3	ncbi_gs3	SRX1545262	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T07:14:00Z	GSM6102562	GRCH38	
A.E07.2	GEO	public	TenX,run,zq	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545263	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T07:57:00Z	GSM6102563	GRCH38	
A.E07.3	GEO	public	run29,TenX	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545264	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T07:22:00Z	GSM6102564	GRCH38	
A.E08	GEO	public	run29,TenX	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545265	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T07:59:00Z	GSM6102565	GRCH38	
A.E09.1	GEO	public	run29,TenX	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545266	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T09:50:00Z	GSM6102566	GRCH38	
A.E09.2	GEO	public	run29,TenX	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545267	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-05T01:19:00Z	GSM6102560	GCA 000001405_29	
A.E10	GEO	public	run29,TenX	gs_ncbi_gs3	gs_ncbi_gs3	SRX1545240	Illumina NovaSeq 6000	Gene Expression	PAIRED	cDNA	TRANSCRIPTOMIC SINGLE CELL	scRNA-seq	Homo sapiens	ILLUMINA	2022-05-06T00:00:00Z	2022-05-04T07:34:00Z	GSM6102562	GRCH38	

**ANEXO II. Metadata correspondiente al dataset B** (muestras de pacientes con endometriosis sin tratamiento hormonal). Abreviaturas: Y, con tratamiento hormonal; N, sin tratamiento hormonal.

<b>id</b>	<b>old.ID</b>	<b>dataset</b>	<b>condition</b>	<b>age</b>	<b>ethnicity</b>	<b>menstrualCycle</b>	<b>sampleLocation</b>	<b>tissue</b>	<b>rASMR_stage</b>	<b>TreatmentState</b>	<b>OralContraceptiveTreatment</b>	<b>OtherDrugTreatment</b>	<b>Assay.Type</b>	<b>AvgSpotLen</b>	<b>SRA.Study</b>	<b>Bases</b>	
B.C01	6	B	Control	58	White	(post-menopausia)	Endometrium	Eutopic Endometrium	NA	Y	Climara 0.05 mg/24hr transdermal patch & Prometrium 100mg capsule nightly	NA	RNA-Seq	310	SRP396798	1,09078E+11	
B.C02	12	B	Control	48	White	NA	Endometrium	Eutopic Endometrium	NA	Y	Aygestin (norethindrone acetate), taken before surgery to help control heavy bleeding	NA	RNA-Seq	308	SRP396798	78484758968	
B.C03	15	B	Control	52	White	luteal	Right ovary	Ovary	NA	N	NA	NA	RNA-Seq	320	SRP396798	50109387200	
B.E01	1	B	Endometriosis	40	Asian	luteal	Left ovary	Endometrioma	4	N	NA	NA	RNA-Seq	310	SRP396798	22710844590	
B.E02	4	B	Endometriosis	32	White	luteal	Anterior cul-de-sac	Endometriosis	1	N	NA	NA	RNA-Seq	187	SRP396798	9350000000	
B.E03	5	B	Endometriosis	32	White	luteal	Posterior cul-de-sac	Endometriosis	1	N	NA	NA	RNA-Seq	187	SRP396798	9350000000	
B.E04.1	7	B	Endometriosis	24	White	NA	Right and left pelvic side wall†	Endometriosis	4	Y	Nuvaring - etonogestrel-ethynodiol 0.12-0.015mg/ 24hr	NA	RNA-Seq	308	SRP396798	20471244024	
B.E04.2	7	B	Endometriosis	24	White	NA	Anterior cul-de-sac	Endometriosis	4	Y	Nuvaring - etonogestrel-ethynodiol 0.12-0.015mg/ 24hr	NA	RNA-Seq	308	SRP396798	35758694664	
B.E04.3	7	B	Endometriosis	24	White	NA	Right ovary	Endometrioma	4	Y	Nuvaring - etonogestrel-ethynodiol 0.12-0.015mg/ 24hr	NA	RNA-Seq	308	SRP396798	1.22665E+11	
B.E04.4	7	B	Endometriosis	24	White	NA	Right and left uterosacral ligament	Endometriosis	4	Y	Nuvaring - etonogestrel-ethynodiol 0.12-0.015mg/ 24hr	NA	RNA-Seq	308	SRP396798	1.53234E+11	
B.E05.1	8	B	Endometriosis	41	White	follicular	Anterior cul-de-sac and right pelvic side wall†	No endometriosis detected	4	N	NA	NA	RNA-Seq	308	SRP396798	98049494344	
B.E05.2	8	B	Endometriosis	41	White	follicular	Left ovary	Endometrioma	4	N	NA	NA	RNA-Seq	308	SRP396798	64395112936	
B.E05.3	8	B	Endometriosis	41	White	follicular	Left uterosacral ligament	Endometriosis	4	N	NA	NA	RNA-Seq	308	SRP396798	54089019292	
B.E05.4	8	B	Endometriosis	41	White	follicular	Rectal serosa	No endometriosis detected	4	N	NA	NA	RNA-Seq	308	SRP396798	45767480828	
B.E06.1	9	B	Endometriosis	31	White	follicular	Right and left pelvic side wall	Endometriosis	3	N	NA	NA	RNA-Seq	308	SRP396798	45444680582	
B.E06.2	9	B	Endometriosis	31	White	follicular	Left ovarian fossa	Endometriosis	3	N	NA	NA	RNA-Seq	308	SRP396798	46440959595	
B.E06.3	9	B	Endometriosis	31	White	follicular	Anterior abdominal wall and cul-de-sac†	Endometriosis	3	N	NA	NA	RNA-Seq	308	SRP396798	53802990164	
B.E06.4	9	B	Endometriosis	31	White	follicular	Right ovary	Endometrioma	3	N	NA	NA	RNA-Seq	308	SRP396798	51708044696	
B.E06.5	9	B	Endometriosis	31	White	follicular	Right parametrum	Endometriosis	3	N	NA	NA	RNA-Seq	308	SRP396798	52884533240	
B.E06.6	9	B	Endometriosis	31	White	follicular	Perirectal fat	Endometriosis	3	N	NA	NA	RNA-Seq	308	SRP396798	48019508152	
B.E07.1	10	B	Endometriosis	40	White	follicular	Endometrium	Eutopic Endometrium	4	N	NA	NA	RNA-Seq	308	SRP396798	20445354468	
B.E07.2	10	B	Endometriosis	40	White	follicular	Left uterosacral ligament	No endometriosis detected	4	N	NA	NA	RNA-Seq	308	SRP396798	23681835100	
B.E07.3	10	B	Endometriosis	40	White	follicular	Right uterosacral ligament	Endometriosis	4	N	NA	NA	RNA-Seq	308	SRP396798	27163837008	
B.E07.4	10	B	Endometriosis	40	White	follicular	Bladder endometriosis	No endometriosis detected	4	N	NA	NA	RNA-Seq	308	SRP396798	50908051656	
B.E08.1	11	B	Endometriosis	42	White	luteal	Left pelvic side wall	Endometriosis	2	N	NA	NA	RNA-Seq	308	SRP396798	21594758416	
B.E08.2	11	B	Endometriosis	42	White	luteal	Left ovarian cyst	Endometrioma	2	N	NA	NA	RNA-Seq	308	SRP396798	1.07191E+11	
B.E08.3	11	B	Endometriosis	42	White	luteal	Endometrium	Eutopic Endometrium	2	N	NA	NA	RNA-Seq	308	SRP396798	1.25138E+11	
