



Caracterización del perfil de expresión génica mediante secuenciación de próxima generación y análisis bioinformático en células mononucleares de sangre periférica de pacientes colombianos con artritis reumatoide y su relación serológica.

Yordan Daniel Rodríguez Mejía.

Universidad Colegio Mayor de Cundinamarca
Facultad de Ciencias de la Salud – Programa de Bacteriología y Laboratorio Clínico.
Bogotá D.C, Colombia
2026.

Caracterización del perfil de expresión génica mediante secuenciación de próxima generación y análisis bioinformático en células mononucleares de sangre periférica de pacientes colombianos con artritis reumatoide y su relación serológica.

Yordan Daniel Rodríguez Mejía.

Trabajo de grado experimental para optar al título de Bacteriólogo y laboratorista clínico.

Director (a):

MSc. David Hernández Mejía.

Codirector (a):

MSc. Omaira Yaneth López Castro.

Grupo de Investigación:

Laboratorio de Ciencias Ómicas y Medicina personalizada (OmiMed) – Instituto Distrital de Ciencia, Biotecnología e Innovación en Salud (IDCBIS)

Universidad Colegio Mayor de Cundinamarca.

Facultad de Ciencias de la Salud - Programa de Bacteriología y Laboratorio Clínico.

Bogotá D.C, Colombia

2026.

Dedicatoria

A Dios:

Gracias por nunca soltarme, por cada día llevarme por su camino y hacerme entender que es a su manera y no a la mía, por ayudarme a seguir adelante y por darme la valentía de salir de lugares de donde nunca quise irme, para entender que había grandes cosas esperándome. A Dios gracias, esto es gracias a ti.

A mi mami:

Quien fue, es y seguirá siendo el más grande ejemplo de amor, resiliencia y superación día tras día. A la mujer que siempre ha creído en mí, incluso cuando ni yo lo hacía. Al amor de mi vida, mi motor, mi motivo de alegría y mi más grande felicidad. Nunca me cansare de agradecerte por tu esfuerzo y amor durante este largo camino, gracias por tu gran esfuerzo cada mañana, gracias por tus palabras y tus bendiciones. Las palabras jamás me alcanzaran para expresar todo el agradecimiento y el infinito amor que siento hacia ti mami. Sin tus esfuerzos y sacrificios esto jamás hubiera sido posible, este logro es tan tuyo como mío, todo lo que soy es gracias a ti, te amo mamita.

A mis hermanos:

Gracias por ser ejemplo de esfuerzo, sacrificio y valentía. A Brandon, Bairon y Duván, gracias por su cariño, nunca me sentí solo y siempre supe que tenía unos hermanos y una madre atrás mío apoyándose y sintiéndose muy orgullosos de mí, aunque no se los diga seguido, los amo y apoyo y esfuerzo, son ustedes ejemplo en mi vida y espero sientan este logro tan suyo como yo lo siento mío, sin su gran apoyo no lo hubiera logrado. Gracias porque desde la distancia son muy importantes para mí.

A mí:

Finalmente, y no menos importante, gracias, Yordan Daniel por tu valentía y entrega cada día, por siempre querer dar un poco más, por tu gran curiosidad y amor por la ciencia. Hoy miro atrás y veo que los sueños que tuviste se hicieron realidad porque trabajaste duro y te esforzaste día a día.

Agradecimientos

Agradecer a la Universidad Colegio Mayor de Cundinamarca, es especial a la Profe Omaira por haber confiado en mi trabajo y en mis capacidades para realizar este trabajo. Espero el camino recompense tu ayuda y gran corazón.

Agradecer infinitamente al Instituto Distrital de Ciencia, Biotecnología e Innovación en Salud (IDCBIS) por abrirme sus puertas para ser tesista y practicante, gracias por su gran apoyo, enseñanza y aporte a mi proceso formativo. Gracias por permitirme llevar a cabo mi tesis en sus instalaciones y acompañado de grandes profesionales que admiro y respeto.

Quiero dar un agradecimiento especial a David Hernández, gracias por tu gran esfuerzo para el desarrollo de esta tesis, gracias por todas tus enseñanzas, gracias por la confianza depositada en mí para llevar a cabo el desarrollo de esta tesis. Gracias por ser una gran persona, por estar dispuesto siempre a enseñar y a mostrarme desde el ejemplo el profesional en el que me quiero convertir en algunos años. Gracias David, sin tu esfuerzo y ayuda este trabajo no hubiera sido posible.

Gracias a Nico por su gran ayuda en el análisis de datos, gracias por siempre estar ahí para ayudarme y guiarme, tu ayuda fue valiosísima. Gracias a Martín por su ayuda en el cultivo celular y reactivación de las células, gracias por tu gran apoyo y compartir todos tus aprendizajes conmigo, tu ayuda fue importantísima en un momento crucial de este trabajo, más allá de eso, gracias por tu amistad y escucharme cuando no sabía que hacer, gracias por las risas y los momentos compartidos.

Al Doctor Gustavo Salguero y el Doctor Gerardo Quintana por confiar en mi trabajo y permitirme hacer parte de este macroproyecto que ayudara a impactar positivamente la vida y bienestar de muchas personas en Colombia. A la Universidad Nacional y A la Fundación Santa Fe de Bogotá por su ayuda en la recolección de las muestras y poner a disposición la información para llevar a cabo este trabajo.

Gracias a mis amigos de la carrera, sin ustedes no hubiera sobrevivido a cinco años de estudio, gracias por ser familia y refugio durante este tiempo, son muy importantes para mí, los quiero mucho. A Julián, gracias por su apoyo en el momento más difícil de mi vida y hacerme saber que estaba para grandes cosas, gracias por tu amor y cariño, esto también es tuyo, desde la distancia.

Finalmente, quiero agradecer a una persona muy especial y que marcó un antes y un después en mi formación académica, el PhD. Fabian Espitia, gracias por abrirme las puertas de tu laboratorio un día y permitirme desarrollarme en el mundo de la ciencia, gracias por tu enseñanza, por tu guía, por el cariño y por el amor con el que día a día me ayudaste a ser mejor persona y a su vez profesional. Espero que el tiempo y el destino nos permita volver a trabajar juntos.

Resumen

El presente estudio tiene como objetivo caracterizar el perfil de expresión génica en células mononucleares de sangre periférica (PBMC) de pacientes colombianos con artritis reumatoide (AR), agrupados según su endofenotipo serológico. Se empleó la secuenciación masiva de ARN (Bulk RNA-Seq) para identificar patrones diferenciales de expresión génica asociados con la actividad clínica y las características serológicas de la enfermedad. La cohorte de pacientes estuvo conformada por 39 pacientes, con al menos cuatro individuos por cada uno de los ocho endofenotipos definidos según la combinación de los marcadores FR, ACPA y ANAs. Se realizó un análisis de expresión diferencial que incluyó control de calidad de lecturas, alineamiento contra el genoma humano de referencia (GRCh 38.p14), cuantificación con featureCounts y análisis diferencial mediante DESeq2 en R. Se identificaron genes diferencialmente expresados que están asociados con endofenotipos serológicos específicos, contribuyendo a la comprensión de la disregulación inmunológica característica de la AR.

Palabras clave: Artritis reumatoide; expresión génica; RNA-seq; endofenotipos serológicos; PBMC; transcriptómica.

Abstract.