B.E09	13	B	Endometriosis	38	White	NA	Endometrium	Eutopic Endometrium	NA	N	NA	NA	Patient on the Wiley protocol, using estrogen and testosterone creams and progesterdon on days 14-28	RNA-Seq	308	SRP396798	80442098044
B.E10.1	14	B	Endometriosis	28	White	luteal	Right ovary	Endometriosis	2	N	NA	NA	RNA-Seq	320	SRP396798	58076140480	
B.E10.2	14	B	Endometriosis	28	White	luteal	Left ovary	Ovary	2	N	NA	NA	RNA-Seq	320	SRP396798	36810443200	
B.E11.1	16	B	Endometriosis	21	White	follicular	Left ovarian fossa	Endometriosis	2	N	NA	NA	RNA-Seq	320	SRP396798	54457723840	
B.E11.2	16	B	Endometriosis	21	White	follicular	Endometrium	Eutopic Endometrium	2	N	NA	NA	RNA-Seq	320	SRP396798	77665318720	
B.E12	17	B	Endometriosis	34	White	luteal	Endometrium	Eutopic Endometrium	4	N	NA	NA	RNA-Seq	320	SRP396798	67821920000	
B.E13.1	18	B	Endometriosis	52	Black	(perimenopausia)	Endometrium	Eutopic Endometrium	4	Y	Progesterone and estrogen pellets	NA	RNA-Seq	320	SRP396798	42938918720	
B.E13.2	18	B	Endometriosis	52	Black	(perimenopausia)	Ovary	4	Y	Progesterone and estrogen pellets	NA	RNA-Seq	320	SRP396798	46900937280		
B.E14.1	19	B	Endometriosis	37	Black	luteal	Left ovarian cyst	Endometrioma	3	N	NA	NA	RNA-Seq	320	SRP396798	52561188800	
B.E14.2	19	B	Endometriosis	37	Black	luteal	Right ovarian fossa	Endometriosis	3	N	NA	NA	RNA-Seq	320	SRP396798	37968836480	
B.E15	20	B	Endometriosis	46	White	follicular	Left uterosacral ligament	Endometriosis	2	N	NA	NA	RNA-Seq	320	SRP396798	15343936320	
B.E16.1	21	B	Endometriosis	36	White	follicular	Rectum	Endometriosis	3	N	NA	NA	RNA-Seq	320	SRP396798	13228626880	
B.E16.2	21	B	Endometriosis	36	White	follicular	Left ovarian fossa	Endometriosis	3	N	NA	NA	RNA-Seq	320	SRP396798	9774401600	
B.E16.3	21	B	Endometriosis	36	White	follicular	Right uterosacral ligament	Endometriosis	3	N	NA	NA	RNA-Seq	320	SRP396798	14462837760	
B.E16.4	21	B	Endometriosis	36	White	follicular	Anterior cul-de-sac	Endometriosis	3	N	NA	NA	RNA-Seq	320	SRP396798	17323236800	
B.E16.5	21	B	Endometriosis	36	White	follicular	Posterior uterus	Endometriosis	3	N	NA	NA	RNA-Seq	320	SRP396798	13275316800	

**ANEXO II (continuación). Metadata correspondiente al dataset B (muestras de pacientes con endometriosis sin tratamiento hormonal).**

<b>Id</b>	<b>BioProject</b>	<b>BioSample</b>	<b>Bytes</b>	<b>Center_Name</b>	<b>Consent</b>	<b>Experiment</b>	<b>GEO_Accesion_exp.</b>	<b>Instrument</b>	<b>LibraryLayout</b>	<b>LibrarySelection</b>	<b>Method</b>	<b>Organism</b>	<b>Platform</b>	<b>ReleaseDate</b>	<b>create_date</b>	<b>Sample_Name</b>
B.C01	PRJNA879764	SAMN30814206	4132717004		public	SRX17538932	GSM6574535	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-14T00:00:00Z	2023-10-13T23:53:00Z	GSM657453-sample8
B.C02	PRJNA879764	SAMN30814178	9242618332		public	SRX17538955	GSM6574535	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:28:00Z	GSM657453-sample31
B.C03	PRJNA879764	SAMN30814178	19847025908		public	SRX17538959	GSM6574540	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:25:00Z	GSM6574540-sample36
B.E01	PRJNA879764	SAMN30814209	8739513260		public	SRX17538929	GSM6574509	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:40:00Z	GSM6574509-sample1
B.E02	PRJNA879764	SAMN30814208	3105497764		public	SRX17538930	GSM6574510	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:39:00Z	GSM6574510-sample5
B.E03	PRJNA879764	SAMN30814207	3037074707		public	SRX17538931	GSM6574511	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:40:00Z	GSM6574511-sample6
B.E04.1	PRJNA879764	SAMN30814204	7936155440		public	SRX17538934	GSM6574514	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:40:00Z	GSM6574514-sample10
B.E04.2	PRJNA879764	SAMN30814203	13704796002		public	SRX17538935	GSM6574515	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:40:00Z	GSM6574515-sample10
B.E04.3	PRJNA879764	SAMN30814202	47266168898		public	SRX17538939	GSM6574516	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-14T00:00:00Z	2023-10-14T00:06:00Z	GSM6574516-sample12
B.E04.4	PRJNA879764	SAMN30814201	58935373736		public	SRX17538936	GSM6574517	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-14T00:00:00Z	2023-10-14T00:24:00Z	GSM6574517-sample13
B.E05.1	PRJNA879764	SAMN30814200	37676741612		public	SRX17538937	GSM6574518	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:27:00Z	GSM6574518-sample14
B.E05.2	PRJNA879764	SAMN30814199	24777738827		public	SRX17538938	GSM6574519	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:41:00Z	GSM6574519-sample15
B.E05.3	PRJNA879764	SAMN30814198	20883183581		public	SRX17538940	GSM6574520	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:14:00Z	GSM6574520-sample16
B.E05.4	PRJNA879764	SAMN30814197	19009281740		public	SRX17538941	GSM6574521	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:18:00Z	GSM6574521-sample17
B.E06.1	PRJNA879764	SAMN30814196	17180921291		public	SRX17538942	GSM6574522	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:19:00Z	GSM6574522-sample18
B.E06.2	PRJNA879764	SAMN30814195	17945105679		public	SRX17538943	GSM6574523	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:13:00Z	GSM6574523-sample19
B.E06.3	PRJNA879764	SAMN30814194	20760949969		public	SRX17538944	GSM6574524	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:09:00Z	GSM6574524-sample20
B.E06.4	PRJNA879764	SAMN30814193	20098757530		public	SRX17538945	GSM6574525	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:12:00Z	GSM6574525-sample21
B.E06.5	PRJNA879764	SAMN30814192	20465963280		public	SRX17538946	GSM6574526	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:15:00Z	GSM6574526-sample22
B.E06.6	PRJNA879764	SAMN30814191	10439764027		public	SRX17538947	GSM6574527	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:17:00Z	GSM6574527-sample23
B.E07.1	PRJNA879764	SAMN30814190	776916520	LAWRENSON LAB, SAMIR OSCHIN COMPREHENSIVE CANCER INSTITUTE AND WOMEN'S CANCER PROGRAM, CEDARS- SINAI MEDICAL CENTER	public	SRX17538948	GSM6574528	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T23:57:00Z	GSM6574528-sample24
B.E07.2	PRJNA879764	SAMN30814189	9069780194		public	SRX17538949	GSM6574529	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T19:53:00Z	GSM6574529-sample25
B.E07.3	PRJNA879764	SAMN30814188	10390503206		public	SRX17538950	GSM6574530	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T19:55:00Z	GSM6574530-sample26
B.E07.4	PRJNA879764	SAMN30814187	19257643921		public	SRX17538951	GSM6574531	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T20:06:00Z	GSM6574531-sample27
B.E08.1	PRJNA879764	SAMN30814186	8211045490		public	SRX17538952	GSM6574532	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T20:04:00Z	GSM6574532-sample28
B.E08.2	PRJNA879764	SAMN30814185	4042057215		public	SRX17538954	GSM6574533	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T20:39:00Z	GSM6574533-sample29
B.E08.3	PRJNA879764	SAMN30814184	47074658107		public	SRX17538953	GSM6574534	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-16T00:00:00Z	2023-10-16T20:58:00Z	GSM6574534-sample30
B.E09	PRJNA879764	SAMN30814182	29937645151		public	SRX17538956	GSM6574536	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:30:00Z	GSM6574536-sample32
B.E10.1	PRJNA879764	SAMN30814181	22967973543		public	SRX17538963	GSM6574537	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:35:00Z	GSM6574537-sample33
B.E10.2	PRJNA879764	SAMN30814180	14443598859		public	SRX17538957	GSM6574538	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:07:00Z	GSM6574538-sample34
B.E11.1	PRJNA879764	SAMN30814177	21755356392		public	SRX17538960	GSM6574541	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:10:00Z	GSM6574541-sample37
B.E11.2	PRJNA879764	SAMN30814176	30560298743		public	SRX17538961	GSM6574542	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:25:00Z	GSM6574542-sample38
B.E12	PRJNA879764	SAMN30814175	26871998505		public	SRX17538962	GSM6574543	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:35:00Z	GSM6574543-sample39
B.E13.1	PRJNA879764	SAMN30814174	17036010337		public	SRX17538964	GSM6574544	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:15:00Z	GSM6574544-sample40
B.E13.2	PRJNA879764	SAMN30814173	1880341978		public	SRX17538965	GSM6574545	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:23:00Z	GSM6574545-sample41
B.E14.1	PRJNA879764	SAMN30814172	20803416391		public	SRX17538966	GSM6574546	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:25:00Z	GSM6574546-sample42
B.E14.2	PRJNA879764	SAMN30814171	15080705666		public	SRX17538967	GSM6574547	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T23:18:00Z	GSM6574547-sample43
B.E15	PRJNA879764	SAMN30814170	6145037743		public	SRX17538968	GSM6574548	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:47:00Z	GSM6574548-sample44
B.E16.1	PRJNA879764	SAMN30814169	5339257200		public	SRX17538969	GSM6574549	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:45:00Z	GSM6574549-sample45
B.E16.2	PRJNA879764	SAMN30814168	3887693064		public	SRX17538970	GSM6574550	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:39:00Z	GSM6574550-sample46
B.E16.3	PRJNA879764	SAMN30814167	5832874168		public	SRX17538971	GSM6574551	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:44:00Z	GSM6574551-sample47
B.E16.4	PRJNA879764	SAMN30814166	6968237303		public	SRX17538972	GSM6574552	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:47:00Z	GSM6574552-sample48
B.E16.5	PRJNA879764	SAMN30814165	5324794565		public	SRX17538973	GSM6574553	Illumina NovaSeq 6000	SINGLE	cDNA	scRNA-seq	Homo sapiens	ILLUMINA	2023-10-13T00:00:00Z	2023-10-13T22:40:00Z	GSM6574553-sample49

**ANEXO III. Script para preparación de los datos de scRNA-seq con Seurat.** Este anexo detalla el proceso de creación y manipulación de objetos Seurat a partir de matrices de conteo de scRNA-seq para el análisis de células individuales. El código carga los datos en formato H5, crea objetos Seurat, calcula porcentajes de genes mitocondriales y ribosomales, y añade metadatos adicionales para cada muestra. Finalmente, los objetos Seurat procesados se guardan para su posterior análisis, facilitando la visualización y el estudio de la expresión génica específica.

```
### Author: Madalina Alexandra Dodu

## *** PAQUETES ***
install.packages("openxlsx")
install.packages('Seurat')
install.packages("hdf5r")
install.packages("tidyverse")

library(hdf5r)
library("openxlsx")
library('Seurat')
library(tidyverse)

## *** PREPARACIÓN DE LOS OBJETOS SEURAT **

# Especifica el directorio donde están tus archivos HDF5.
files <- list.files(path = "data/raw/nombre_directorio", pattern = "*.h5", full.names = TRUE)

# Extrae los nombres de los archivos sin ruta ni extensión, para utilizarlos como nombres de muestra.
sample_names <- sapply(basename(files), function(x) sub("\\.h5$", "", x))

# Carga el archivo de metadatos desde un archivo Excel.
metadata <- read.xlsx("metadata/metadata.xlsx", sheet = 1)
ids <- metadata$id # Extrae los IDs de las muestras desde la metadata.