The aim of this study is to characterize the gene expression profile in peripheral blood mononuclear cells (PBMCs) from Colombian patients with rheumatoid arthritis (RA), grouped according to their serological endophenotype. Bulk RNA-Seq was used to identify differential gene expression patterns associated with clinical activity and the serological characteristics of the disease. The patient cohort consisted of 39 patients, with at least four individuals for each of the eight endophenotypes defined based on the combination of the FR, ACPA, and ANA markers. A differential expression analysis was performed, including read quality control, alignment against the human reference genome (GRCh 38.p14), quantification with featureCounts, and differential analysis using DESeq2 in R. Differentially expressed genes associated with specific serological endophenotypes were identified, contributing to the understanding of the immunological dysregulation characteristic of RA.

Keywords: Rheumatoid arthritis; gene expression; RNA-seq; serological endophenotypes; PBMC; transcriptomics.

Tabla de contenido

1. Planteamiento del problema	9
2. Justificación	9
3. Objetivos.....	10
Objetivo general.	10
Objetivos específicos.....	10
4. Marco teórico.	11
4.1 Panorama general de la artritis reumatoide.	11
4.2 Bases celulares y moleculares de artritis reumatoide.....	12
4.3. Mecanismos moleculares de pérdida de tolerancia	18
4.4 Transcriptómica en artritis reumatoide.	21
5. Estado del arte.....	23
6. Metodología.	24
6.1 Tipo de muestras.	24
6.2 Clasificación de las muestras.	24
6.3 Extracción de ARN.	25
6.4 Cuantificación y análisis de la integridad del ARN.....	25
6.5. Preparación de las librerías.	26
6.6. Secuenciación.....	26
6.7. Análisis bioinformático.	26
6.8. Análisis estadístico.....	26
7. Resultados.....	28
7.1. Caracterización clínica y serológica de la cohorte de pacientes.	28
7.2. Asociación entre los marcadores serológicos y la actividad clínica de la enfermedad.	28
7.3. Actividad clínica de la enfermedad según endofenotipo.	29
7.4. Estructura global del perfil transcriptómico de las muestras según endofenotipo clínico. 34	
7.5. Patrones de expresión génica intragrupo según endofenotipo clínico.	37
7.6. Análisis comparativo de la expresión génica entre endofenotipos clínicos y optimización del conjunto de datos.	39
7.7. Reevaluación de la estructura transcriptómica tras la depuración del conjunto de datos.	43
7.8. Análisis de enriquecimiento funcional de los genes diferencialmente expresados.....	48
7.9. Reclasificación transcriptómica de las muestras basada en perfiles de expresión génica. 54	
8. Discusión	69
9. Conclusiones	72
10. Limitaciones del estudio.....	73
11. Bibliografía.....	73

Tabla de Figuras

Figura 1. Tasa de incidencia y prevalencia de la artritis reumatoide según la edad y el sexo..	12
Figura 2. Papel hipotético de la citrulinación en la AR.....	15
Figura 3. Interacción intercelular en la patología de la AR.....	17
Figura 4. La reacción química de la citrulinación	18
Figura 5. Relación entre los NET y la autoinmunidad	20
Figura 6. Relación entre los títulos de ANAs y la AR.....	21
Figura 7. Correlación entre los marcadores serológicos FR, ACPA y ANA con la actividad clínica de la enfermedad medida por CDAI	29
Figura 8. Distribución del factor reumatoide (FR) según endofenotipo clínico.....	30
Figura 9. Distribución de los niveles de ACPA según endofenotipo clínico.....	37
Figura 10. Distribución de los niveles de ANAs según endofenotipo clínico	32
Figura 11. Distribución del CDAI según endofenotipo clínico	33
Figura 12. Comparaciones con diferencias serológicas significativas entre endofenotipos clínicos..	34
Figura 13. Análisis de componentes principales (PCA) de las 39 muestras incluidas en el estudio..	35
Figura 14. Heatmap global de expresión génica basado en los 2000 genes con mayor variabilidad.....	36
Figura 15. Matriz de correlación entre muestras basada en la matriz de expresión transformada mediante VST	37
Figura 16. Heatmaps intragrupo de expresión génica en PBMC según endofenotipo clínico..	38
Figura 14. PCA de las comparaciones entre el endofenotipo 1 y los demás endofenotipos clínicos.	40
Figura 15. Análisis de expresión diferencial entre el endofenotipo 1 y los demás endofenotipos clínicos	42
Figura 16. PCA de las comparaciones entre el endofenotipo 1 y los demás endofenotipos clínicos tras la depuración del conjunto de datos.	44
Figura 17. Número de genes diferencialmente expresados entre el endofenotipo 1 y los endofenotipos 2 a 8.....	45
Figura 18. Volcano plots de la expresión génica diferencial entre el endofenotipo 1 y los endofenotipos 2 a 7.....	47
Figura 19. Ranking de los 10 genes diferencialmente expresados más significativos en cada comparación entre el endofenotipo 1 y los endofenotipos 2 a 7	54
Figura 20. Enriquecimiento funcional de procesos biológicos (GO-BP) en comparaciones entre endofenotipos clínicos.	49
Figura 21. Enriquecimiento funcional de procesos biológicos (GO-BP) en comparaciones entre endofenotipos clínicos	51

Figura 22. Enriquecimiento funcional de los genes diferencialmente expresados basado en términos de Gene Ontology (Proceso biológico) para las comparaciones de los endofenotipos ENDO_1 vs ENDO_7.....	52
Figura 23. Enriquecimiento de rutas KEGG de los genes diferencialmente expresados en las comparaciones ENDO_2–ENDO_7 frente a ENDO_1.	53
Figura 24. Análisis de componentes principales (PCA) de las muestras basado en datos de expresión génica transformados mediante VST, tras la reclasificación en clústers transcriptómicos (CLUS_1–CLUS_4). Cada punto representa una muestra individual, coloreada según su asignación a clúster.....	55
Figura 25. Heatmap global de expresión génica basado en los 2000 genes más variables tras transformación VST de los datos de RNA-Seq.	56
Figura 26. Análisis de componentes principales (PCA) por pares entre los clústers transcriptómicos reclasificados (CLUS_1–CLUS_4).....	57
Figura 27. Número de genes diferencialmente expresados (DEGs) en las comparaciones por pares entre los clusters transcriptómicos (CLUS_1–CLUS_4).....	59
Figura 28. Volcano plots de expresión génica diferencial para las comparaciones entre clusters (CLUS_1–CLUS_4).....	61
Figura 29. Ranking de los 10 genes diferencialmente expresados más significativos en las comparaciones por pares entre los clústeres transcriptómicos (CLUS_1–CLUS_4).	61
Figura 30. Enriquecimiento funcional basado en Gene Ontology (proceso biológico, BP) para las comparaciones entre clústers (CLUS_1–CLUS_4)..	63
Figura 31. Redes funcionales de enriquecimiento basadas en Gene Ontology para procesos biológicos (Biological Process, BP) en las comparaciones entre clústers transcriptómicos (CLUS_1–CLUS_4).....	64
Figura 32. Enriquecimiento de rutas KEGG de los genes diferencialmente expresados en las comparaciones entre clusters transcriptómicos (CLUS_1–CLUS_4).	65
Figura 33. Distribución de FR, ACPA y ANAs según clúster. Se presenta la distribución de los niveles de factor reumatoide (FR), anticuerpos anti-péptidos citrulinados (ACPA) y anticuerpos antinucleares (ANAs) entre los clústeres transcriptómicos.....	67
Figura 34. Distribución del índice de actividad clínica de la enfermedad (CDAI) según los clústeres transcriptómicos (CLUS_1–CLUS_4).....	69