# Crea una lista vacía para almacenar los objetos Seurat que se generarán.
seurat_list <- list()

# Bucle para cargar y procesar cada archivo H5 individualmente.
for (i in seq_along(files)) {
  file <- files[i]
  sample_name <- sample_names[i]
  id <- ids[i]

  # Lee los datos del archivo HDF5 y crea un objeto Seurat a partir de ellos.
  seurat_data <- Read10X_h5(file)
  seurat_obj <- CreateSeuratObject(counts = seurat_data,
                                      min.features = 200,
                                      project = sample_name)

  # Renombra las celdas del objeto Seurat para asegurar que cada celda tenga un nombre único.
  cell_ids <- colnames(seurat_obj)
  new_cell_ids <- paste(id, cell_ids, sep = "_")
  seurat_obj <- RenameCells(seurat_obj, new.names = new_cell_ids)

  # Añade el objeto Seurat a la lista de objetos.
  seurat_list <- append(seurat_list, list(seurat_obj))
}

# Limpieza de memoria: Elimina las variables que ya no se necesitan para liberar espacio.
rm(seurat_data, seurat_obj, file, files, i, sample_name, sample_names, cell_ids, id, ids, new_cell_ids)

# Bucle para calcular porcentajes de genes mitocondriales y ribosomales, y añadir metadata a cada objeto Seurat.
for (i in seq_along(seurat_list)) {

  # Extrae el objeto Seurat actual de la lista.
  current_sample <- seurat_list[[i]]

  # Calcula el porcentaje de genes mitocondriales en el objeto Seurat.
  current_sample$percent.mt <- PercentageFeatureSet(object = current_sample, pattern = "^\MT-")
```

```
# Calcula el porcentaje de genes ribosomales en el objeto Seurat.
current_sample$percent.Ribosomal <- PercentageFeatureSet(object = current_sample, pattern = "^\$RP[LS]")

# Obtiene los nombres de las células del objeto Seurat.
cell_ids <- colnames(current_sample)

# Crea etiquetas de muestra basadas en los nombres de las células.
sample_labels <- sapply(strsplit(cell_ids, "_"), `[, 1])

# Crea un dataframe de metadatos con los IDs de las células.
metadata_seurat <- data.frame(row.names = cell_ids, id = sample_labels)

# Guarda los nombres originales de las filas para usarlos más adelante.
original_row_names <- row.names(metadata_seurat)

# Combina los metadatos originales con los nuevos metadatos basados en los IDs.
merged_metadata <- merge(metadata_seurat, metadata, by = "id", all.x = TRUE)

# Restaura los nombres de las filas originales en los metadatos combinados.
row.names(merged_metadata) <- original_row_names

# Añade los metadatos combinados al objeto Seurat.
current_sample <- AddMetaData(current_sample, metadata = merged_metadata)

# Renombra las columnas de 'nCount_RNA' a 'nUMI' y 'nFeature_RNA' a 'nGene' para mayor claridad.
colnames(current_sample@meta.data)[colnames(current_sample@meta.data) == "nCount_RNA"] <- "nUMI"
colnames(current_sample@meta.data)[colnames(current_sample@meta.data) == "nFeature_RNA"] <- "nGene"

# Guarda el objeto Seurat modificado de nuevo en la lista.
seurat_list[[i]] <- current_sample
}

# Visualiza los metadatos del primer objeto Seurat para verificación.
View(seurat_list[[1]]@meta.data)

# Limpieza de memoria: Elimina variables temporales para liberar espacio.
rm(i, metadata, merged_metadata, metadata_seurat, current_sample, cell_ids, original_row_names, sample_labels)

# Guarda la lista de objetos Seurat en un archivo RDS para su uso posterior.
saveRDS(seurat_list, file = "data/processed/seurat_list_raw.rds")
```

**ANEXO IV. Script para filtrado de calidad y eliminación de dobletes en datos scRNA-seq.** Este anexo describe el proceso de filtrado y pre-procesamiento de datos scRNA-seq mediante Seurat, con el objetivo de asegurar la calidad de los datos y eliminar células potencialmente problemáticas como los dobletes. Primero, se realiza un filtrado inicial basado en la cantidad de genes y la proporción de ARN mitocondrial para mantener únicamente células de alta calidad. Luego, se implementa un análisis más detallado para identificar y remover dobletes, utilizando la herramienta DoubletFinder tras realizar una reducción de dimensionalidad mediante PCA y UMAP. Este procedimiento garantiza la limpieza de los datos antes de su análisis posterior.

```
### Author: Madalina Alexandra Dodu

## *** PAQUETES ***
install.packages('Seurat')
install.packages("tidyverse")
install.packages("devtools")

library(devtools)
install_github('chris-mcginnis-ucsf/DoubletFinder')

library(DoubletFinder)
library('Seurat')
library(tidyverse)

## *** CONTROL DE CALIDAD **

## FILTRADO DE CALIDAD
# --- DATASET A ---
# Se filtra cada objeto Seurat en la lista, eliminando células con menos de 500 genes detectados,
# menos de 1000 UMI, más de 100000 UMI, o con un porcentaje de genes mitocondriales mayor al 25%.
seurat_list <- lapply(seurat_list, function(x) {
  subset(x, subset = nGene >= 500 & nUMI >= 1000 & nUMI <= 100000 & percent.mt < 25)
})

# --- DATASET B ---
# Se filtra cada objeto Seurat en la lista, eliminando células con menos de 200 genes detectados y un porcentaje
# mayor al 20%.
seurat_list <- lapply(seurat_list, function(x) {
  subset(x, subset = nGene >= 200 & percent.mt < 20)
})

## FILTRADO DE DOBLETES
# Se inicia un bucle para iterar a través de cada objeto Seurat en la lista con el fin de identificar y eliminar
# dobletes.
for (i in 1:length(seurat_list)) {
  # Imprime el índice de la muestra actual para seguimiento.
  print(paste0("Sample ", i))

  # Pre-procesamiento del objeto Seurat siguiendo el flujo de trabajo estándar de Seurat.
  sample <- NormalizeData(seurat_list[[i]]) # Normaliza los datos de expresión génica.
  sample <- FindVariableFeatures(sample)      # Identifica los genes con mayor variabilidad.
  sample <- ScaleData(sample)                 # Escala los datos para que los valores sean comparables.
  sample <- RunPCA(sample)                   # Realiza un Análisis de Componentes Principales (PCA).

  # Construcción del gráfico de vecinos y agrupamiento.
  sample <- FindNeighbors(object = sample, dims = 1:20) # Encuentra vecinos similares basado en las dimensiones
  especificadas.
  sample <- FindClusters(object = sample)               # Agrupa las células en clústeres basados en similitudes.
  sample <- RunUMAP(sample, dims = 1:20)                # Ejecuta UMAP para visualizar los clústeres en un espacio
  de menor dimensión.

  # Identificación de pK óptimo para DoubletFinder (sin verdad de terreno)
  sweep.list <- paramSweep(sample, PCs = 1:20, sct = FALSE) # Realiza una búsqueda de parámetros para
  DoubletFinder.
  sweep.stats <- summarizeSweep(sweep.list, GT = FALSE)       # Resumen de las estadísticas de la búsqueda de
  parámetros.
  bcmvn <- find.pK(sweep.stats)                            # Identifica el valor óptimo de pK para maximizar la
  métrica BC.
```

```
# Se selecciona el pK óptimo basado en la distribución del coeficiente de bimodalidad.
pK <- bcmvn %>%
  filter(BCmetric == max(BCmetric)) %>%
  select(pK)
pK <- as.numeric(as.character(pK[[1]]))

## Estimación de la proporción de dobletes homotípicos
annotations <- sample@meta.data$seurat_clusters # Extrae las anotaciones de clústeres.
homotypic.prop <- modelHomotypic(annotations) # Modela la proporción de dobletes homotípicos.
nExp.poi <- round(0.075 * nrow(sample@meta.data)) ## Supone una tasa de formación de dobletes del 7.5% - ajustar
según el dataset.
nExp.poi.adj <- round(nExp.poi * (1 - homotypic.prop)) # Ajusta la estimación de dobletes esperados.

# Corre DoubletFinder para identificar dobletes
sample <- doubletFinder(seu = sample,
                           PCs = 1:20,
                           pN = 0.25,
                           pK = pK,
                           nExp = nExp.poi.adj, reuse.pANN = FALSE, sct = FALSE)

# Actualiza el metadata con los resultados de DoubletFinder
metadata <- sample@meta.data
colnames(metadata)[44] <- "doublet_finder" # Renombra la columna para identificar dobletes.
sample@meta.data <- metadata

# Filtra las células etiquetadas como singlets.
last_col_name <- colnames(metadata)[ncol(metadata)]
is_singlet <- FetchData(sample, vars = last_col_name) == "Singlet"

# Subconjunto de células singlets y guardar el resultado en la lista original.
singlets <- subset(sample, cells = which(is_singlet))
seurat_list[[i]] <- singlets
remove(singlets)
}

# Guarda la lista de objetos Seurat filtrados y sin dobletes
saveRDS(seurat_list, file = "data/processed/seurat_list_filtered_noDoublets.rds")

# Limpia el entorno eliminando las variables que ya no son necesarias
rm(bcmvn, is_singlet, metadata, sample, sweep.stats, sweep.list, annotations,
  homotypic.prop, i, last_col_name, nExp.poi, nExp.poi.adj, pK, current_sample)
```

**ANEXO V. Script para normalización y detección del ciclo celular, combinación de datos y clustering en datos de scRNA-seq.** Este anexo describe el proceso de normalización, anotación del ciclo celular, combinación de datos de múltiples muestras, y el posterior análisis de agrupación de datos de scRNA-seq.

```
### Author: Madalina Alexandra Dodu

## *** PAQUETES ***
install.packages('Seurat')
install.packages("tidyverse")

library('Seurat')
library(tidyverse)

## *** NORMALIZACIÓN Y DETERMINACIÓN DEL CICLO CELULAR ***
# Cargar la lista de objetos Seurat previamente filtrados y sin dobletes
seurat_list <- readRDS(file = "data/processed/seurat_list_filtered_noDoublets.rds")

# Cargar las listas de genes asociados a las fases del ciclo celular (S y G2M) (vienen con Seurat)
s.genes <- cc.genes$s.genes
g2m.genes <- cc.genes$g2m.genes

# Bucle para normalizar los datos y determinar la fase del ciclo celular de cada célula
for (i in 1:length(seurat_list)) {
  # Normalización de los datos de expresión génica
  seurat_list[[i]] <- NormalizeData(seurat_list[[i]], verbose = TRUE)

  # Anotación del ciclo celular basado en las fases S y G2M
  seurat_list[[i]] <- CellCycleScoring(seurat_list[[i]], g2m.features = g2m.genes, s.features = s.genes)
}

## *** COMBINACIÓN DE LOS DATOS DE LAS MUESTRAS ***
# Combinar todos los objetos Seurat de la lista en un único objeto para análisis conjunto
combined_dataset <- merge(seurat_list[[1]], y = seurat_list[2:length(seurat_list)],
                           project = "Endometriosis_Dataset")

# Limpiar el metadata eliminando columnas innecesarias creadas durante la eliminación de dobletes
columns_to_remove <- c(grep("^pANN", colnames(combined_dataset@meta.data), value = TRUE),
                       grep("^DF", colnames(combined_dataset@meta.data), value = TRUE))

# Eliminar las columnas de la metadata
combined_dataset@meta.data <- combined_dataset@meta.data[, !(colnames(combined_dataset@meta.data)
                                                       %in% columns_to_remove)] 

# Guardar el objeto combinado y limpio
saveRDS(combined_dataset, file = "data/processed/seurat_merged_postQC_cc.rds")

## *** CLUSTERING ***
# Identificación de las características genéticas más variables para su uso en clustering
combined_dataset <- FindVariableFeatures(combined_dataset, selection.method = "vst", nfeatures = 2000)

# Escalar los datos para que los valores sean comparables entre sí
combined_dataset <- ScaleData(combined_dataset, features = rownames(combined_dataset)) # Dataset A
combined_dataset <- ScaleData(combined_dataset) # Dataset B

# Realizar un Análisis de Componentes Principales (PCA) para reducir la dimensionalidad del dataset
combined_dataset <- RunPCA(combined_dataset, features = VariableFeatures(combined_dataset))

# Guardar el objeto Seurat después de PCA para posibles análisis posteriores
saveRDS(combined_dataset, file = "data/processed/seurat_merged_postPCA.rds")

# Volver a cargar el objeto Seurat post-PCA si es necesario
combined_dataset <- readRDS("data/processed/seurat_merged_postPCA.rds")

# Realizar el clustering en base a los resultados del PCA
ElbowPlot(combined_dataset) # Se utiliza el codo para determinar el número óptimo de PCs para el clustering

# Encontrar vecinos cercanos y realizar clustering de las células en base a 10 dimensiones principales
combined_dataset <- FindNeighbors(combined_dataset, dims = 1:10, verbose = FALSE)
```



```
combined_dataset <- FindClusters(combined_dataset,
                                resolution = c(0.4, 0.6, 0.8, 1.0, 1.4))

# Ejecutar UMAP para visualizar los clústeres en un espacio de menor dimensión
combined_dataset <- RunUMAP(combined_dataset, dims = 1:10, reduction = "pca", verbose = FALSE)

# Guardar el objeto Seurat final con los clústeres identificados
saveRDS(combined_dataset, file = "data/processed/seurat_merged_clustered.rds")
```



**ANEXO VI. Script para anotación de clústeres en un objeto Seurat.** Este script realiza la anotación de clústeres en un conjunto de datos Seurat combinado. Primero, une las capas del conjunto de datos para la anotación de clústeres y guarda el objeto resultante. Luego, realiza un análisis de marcadores de clústeres utilizando la prueba de Wilcoxon, identifica los principales 10 marcadores y guarda los resultados en un archivo Excel.