Lista de tablas

Tabla 1. Clasificación por endofenotipos.....	25
Tabla 2. Características clínicas basales de los pacientes en el estudio.	28
Tabla 3. Muestras con perfiles de expresión génica discordantes respecto a su endofenotipo serológico.....	39
Tabla 4. Distribución de las muestras por clúster transcriptómico (CLUS_1–CLUS_4) y endofenotipo serológico.	54

1. Planteamiento del problema

La artritis reumatoide (AR) es una enfermedad autoinmune sistémica caracterizada por inflamación sinovial persistente, destrucción progresiva del cartílago y del hueso, y una elevada heterogeneidad clínica y serológica. Esta heterogeneidad se traduce en variabilidad en la presentación clínica, la progresión del daño articular y la respuesta terapéutica, lo que representa un desafío para el manejo individualizado de los pacientes (1). Durante mucho tiempo los marcadores serológicos clásicos utilizados en la rutina clínica —factor reumatoide (FR) y anticuerpos anti-péptidos cíclicos citrulinados (ACPA)— aunque han sido fundamentales para el diagnóstico, no siempre permiten tener un panorama claro del estado del paciente ni predecir de manera precisa la severidad de la enfermedad ni la respuesta a los tratamientos (2).

Estudios recientes sugieren que la integración de nuevos marcadores, como los anticuerpos antinucleares (ANA) y los patrones de expresión génica, podría mejorarla estratificación de los pacientes y aportar información sobre los mecanismos moleculares subyacentes a cada endofenotipo serológico (3). Sin embargo, la gran mayoría de los estudios transcriptómicos en AR se han desarrollado en poblaciones europeas y asiáticas, con una representación reducida de pacientes colombianos y latinoamericanos (4). Esta subrepresentación limita la comprensión de cómo la diversidad genética y ancestral de la población colombiana puede influir en los perfiles de expresión génica y en la fisiopatología de la enfermedad.

Bajo este panorama, se hace necesario caracterizar el perfil transcriptómico de pacientes colombianos con AR a partir de células mononucleares de sangre periférica (PBMC), con el fin de explorar su relación con los endofenotipos serológicos definidos por la combinación de FR, ACPA y ANA. Este estudio permite explorar la relación entre los perfiles de expresión génica y la actividad clínica de la enfermedad, generando conocimiento sobre los mecanismos inmunológicos diferenciales y las rutas moleculares que explican la heterogeneidad clínica observada. En este contexto, surge la siguiente pregunta de investigación: ¿Cuáles son los patrones de expresión génica en células mononucleares de sangre periférica de pacientes colombianos con artritis reumatoide y cómo varían entre los endofenotipos serológicos definidos por FR, ACPA y ANA?

2. Justificación

El estudio de la expresión génica en PBMC de pacientes con AR mediante RNA-seq ofrece una oportunidad única para comprender las bases moleculares de la enfermedad desde una perspectiva sistémica y no invasiva. Las PBMC reflejan alteraciones inmunológicas relevantes en la AR, incluyendo el desequilibrio Th17/Treg, la activación monocítica y la respuesta mediada por interferones tipo I, procesos estrechamente vinculados con la actividad inflamatoria (5,6).

A nivel nacional, la evidencia transcriptómica en cohortes colombianas con AR es escasa. Considerando la diversidad genética de la población colombiana, producto del mestizaje entre ancestría europea, indígena y africana, su análisis transcriptómico puede aportar

evidencia sobre posibles firmas de expresión génica particulares, así como sobre variaciones en rutas moleculares previamente descritas. Además, la integración de datos serológicos, clínicos y transcriptómicos puede aportar elementos para la construcción de modelos más sólidos de estratificación clínica y molecular de los pacientes, contribuyendo a la identificación de estrategias de diagnóstico y seguimiento más precisas.

Desde una perspectiva científica, este trabajo contribuye al conocimiento de los mecanismos moleculares e inmunológicos implicados en la patogénesis de la AR. Desde un enfoque aplicado, aporta elementos que pueden orientar futuras investigaciones dirigidas a la identificación de biomarcadores pronósticos asociados con la actividad de la enfermedad o la respuesta terapéutica, lo que podría favorecer el desarrollo de estrategias de diagnóstico y seguimiento más precisas en el contexto de la medicina personalizada. Bajo este contexto, se plantea la hipótesis de que los pacientes colombianos con artritis reumatoide presentan patrones diferenciales de expresión génica en PBMC asociados con los endofenotipos serológicos definidos por FR, ACPA y ANA.

3. Objetivos.

Objetivo general.

Caracterizar los perfiles de expresión génica de células mononucleares de sangre periférica en una cohorte de pacientes colombianos con artritis reumatoide y establecer su relación con los endofenotipos serológicos.

Objetivos específicos.

1. Identificar los patrones de expresión génica en PBMC de los diferentes endofenotipos serológicos mediante secuenciación de próxima generación (NGS)
2. Elaborar un análisis bioinformático comparativo entre los niveles de expresión génica en células mononucleares de sangre periférica en los diferentes endofenotipos.
3. Realizar un análisis funcional de los genes diferencialmente expresados para identificar rutas biológicas asociadas con los diferentes endofenotipos serológicos y su relación con indicadores clínicos de actividad de la enfermedad.

4. Marco teórico.

4.1 Panorama general de la artritis reumatoide.

La artritis reumatoide (AR) es una enfermedad crónica, inflamatoria y autoinmune caracterizada por sinovitis persistente que conduce a daño articular progresivo e irreversible. Constituye una de las principales causas de dolor articular y discapacidad en personas adultas a nivel mundial (7). En ausencia de un tratamiento adecuado y oportuno, la artritis reumatoide puede comprometer órganos y sistemas extraarticulares. Entre las complicaciones más frecuentes se encuentran las enfermedades cardiovasculares, las cuales en determinados casos pueden resultar fatales (7). La aparición de estas complicaciones se encuentra estrechamente asociada con la actividad inflamatoria de la enfermedad (8).

La AR es una enfermedad sistémica compleja cuya patogénesis involucra múltiples interacciones entre distintas poblaciones celulares localizadas tanto en el compartimento sinovial de la articulación afectada como en la sangre periférica, lo que dificulta su comprensión integral (9,10). Entre las células implicadas se incluyen los sinoviocitos semejantes a fibroblastos (FLS, por sus siglas en inglés), células del sistema inmune innato y adaptativo, así como células involucradas en la osteogénesis y la remodelación ósea. Durante la progresión de la enfermedad, estas poblaciones celulares experimentan variaciones en número, fenotipo y comportamiento en respuesta a los cambios del microambiente articular. Estas alteraciones influyen en la secreción de citocinas, en la regulación de las redes de señalización intracelular y en el mantenimiento de la homeostasis tisular, favoreciendo el inicio, la progresión y la persistencia de la AR (11).