```
### Author: Madalina Alexandra Dodu

## *** PAQUETES ***
install.packages("openxlsx")
install.packages('Seurat')
install.packages("devtools")
install.packages("tidyverse")

library(devtools)
install_github("immunogenomics/presto")

library("openxlsx")
library('Seurat')
library(tidyverse)
library("presto")

## *** ANOTACIÓN DE CLÚSTERES **

# Combinar diferentes capas del conjunto de datos Seurat en un solo objeto
joined <- JoinLayers(combined_dataset) # Presto da error si no se realiza esta unión

# Guardar el objeto unido como un archivo RDS para su posterior uso
saveRDS(joined, file = "data/processed/joined_clustered.rds")

# Análisis de marcadores de clústeres utilizando la prueba de Wilcoxon. Los grupos se especifican mediante el
# parámetro `group_by`, que en este caso es "RNA_snn_res.0.4".
markers <- wilcoxauc(joined, group_by = "RNA_snn_res.0.4")

# Identificación de los principales 10 marcadores de cada clúster
# Se seleccionan los 10 principales marcadores que están presentes en al menos el 70% de las observaciones dentro
# de cada grupo.
top10_markers <- top_markers(markers, n = 10, pct_in_min = 70)

# Guardar los resultados de los principales marcadores en un archivo Excel para su revisión y análisis
write.xlsx(top10_markers, file = "results/top10_markers.xlsx")

# Eliminación de clústeres de mala calidad, en este caso, clúster 15
Idents(object = joined) <- "RNA_snn_res.0.4"
joined <- subset(joined, idents = 15, invert = TRUE)

# Importación de la tabla con la anotación
annot_df <- read.xlsx("results/top10_annot.xlsx", sheet = 2)

# ANOTACIÓN
levels(Idents(joined)) <- as.numeric(levels(Idents(joined)))

if(all(levels(Idents(joined)) %in% annot_df$cluster)) {
  # Crear un vector para los nuevos niveles de acuerdo con el orden en annot_df
  new_levels <- annot_df$annotation_abv[match(levels(Idents(joined)), annot_df$cluster)]
  levels(Idents(joined)) <- new_levels
} else {
  stop("Los niveles de los clústeres no coinciden con el DataFrame.")
}

# Adición de la anotación a una nueva columna denominada `cell_type`
joined$cell_type <- Idents(joined)
```

**ANEXO VII.** Códigos empleados para la generación del gráfico de codo y UMAPs sin y con anotación, empleando colores personalizados.

```
### Author: Madalina Alexandra Dodu

## *** PAQUETES ***
install.packages("openxlsx")
install.packages('Seurat')
install.packages("tidyverse")

install.packages("devtools")
devtools::install_github("thomasp85/patchwork")

library("openxlsx")
library('Seurat')
library(tidyverse)
library(RColorBrewer)
library("patchwork")

## ELBOW PLOTS
A_joined_raw <- readRDS(file = "data/processed/A_seurat_list_raw.rds")
B_joined_raw <- readRDS(file = "data/processed/B_seurat_list_raw.rds")

elbowA <- ElbowPlot(A_joined_raw) + theme_article() +
  ggtitle("Dataset A")
# annotate("rect", xmin = 6, xmax = 12.5, ymin = 3 , ymax = 5.5, alpha = 0, color="pink") # añade recuadro

elbowB <- ElbowPlot(B_joined_raw) + theme_article() +
  ggtitle("Dataset B")
# annotate("rect", xmin = 5.5, xmax = 12.5, ymin = 2.2 , ymax = 4.2, alpha = 0, color="pink") # añade recuadro

elbow_plots <- (elbowA | elbowB) +
  plot_annotation(tag_levels = 'A') &
  theme(plot.tag = element_text(size = 14, face = "bold"))

# Guardar los plots en alta resolución
ggsave(filename = "results/elbow_plots.png",
       plot = elbow_plots,
       width = 40,           # Ancho en centímetros
       height = 16,          # Alto en centímetros
       units = "cm",         # Unidades en centímetros
       dpi = 300)

## VISUALIZACIÓN UMAP
A_joined <- readRDS(file = "data/processed/A_seurat_merged_clustered.rds")
Idents(object = A_joined) <- "RNA_snn_res.0.4"

B_joined <- readRDS(file = "data/processed/B_seurat_merged_clustered.rds")
Idents(object = B_joined) <- "RNA_snn_res.0.4"

# Ordena los clústeres para que aparezcan luego ordenados en la leyenda
cluster_numeric <- as.numeric(as.character(Idents(A_joined)))
A_joined$RNA_snn_res.0.4 <- factor(as.character(cluster_numeric),
                                      levels = sort(unique(cluster_numeric)))

cluster_numeric <- as.numeric(as.character(Idents(B_joined)))
B_joined$RNA_snn_res.0.4 <- factor(as.character(cluster_numeric),
                                      levels = sort(unique(cluster_numeric)))

A_umap <- DimPlot(A_joined,
                    reduction = "umap",
                    label = TRUE,
                    label.size = 5,
                    raster=FALSE) +
  labs(
    x = "UMAP_1",           # Nombre del eje X
    y = "UMAP_2",           # Nombre del eje Y
  ) +
```

```

ggtitle("Clústeres en el dataset A") +
theme_article() +
theme(
  legend.text = element_text(size = 11) # Tamaño del texto de la leyenda
)
print(A_umap)

B_umap <- DimPlot(B_joined,
                    reduction = "umap",