La AR constituye un problema relevante de salud pública a nivel mundial. Su prevalencia se estima en alrededor de 240 casos por cada 100.000 habitantes, cifra que podría aumentar como consecuencia del envejecimiento progresivo de la población global (12). En Colombia, la incidencia y prevalencia reportadas entre 2017 y 2022 fueron de 34,7–51,4 por cada 100.000 habitantes y de 0,282–0,382 por cada 100 habitantes, respectivamente. Solo en 2022, las consultas relacionadas con AR representaron un gasto cercano a los 23 millones de dólares para el sistema de salud colombiano (13).

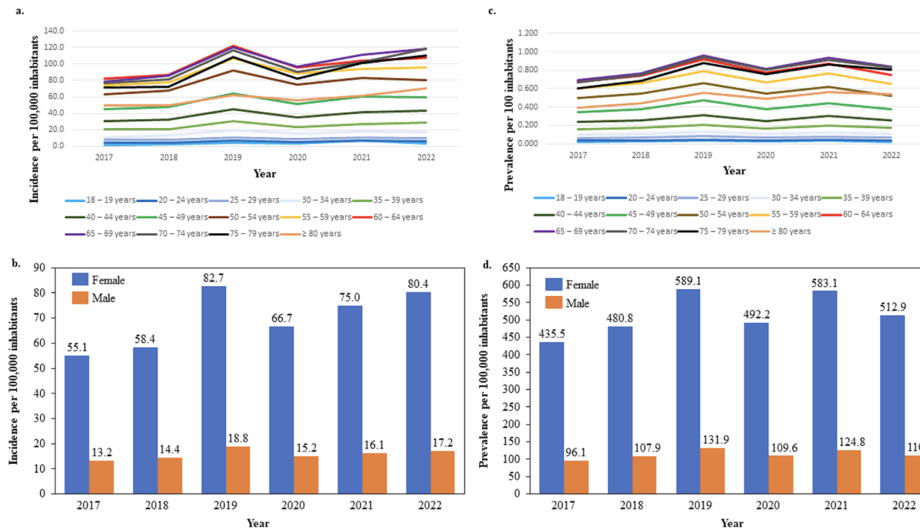


Figura 1. Tasa de incidencia y prevalencia de la artritis reumatoide según la edad y el sexo. Los paneles (a) y (c) muestran la distribución por grupos etarios, evidenciando un incremento progresivo de la enfermedad en edades mayores. Los paneles (b) y (d) muestran la distribución por sexo, destacando una mayor incidencia y prevalencia en mujeres en todos los años evaluados. Tomada de Fernández-Ávila et al (13).

4.2 Bases celulares y moleculares de artritis reumatoide.

4.2.1 Sinoviocitos semejantes a fibroblastos (FLS).

Los sinoviocitos semejantes a fibroblastos (FLS, por sus siglas en inglés) son células derivadas de precursores mesenquimales altamente especializadas que conforman la capa de revestimiento sinovial en las articulaciones diartrodiales. En condiciones fisiológicas, los FLS secretan componentes de la matriz extracelular (ECM) y del líquido sinovial, lo que asegura la lubricación y nutrición del cartílago articular y contribuye al mantenimiento de la homeostasis articular (14,15).

En el contexto de la AR, los FLS adquieren un fenotipo patológico caracterizado por una reducción en la apoptosis, así como por una proliferación, migración e invasión desreguladas. Además, incrementan la secreción de mediadores proinflamatorios y metaloproteinasas de matriz (MMP), favoreciendo la inflamación crónica y la degradación del cartílago (14,16,17). Estudios previos han reportado una sobreexpresión constitutiva del supresor tumoral p53, así como mutaciones en el gen TP53 en FLS aislados de articulaciones afectadas por AR. Este comportamiento aberrante parece ser un fenómeno localizado, ya que no se observa de manera generalizada en FLS provenientes de articulaciones sanas del mismo individuo (18).

Los FLS en la AR no se comportan de manera pasiva. Estas células son capaces de adoptar un estado funcional persistente, que lo lleva a adoptar un perfil patológico que combina proliferación descontrolada, resistencia a apoptosis, secreción de MMPs y producción de citocinas y quimiocinas que sostienen el estado invasivo en las articulaciones. Mecanismos epigenéticos y cambios transcripcionales en estas células fijan ese fenotipo agresivo, lo que explica su capacidad para invadir cartílago y hueso. En consecuencia, en los últimos años se

ha considerado su posible papel como una diana terapéutica emergente. Análisis transcriptómicos han permitido identificar perfiles específicos de FLS asociados con mayor capacidad invasiva y resistencia a las terapias (18).

4.2.2 Inmunidad adaptativa.

El cambio funcional de las células inmunitarias desempeña un papel central en la patogénesis de la AR. En esta patología, subpoblaciones específicas de células inmunes no solo inician, sino que también mantienen la inflamación autoinmune crónica. Entre las más relevantes se encuentran los linfocitos T CD4+, que coordinan respuestas proinflamatorias, y las células de la línea monocítica, que contribuyen a la activación de macrófagos y a la producción sostenida de citocinas proinflamatorias (19).

Las células T CD4+ tienen la capacidad de infiltrarse en el microambiente sinovial inflamado y desempeñar un papel clave en la progresión de la AR. Este grupo celular incluye subpoblaciones con funciones diferenciadas: células T colaboradoras (Th), células T reguladoras (Tregs) y células T foliculares auxiliares (Tfh). Cada una se distingue por sus perfiles de secreción de citocinas y, en consecuencia, por su contribución a distintos procesos patológicos (20). Un punto crítico en la inmunopatogénesis de la AR es el desequilibrio entre las células Th17 patógenas, productoras de citocinas proinflamatorias como IL-17A, IL-17F e IL-22 (9), y las Tregs FoxP3+, caracterizadas por su baja capacidad proliferativa y por la secreción de citocinas inmunosupresoras como IL-4, IL-10 y TGF- β . La pérdida progresiva de la función supresora de las Tregs favorece la expansión de respuestas inflamatorias descontroladas, lo que contribuye a la perpetuación del daño articular en la AR (21).

Los linfocitos B desempeñan un papel central en la patogénesis de la AR, que va más allá de la producción de autoanticuerpos. Estas células también participan en la presentación de antígenos y en la liberación de citocinas que modulan el microambiente sinovial inflamatorio. La estimulación de los linfocitos B a través de los receptores tipo Toll (TLR por sus siglas en inglés) favorece su diferenciación hacia células plasmáticas capaces de sintetizar autoanticuerpos (7). Entre los autoanticuerpos más relevantes en la AR se encuentran el factor reumatoide (FR), los anticuerpos anti-péptidos cíclicos citrulinados (anti-CCP o ACPA) y, más recientemente, se ha reportado la utilidad de incluir los anticuerpos antinucleares (ANAs) en la caracterización serológica de estos pacientes (22).

La respuesta adaptativa en la AR se dirige principalmente contra péptidos citrulinados y otros neoantígenos; la activación de Tfh y la estimulación a linfocitos B facilita la maduración de clones productores de ACPA. Diversos estudios transcriptómicos muestran que perfiles de expresión en subpoblaciones de linfocitos T y B se correlacionan con presencia y títulos de ACPA, sosteniendo la idea de endofenotipos serológicos con huellas transcripcionales distinguibles (23).

Aunque ningún autoanticuerpo posee valor predictivo absoluto de manera individual, la combinación de varios marcadores permite mejorar la estratificación de los pacientes y orientar un tratamiento más efectivo. Este enfoque es coherente con los principios de la

medicina personalizada, ya que posibilita la instauración temprana de intervenciones terapéuticas adaptadas a las características inmunológicas de cada paciente (11).