                    label = TRUE,
                    label.size = 5,
                    raster=FALSE) +
labs(
  x = "UMAP_1", # Nombre del eje X
  y = "UMAP_2", # Nombre del eje Y
) +
ggtitle("Clústeres en el dataset B") +
theme_article() +
theme(
  legend.text = element_text(size = 11), # Tamaño del texto de la leyenda
)
print(B_umap)

## UMAPS FASE DEL CICLO CELULAR
A_plot_phases <- DimPlot(A_joined,
                           reduction = "umap", group.by = "Phase") +
labs(
  x = "UMAP_1", # Nombre del eje X
  y = "UMAP_2", # Nombre del eje Y
) +
ggtitle("Fases del ciclo celular (dataset A)") +
theme_article() +
theme(
  legend.text = element_text(size = 11), # Tamaño del texto de la leyenda
)
print(A_plot_phases)

B_plot_phases <- DimPlot(B_joined,
                           reduction = "umap", group.by = "Phase",
                           raster=FALSE) +
labs(
  x = "UMAP_1", # Nombre del eje X
  y = "UMAP_2", # Nombre del eje Y
) +
ggtitle("Fases del ciclo celular (Dataset B)") +
theme_article() +
theme(
  legend.text = element_text(size = 11), # Tamaño del texto de la leyenda
)
print(B_plot_phases)

## UMAPS CON ANOTACIÓN
# Variables con colores para tener los tipos celulares/condiciones siempre del mismo color
cell_colors <- c(
  # endoteliales
  "Cél. endoteliales" = "#FF3333",

  #estromales
  "Cél. estromales endom. (sec.)" = "#F871FF",
  "Cél. estromales endom." = "#B83779FF",
  "Cél. estromales endom. (prolif.)" = "#EE2B6D",
  "Cél. glandulares" = "#FAB0B0",
  "Cél. estrom. decid." = "#822581FF",
  "Cél. estromales endom. (menstr.)" = "#FF65AC",
  "Cél. estromales" = "#B83779FF",

  #inmunes
  "Cél. T" = "#1AD1D3FF",
)

```

```

"Cél. γ/δ & NK" = "#00CCFF",
"Monocitos" = "#B684FF",
"Macrófagos" = "#8B93FF",
"Cél. B" = "#2B748EFF",
"Cél. dendríticas" = "#4770E8FF",
"Cél. plasmáticas" = "#42A0FF",
"Cél. Inmunes" = "#00B5ED",
"Cél. inmunes" = "#00B5ED",

#epiteliales
"Cél. no ciliadas" = "#FAF400",
"Cél. epiteliales" = "#FFD521",
"Cél. no ciliadas SPDEF+" = "#FEA832FF",
"Cél. ciliadas" = "#FB8222FF",
"Cél. epitel. prolif." = "#FEAB30",

#musculares
"Cél. musc. lisas" = "#DF6767",

#fibroblastos
"Fibroblastos estrom." = "#85C441",
"Fibroblastos" = "#1B8E00",
"Fibroblastos estrom. C7+" = "#33CC33"
)

treatment_colors <- c(
  "Control_N" = "#26828EFF",
  "Control_Y" = "#B4DE2CFF",
  "Endometriosis_N" = "#721F81FF",
  "Endometriosis_Y" = "#D6456CFF"
)

A_joined <- readRDS("data/processed/A_joined_clustered.rds")
B_joined <- readRDS("data/processed/B_joined_clustered.rds")

Idents(object = A_joined) <- "cell_type"
Idents(object = B_joined) <- "cell_type"

A_umap_ann <- DimPlot(A_joined, reduction = "umap", cols = cell_colors) +
  labs(
    x = "UMAP_1",           # Nombre del eje X
    y = "UMAP_2"            # Nombre del eje Y
  ) +
  ggtitle("Dataset A") +
  theme_article() +
  theme(
    legend.text = element_text(size = 11)  # Tamaño del texto de la leyenda
  )
A_umap_ann <- LabelClusters(A_umap_ann, id = "ident", fontface = "bold")
print(A_umap_ann)

B_umap_ann <- DimPlot(B_joined, reduction = "umap", cols = cell_colors, raster=FALSE) +
  labs(
    x = "UMAP_1",           # Nombre del eje X
    y = "UMAP_2"            # Nombre del eje Y
  ) +
  ggtitle("Dataset B") +
  theme_article() +
  theme(
    legend.text = element_text(size = 11)  # Tamaño del texto de la leyenda
  )
B_umap_ann <- LabelClusters(B_umap_ann, id = "ident", fontface = "bold")
print(B_umap_ann)

```

**ANEXO VIII.** Script para generar una firma génica asociada al dolor en dos conjuntos de datos de scRNA-seq (A y B) y caracterización de tipos celulares con firma del dolor y genes asociados a este síntoma. El objetivo es identificar los genes relacionados con el dolor presentes en diferentes tipos celulares dentro de estos datasets, utilizando la puntuación de módulos UCell. Además, se realiza un análisis de expresión diferencial entre los tipos celulares clave y se extraen los genes más significativos de cada uno. El script también procesa los datos para escalar las expresiones y eliminar clústeres, lo que optimiza la detección de los genes de dolor en tipos celulares específicos. Los resultados obtenidos incluyen los 5 genes más representativos para cada tipo celular de ambos datasets, que pueden estar involucrados en la señalización de dolor en el contexto de la endometriosis.

```
## *** CARACTERIZACIÓN DE TIPOS CELULARES CON FIRMA DE DOLOR - DATASET A ***

# Definir los tipos celulares para el análisis de expresión diferencial de genes
# relacionados con el dolor en dataset A
A_painGenes_typeCells_expression <- c("Cél. estromales endom. (sec.)", # Subtipos celulares clave
                                         "Monocitos",
                                         "Macrófagos",
                                         "Fibroblastos estrom.")