4.2.3 Inmunidad innata.

La inmunidad innata en la AR tiene como principales poblaciones celulares a los macrófagos y los monocitos, ambos considerados efectores claves de la sinovitis. En condiciones fisiológicas, los macrófagos residentes en el tejido sinovial cumplen funciones esenciales como la eliminación de desechos, la remodelación tisular y el mantenimiento de la tolerancia inmunológica. Sin embargo, en un entorno articular inflamado adquieren un fenotipo activado, que se caracteriza por la secreción de citocinas proinflamatorias que ayudan al sostenimiento de la inflamación crónica en el sinovio (24). Por su parte, los monocitos circulantes migran desde la médula ósea hacia las articulaciones comprometidas, donde se diferencian en macrófagos sinoviales, reforzando así la retroalimentación inflamatoria que perpetúa la patogénesis de la AR (25,26).

Las células efectoras de la inmunidad innata, como las células dendríticas (DC) y los neutrófilos, desempeñan un papel crítico en la patogénesis de la AR. En el desarrollo de la AR es característico el aumento de neutrófilos en el líquido sinovial; este aumento contribuye a la sinovitis sostenida, no solo por la síntesis de proteasas, prostaglandinas y especies reactivas de oxígeno (ROS) (27), sino también por la formación de trampas extracelulares de neutrófilos (NETs), las cuales liberan autoantígenos citrulinados. La NETosis es un mecanismo por el cual neutrófilos liberan cromatina recubierta de proteínas (histonas, vimentina, fibrinógeno) que frecuentemente son citrulinados; esto genera un depósito extracelular de autoantígenos que puede estimular respuestas adaptativas. PAD4 es crítico para la citrulinación de histonas y la descondensación de la cromatina previa a la liberación de NETs; la actividad de PAD4 se asocia con formación de NETs y con marcadores de actividad clínica en AR (28,29).

La NETosis deja una huella transcripcional detectable en sangre: aumento de PADI4 y genes de activación neutrofílica y respuesta oxidativa que pueden aparecer en PBMC o en perfiles celulares subyacentes detectables por RNA-seq, lo que permite correlacionar la actividad de NETosis con estados serológicos y de enfermedad (Figura 2) (30,31).

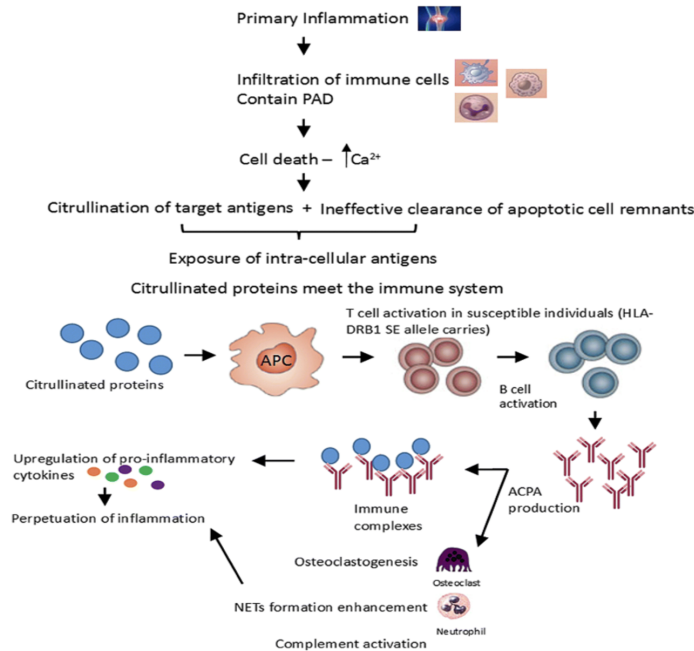


Figura 2. Papel hipotético de la citrulinación en la AR. Infiltración de células inmunitarias que contienen PAD en el tejido inflamado. La activación de PAD, debido a la alta concentración intracelular de calcio durante la muerte celular, promueve la citrulinación de los antígenos diana. La eliminación ineficaz de los restos celulares apoptóticos promueve una accesibilidad prolongada de los antígenos intracelulares que entran en contacto con el sistema inmunitario. Las células presentadoras de antígenos (CPA) procesarán y presentarán péptidos citrulinados a las células T. En individuos susceptibles, las células T activadas activarán las células B para generar autoanticuerpos. Esto dará lugar, en última instancia, a la formación de inmunocomplejos, seguida de la regulación positiva de citocinas proinflamatorias. Tomada de Brito Rocha, S et al 2019 (32).

Las DC pueden contribuir a la tolerancia inmune mediante la generación y mantenimiento de células T reguladoras (Treg) y al promover la anergia o supresión funcional de células T (33). Sin embargo, en el microambiente sinovial inflamatorio de la AR, las DC maduras sobre expresan moléculas coestimuladoras como CD80, CD86 y CD40, presentan antígenos a través del MHC clase I y II, y favorecen la diferenciación y expansión de células T autorreactivas, especialmente hacia fenotipos Th1 y Th17. Esta dualidad (tolerancia vs autoreactividad) está mediada por citocinas proinflamatorias (TNF- α , IL-1, IL-6, IL-17), daño oxidativo, inmunocomplejos y autoanticuerpos (34).

4.2.4 Células del tejido óseo y cartílago.

La AR es una enfermedad compleja que involucra un amplio repertorio de células en múltiples procesos fisiológicos. Los osteoclastos, osteoblastos y condrocitos son células del tejido óseo y condral implicadas en la progresión de la AR.

Los osteoclastos son células grandes, multinucleares que se diferencian a partir de progenitores del linaje monocito-macrófago (35). Estas células degradan la matriz ósea mediante enzimas como la fosfatasa ácida resistente al tartrato (TRAcP), catepsina K y la metaloproteinasas de matriz 9 (MMP-9). Durante la progresión de la AR, los osteoclastos migran y se establecen en la interfaz situada entre el sinovio inflamado y la superficie ósea periarticular. En este microambiente inflamatorio predominan diversas citocinas proinflamatorias, entre ellas el factor de necrosis tumoral alfa (TNF- α) y las interleucinas IL-

1, IL-6 e IL-17, así como la presencia de autoanticuerpos (35,36). En conjunto, estos mediadores inflamatorios, junto con la activación de la vía de señalización del ligando del receptor activador del factor nuclear kappa B (RANKL), favorecen la diferenciación y activación de los osteoclastos. Como resultado, se intensifica la actividad osteoclástica y se acelera la resorción ósea, contribuyendo al daño estructural característico de la enfermedad (37).

Por otro lado, los osteoblastos se originan a partir de células madre mesenquimales (MSC) presentes en el estroma de la médula ósea. Estas células especializadas son indispensables para la síntesis de matriz ósea y su posterior mineralización. Sin embargo, en el microambiente inflamatorio sinovial de la AR, la diferenciación y función osteoblástica se encuentran fuertemente inhibidas por citocinas proinflamatorias como TNF- α e IL-1, así como por la sobreexpresión de antagonistas de la vía Wnt. Este desequilibrio contribuye a la pérdida ósea artrítica grave y progresiva en los sitios de erosión periarticular (35,38).

La destrucción irreversible del cartílago articular es otra característica central de la AR (9). El cartílago está formado por condrocitos, que son las células especializadas del tejido, y por la matriz extracelular que estas células producen y mantienen. Dicha matriz está compuesta principalmente por colágeno tipo II y proteoglicanos, moléculas que proporcionan al cartílago su capacidad para resistir fuerzas mecánicas y soportar la compresión, lo que resulta esencial para su función estructural y protectora en las articulaciones (39). Este tejido avascular depende exclusivamente del líquido sinovial para su nutrición y mantenimiento. En la AR, la invasión del *pannus* sinovial y la alteración del líquido sinovial generan un ambiente proinflamatorio con altas concentraciones de TNF- α e IL-1. Estas señales inflamatorias inducen a los condrocitos a incrementar la expresión de metaloproteinasas de matriz (MMPs) y de enzimas de la familia ADAMTS. Las MMPs degradan principalmente el colágeno tipo II, mientras que las ADAMTS actúan sobre el agregano, el principal proteoglicano del cartílago. La actividad conjunta de estas enzimas favorece la degradación de la matriz extracelular y acelera la pérdida progresiva del cartílago articular (40,41). La Figura 3 presenta un esquema que ilustra la interacción entre estas células en un microambiente proinflamatorio, característico de la AR.

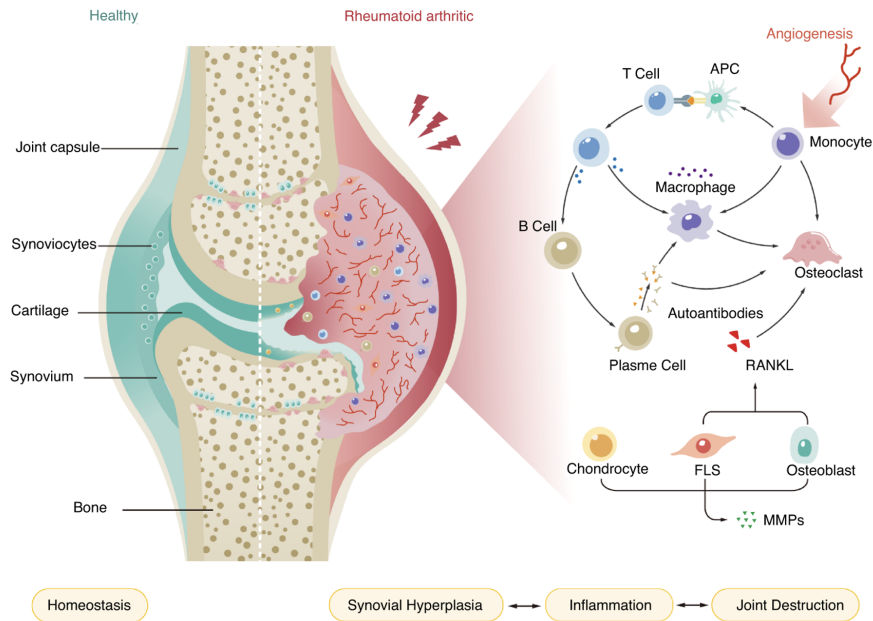


Figura 3. Interacción intercelular en la patología de la AR. A la izquierda se compara una articulación sana con una articulación afectada por artritis reumatoide, evidenciando hiperplasia sinovial, infiltración celular, angiogénesis y degradación progresiva del cartilago y el hueso. A la derecha se ilustran los principales mecanismos inmunológicos implicados, incluyendo la activación de células T y B, la diferenciación de células plasmáticas productoras de autoanticuerpos, la activación de macrófagos y monocitos, y la inducción de osteoclastos mediada por RANKL. Estos procesos conducen a inflamación crónica, producción de metaloproteinasas (MMPs) y destrucción articular progresiva. Tomada de Peng, X., et al 2023 (11).

Como resumen, entendemos que la AR es una patología caracterizada por la interacción entre múltiples tipos celulares que favorecen su aparición y progresión. El proceso se inicia con la pérdida de tolerancia inmune, la infiltración de linfocitos T y B desde la sangre periférica hacia el tejido sinovial y la expansión descontrolada de fibroblastos sinoviales residentes (AR-FLS), que presentan alteraciones en el ciclo de muerte celular (43).

La combinación de estos factores conduce a una sobreproducción de mediadores proinflamatorios, responsables del establecimiento y mantenimiento de la sinovitis crónica, condición distintiva de la enfermedad (10,42). Los AR-FLS activados secretan factores proangiogénicos, promoviendo angiogénesis y remodelación vascular patológica, lo que facilita la infiltración transendotelial de linfocitos T y B hacia la articulación (43).

La interacción entre linfocitos T y células presentadoras de antígeno (CPA) promueve la expansión de las subpoblaciones Th1 y Th17. Estas células producen citocinas proinflamatorias, entre ellas IL-17, IFN- γ y TNF- α , que favorecen la activación de linfocitos B y macrófagos. Posteriormente, los linfocitos B activados se diferencian en células plasmáticas, responsables de la producción de autoanticuerpos como el factor reumatoide (FR) y los anticuerpos contra proteínas citrulinadas (ACPA), los cuales pueden formar complejos inmunitarios. Estos complejos estimulan a los macrófagos para liberar citocinas proinflamatorias, principalmente TNF- α , IL-1 e IL-6, que mantienen el proceso inflamatorio y favorecen la resorción ósea mediada por osteoclastos (44,45).

En paralelo los AR-FLS experimentan expansión clonal e impulsan la transformación hiperplásica de la capa de revestimiento sinovial hacia un *pannus* invasivo, el cual, nutrido

por la angiogénesis, invade cartílago y hueso en la interfaz osteocondral, provocando la destrucción estructural de la articulación. Los fibroblastos sinoviales, condrocitos y osteoblastos tienen la capacidad de expresar RANKL, molécula clave que promueve la diferenciación de progenitores monocito-macrófago hacia osteoclastos. Este eje RANK/RANKL/OPG es crítico en la erosión ósea, ya que la disminución relativa de osteoprotegerina (OPG) permite un predominio de RANKL y una osteoclastogénesis exacerbada. Además, en el microambiente inflamatorio, FLS, condrocitos y osteoblastos secretan metaloproteinasas de matriz (MMPs) que degradan colágeno tipo II, contribuyendo a la destrucción del cartílago articular (16,46).

4.3. Mecanismos moleculares de pérdida de tolerancia

4.3.1 Citrulinación y generación de neoantígenos.

La citrulinación es una modificación postraduccional en la que las enzimas PAD2 y PAD4 convierten residuos de arginina en citrulina, eliminando una carga positiva (**Figura 4**). Este cambio modifica la estructura proteica, su estabilidad y el procesamiento por proteasas lisosomales, lo que expone regiones ocultas del polipéptido y genera epítomos crípticos. En la AR, proteínas propias como fibrinógeno, vimentina y hnRNP A2/B1 pueden sufrir citrulinación. Este cambio modifica el conjunto de péptidos derivados de estas proteínas que son presentados por las moléculas del MHC-II al sistema inmunitario. Como resultado, algunos epítomos dominantes dejan de presentarse, mientras que aparecen nuevos epítomos o neoantígenos que antes no eran reconocidos. Algunos de estos neoantígenos pueden originarse incluso en regiones de la proteína que no contienen citrulina, debido a cambios en el procesamiento antigénico. Este fenómeno contribuye a la ruptura de la tolerancia inmunológica y a la activación de respuestas autoinmunes características de la AR. Así, la respuesta inmune no depende solo de péptidos modificados, sino también de la exposición de secuencias nativas previamente inaccesibles. La activación excesiva de PAD4 en neutrófilos, monocitos y células sinoviales bajo condiciones inflamatorias facilita este proceso, promoviendo ruptura de la tolerancia y amplificación de la respuesta autoinmune (47).