# Inicializar una lista para almacenar los resultados de la expresión diferencial por tipo celular
A_DifferentialResults_pain <- list()

# Realizar la escalación de datos tras la selección de las características más variables,
# después de haber eliminado anteriormente los clústeres de baja calidad
A_scores <- FindVariableFeatures(A_scores, selection.method = "vst", nfeatures = 2000)
A_scores <- ScaleData(A_scores, features = rownames(A_scores))

# Guardar el objeto con las puntuaciones actualizadas
saveRDS(A_scores, file = "data/processed/A_pain_scores_UCell.rds")

## *** CARACTERIZACIÓN DE GENES DE LA VÍA DEL DOLOR - DATASET A ***

# Bucle que realiza el análisis de expresión diferencial por cada tipo celular en el dataset A
for (cell_type in A_painGenes_typeCells_expression) {
  A_DifferentialResults_pain[[cell_type]] <- FindMarkers(A_scores, ident.1 = cell_type,
                                                          features = A_pain_genes_in_seurat)
}

# Extraer los 5 genes más significativos de cada tipo celular
top_genes_pain_A <- unique(unlist(lapply(A_DifferentialResults_pain,
                                           function(x) head(rownames(x[order(x$p_val_adj),]), 5)))))

# Filtrar las células de interés para los análisis posteriores y escalar los datos de los genes seleccionados
A_scores_Pain_Cells <- subset(A_scores, idents = A_painGenes_typeCells_expression)
A_scores_Pain_Cells <- ScaleData(A_scores_Pain_Cells, features = top_genes_pain_A)

## *** CARACTERIZACIÓN DE TIPOS CELULARES CON FIRMA DE DOLOR - DATASET B ***

# Repetir los pasos anteriores para el dataset B, ajustando a los tipos celulares específicos de B
Idents(B_scores) <- "cell_type"
B_painGenes_typeCells_expression <- c("Cél. estromales endom. (sec.)",
                                         "Cél. estrom. decid.",
                                         "Cél. estromales endom. (menstr.)",
                                         "Cél. estromales endom. (prolif.)",
                                         "Monocitos",
                                         "Fibroblastos estrom. C7+",
                                         "Fibroblastos estrom.")

# Inicializar lista para almacenar resultados del dataset B
B_DifferentialResults_pain <- list()

# Escalar los datos del dataset B tras la selección de las características más variables
B_scores <- FindVariableFeatures(B_scores, selection.method = "vst", nfeatures = 2000)
B_scores <- ScaleData(B_scores)

# Guardar el objeto actualizado para dataset B
saveRDS(B_scores, file = "data/processed/B_pain_scores_UCell.rds")

## *** CARACTERIZACIÓN DE GENES DE LA VÍA DEL DOLOR - DATASET B ***

# Bucle para realizar el análisis de expresión diferencial en cada tipo celular de dataset B
for (cell_type in B_painGenes_typeCells_expression) {
  B_DifferentialResults_pain[[cell_type]] <- FindMarkers(B_scores, ident.1 = cell_type,
                                                          features = B_pain_genes_in_seurat)
}

# Extraer los 5 genes más significativos de cada tipo celular
top_genes_pain_B <- unique(unlist(lapply(B_DifferentialResults_pain,
                                           function(x) head(rownames(x[order(x$p_val_adj),]), 5))))
```

```
# Filtrar y escalar los datos de los genes seleccionados en células relevantes para el dataset B
B_scores_Pain_Cells <- subset(B_scores, idents = B_painGenes_typeCells_expression)
B_scores_Pain_Cells <- ScaleData(B_scores_Pain_Cells, features = top_genes_pain_B)

## *** VISUALIZACIÓN ***

# HEATMAPS para ver expresión de genes
heatmapA_pain.Cells <- DoHeatmap(A_scores_Pain_Cells, features = top_genes_pain_A,
                                    group.bar = TRUE, size = 3, group.colors = cell_colors)
heatmapA_pain.Cells <- heatmapA_pain.Cells +
  guides(color = FALSE) +
  labs(fill = "Expresión",
       y = "Genes marcadores relacionados con el dolor",
       title = "Expresión en las celulares con firma de genes del dolor (Dataset A)") +
  scale_fill_viridis(option="magma") +
  theme(text = element_text(face = "bold"))

heatmapB_pain.Cells <- DoHeatmap(B_scores_Pain_Cells, features = top_genes_pain_B,
                                    group.bar = TRUE, size = 3, group.colors = cell_colors)
heatmapB_pain.Cells <- heatmapB_pain.Cells +
  guides(color = FALSE) +
  labs(fill = "Expresión",
       y = "Genes marcadores relacionados con el dolor",
       title = "Expresión en células con firma de genes del dolor (Dataset B)") +
  scale_fill_viridis(option="magma") +
  theme(text = element_text(face = "bold"))

# DOTPLOTS para visualización de expresión por tratamiento y condición
A_dotplot <- DotPlot(A_scores_Pain_Cells, features = top_genes_pain_A,
                      group.by = "cell_type", split.by = "condition_treatment",
                      cols = "Spectral") +
  labs(
    y = "Tipo celular_Condición_Tratamiento",
    x = "Genes asociados al dolor"
  ) + theme_article() +
  theme(
    axis.title.x = element_text(size = 12, face = "bold"),
    axis.title.y = element_text(size = 12, face = "bold"),
    legend.title = element_text(size = 10, face = "bold"),
  ) + RotatedAxis() +
  guides(size = guide_legend(title = "% expresado (céls.)"),
         colour = guide_colourbar("Expresión media"))
print(A_dotplot)

B_dotplot <- DotPlot(B_scores_Pain_Cells, features = top_genes_pain_B,
                      group.by = "cell_type", split.by = "condition_treatment",
                      cols = "Spectral") +
  labs(
    y = "Tipo celular_Condición_Tratamiento",
    x = "Genes asociados al dolor"
  ) + theme_article() +
  theme(
    axis.title.x = element_text(size = 12, face = "bold"),
    axis.title.y = element_text(size = 12, face = "bold"),
    legend.title = element_text(size = 10, face = "bold"),
  ) + RotatedAxis() +
  guides(size = guide_legend(title = "% expresado (céls.)"),
         colour = guide_colourbar("Expresión media"))
print(B_dotplot)
```

**ANEXO IX.** Top 10 genes diferencialmente expresados en cada clúster obtenido en cada conjunto de datos (A y B). Los clústeres 18 del dataset A y 15 del dataset B son los que se han descartado para los próximos pasos del análisis.

**ANEXO X.** Representación de los primeros genes diferencialmente expresados de cada clúster.