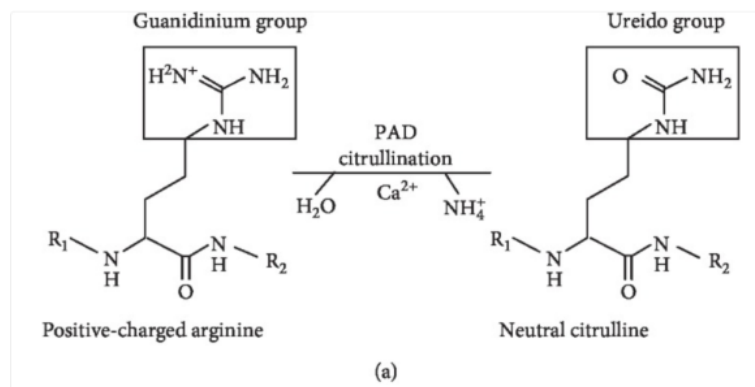


Figura 4. La reacción química de la citrulinación. Se muestra la pérdida de carga positiva y la generación de citrulina, proceso clave en la formación de autoantígenos en artritis reumatoide. Imagen tomada de: Alghamdi M, et al _2019 (48).

La relevancia clínica de la citrulinación radica en que los neoantígenos generados por este proceso son presentados con mayor eficiencia por alelos HLA-DRB1 que contienen el denominado “shared epitope”, lo que favorece la activación de linfocitos T autorreactivos que proporcionan señales de coestimulación a los linfocitos B, promoviendo la producción de anticuerpos anti-proteínas citrulinadas (ACPA). A nivel molecular, estudios estructurales y funcionales han demostrado que la sustitución de arginina por citrulina en posiciones críticas del péptido (P4/P6) incrementa su estabilidad dentro del surco de unión de HLA-DRB1, potenciando la activación de células T autorreactivas. Este eje HLA–citrulina sustenta la hipótesis de que determinados endofenotipos serológicos de la artritis reumatoide presentan correlaciones antigénicas y transcripcionales específicas (49,50).

4.3.2 NETosis

La NETosis es un mecanismo especializado de activación de los neutrófilos que se caracteriza por la descondensación de la cromatina y posteriormente la liberación de trampas extracelulares de neutrófilos (NETs). Las NETs son estructuras compuestas por ADN nuclear, histonas citrulinadas y proteínas granulares con alta actividad inmunomoduladora (51). En la AR, la activación de la NETosis ocurre en respuesta a la presencia de estímulos inflamatorios en el microambiente sinovial, incluyendo citocinas proinflamatorias, inmunocomplejos y señales derivadas del daño tisular, lo que favorece la activación de enzimas como PAD4 y la remodelación de la cromatina necesaria para la liberación de NETs (52). Una vez liberadas, las trampas extracelulares de neutrófilos (NETs) funcionan como estructuras proinflamatorias que intensifican la respuesta inmunitaria local. Estas estructuras activan fibroblastos sinoviales y macrófagos mediante la estimulación de receptores de reconocimiento de patrones (PRR). Como resultado, se induce una expresión sostenida de quimiocinas quimiotácticas, entre ellas CXCL8, CCL2 y CXCL1, las cuales favorecen el reclutamiento continuo de neutrófilos y monocitos hacia el sinovio inflamado, perpetuando así el proceso inflamatorio (53). Este aumento de neutrófilos infiltrando el sinovio refuerza la liberación adicional de NETs, estableciendo un sistema de retroalimentación positiva que mantiene la activación inflamatoria sinovial (54). La NETosis es un mecanismo central de organización y perpetuación de la inflamación sinovial en la AR, al sostener rutas de reclutamiento celular, activación persistente de células residentes sinoviales y cronificación del daño articular (Figura 5) (52,53).

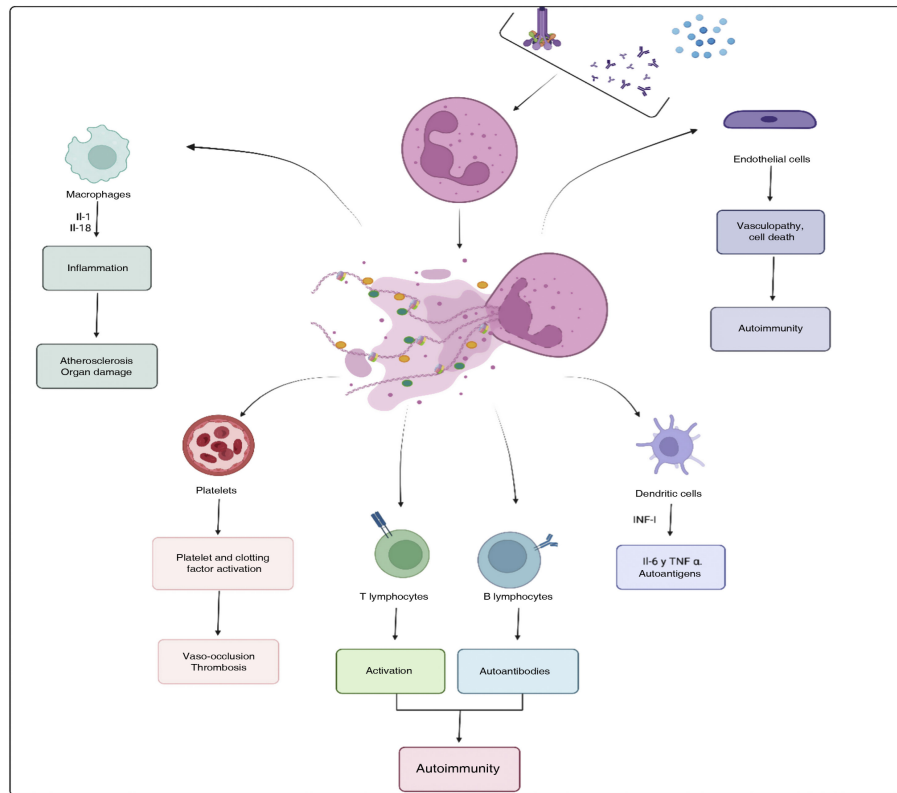


Figura 5. Relación entre los NET y la autoinmunidad. Las NET contienen proteínas inmunoestimuladoras capaces de activar otras células inmunitarias (macrófagos, células dendríticas, linfocitos); además, contienen autoantígenos que pueden desencadenar respuestas autoinmunes mediadas por autoanticuerpos. Por otra parte, las proteínas unidas a las NET pueden dañar las células endoteliales y favorecer la trombosis. Los autoanticuerpos, los complejos inmunes, los autoantígenos, los factores activadores del complemento y las citocinas pueden activar a los neutrófilos e inducir la NETosis. Imagen tomada de: Frade-Sosa B et al _2023 (54).

4.3.3 Endofenotipos serológicos en artritis reumatoide.

Actualmente, el diagnóstico de la AR se basa en los criterios establecidos conjuntamente por el American College of Rheumatology (ACR) y la European League Against Rheumatism (EULAR). Estos criterios consideran cuatro componentes principales: el grado de afectación articular, los marcadores serológicos —especialmente el factor reumatoide (FR) y los anticuerpos contra proteínas citrulinadas (ACPA)—, los reactantes de fase aguda, y la duración de los síntomas clínicos (43,44). Sin embargo, en los últimos años se ha postulado la relevancia de los anticuerpos antinucleares (ANA) como un posible criterio diagnóstico complementario en la AR.

Los ANA representan una amplia gama de autoanticuerpos dirigidos contra diversos componentes nucleares y citoplasmáticos de la célula (45). El reporte de ANA incluye no solo el título o intensidad de fluorescencia, sino también el patrón característico que producen, el cual puede involucrar el núcleo, el citoplasma o estructuras asociadas a células en mitosis. Para estandarizar la nomenclatura de estos patrones, la International Consensus on ANA Patterns (ICAP) propuso sustituir el término ANA por anticuerpos anticelulares (AC), estableciendo 29 patrones distintos de tinción, clasificados como AC1–AC29 (46). Cada patrón se asocia con la presencia de uno o varios autoanticuerpos específicos (47).

En este contexto, Liu et al. demostraron que un título elevado de ANA podría estar asociado con un mayor riesgo de desarrollar AR, en particular con tres patrones específicos: nuclear homogéneo, nuclear moteado y citoplasmático moteado, siendo el nuclear homogéneo el de mayor asociación (**Figura 6**). Estos hallazgos sugieren que los ANA podrían constituir un marcador serológico adicional de riesgo en la AR y abrir nuevas perspectivas para su inclusión en algoritmos diagnósticos (47).

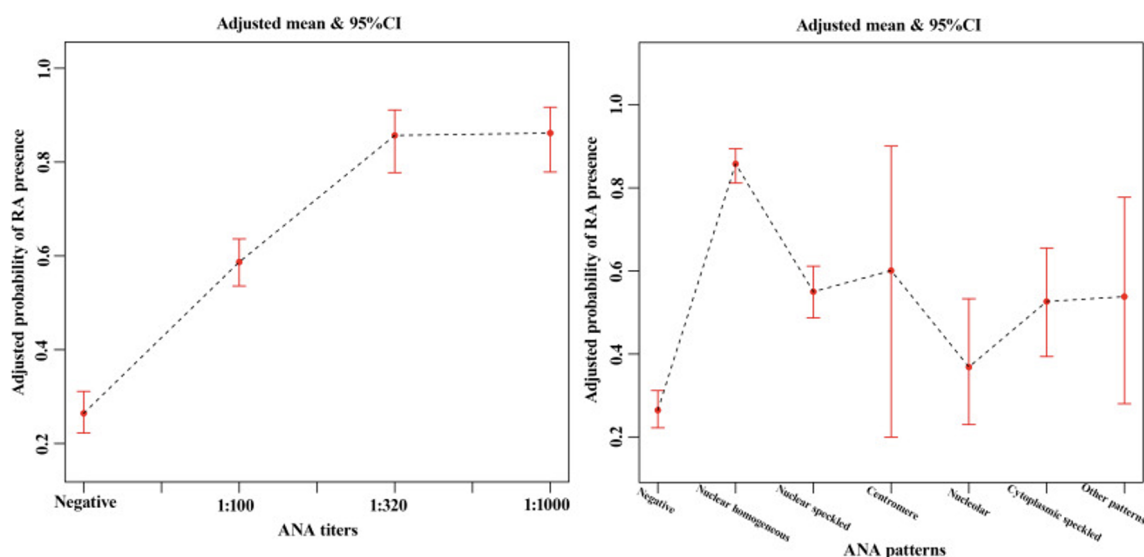


Figura 6. Relación entre los títulos de ANAs y la AR. El ajuste de la curva suave mostró la asociación entre los títulos de ANA (A), los patrones de ANA (B) y la artritis reumatoide tras ajustar por sexo y edad en la cohorte emparejada. Imagen tomada de: Liu F et al _2024 (55)

4.4 Transcriptómica en artritis reumatoide.

4.4.1 ARN mensajero (ARNm)

El ARN mensajero (ARNm) es el producto central de la expresión génica y su regulación ocurre a lo largo de un proceso altamente controlado que incluye transcripción, maduración, exportación nuclear, traducción y degradación. En eucariotas, el ARNm se genera inicialmente como pre-ARNm, este proceso es llevado a cabo por la ARN polimerasa II, posteriormente sufre un procesamiento co-transcripcional que determina su identidad funcional. Este procesamiento comprende varias modificaciones del ARN mensajero, entre ellas la adición del cap en el extremo 5', el splicing o eliminación de intrones y la formación del extremo 3' mediante corte y poliadenilación. Estos eventos no solo cumplen una función estructural en el transcrito, sino que también regulan su estabilidad, su localización dentro de la célula y la eficiencia con la que será traducido en proteína (56).

Un mecanismo particularmente relevante es la poliadenilación alternativa (APA), que genera distintas versiones de un mismo ARNm con regiones 3'UTR de diferente longitud. Estas variaciones influyen en la vida media del transcrito, su localización celular y su interacción con microARNs y proteínas reguladoras. Se ha demostrado que la desregulación de la APA es frecuente en enfermedades inflamatorias crónicas, donde favorece la expresión prolongada de genes proinflamatorios (56). Tras su maduración, el ARNm es exportado del núcleo al

citoplasma mediante mecanismos altamente regulados. Alteraciones en este proceso afectan la disponibilidad del ARNm para ser traducido y modifican la respuesta celular frente a estímulos inflamatorios (57).

En la AR, la inflamación persistente no depende únicamente de qué genes se activan, sino también de cómo se procesan y cuánto tiempo permanecen activos sus ARNm. Estudios realizados en tejido sinovial han evidenciado una expresión persistente de ARN mensajeros relacionados con citocinas, quimiocinas y distintas vías inmunitarias. Este patrón sugiere que los mecanismos que regulan la estabilidad y degradación del ARNm podrían estar alterados, favoreciendo así el mantenimiento de la respuesta inflamatoria (58). Estos hallazgos pueden indicar que la regulación en el ciclo de vida del ARNm contribuye de forma directa a mantener el estado inflamatorio crónico característico de la AR.

4.4.2 Fundamentos de la transcriptómica.

Existe un creciente interés en el análisis del transcriptoma como herramienta para comprender los mecanismos patogénicos de la artritis AR y para identificar nuevas estrategias terapéuticas. El perfil de expresión génica permite cuantificar simultáneamente los niveles de ARNm de miles de genes, proporcionando una visión integral del estado funcional de las células y contribuyendo al desarrollo de enfoques de medicina estratificada (59). Las microarrays de ADN han sido ampliamente utilizadas durante más de dos décadas para el análisis de la expresión génica. Sin embargo, presentan limitaciones importantes, como un rango dinámico restringido, dificultades para detectar señales de baja intensidad y problemas asociados a la hibridación de fondo. Estas limitaciones se han superado en gran medida con la secuenciación de ARN (RNA-seq), una tecnología de secuenciación de próxima generación (NGS) que permite cuantificar la abundancia de transcritos a escala genómica. En comparación con las microarrays, RNA-seq ofrece un rango dinámico más amplio, menor señal de fondo, mayor sensibilidad y una alta reproducibilidad (60).

Las PBMC (peripheral blood mononuclear cells) incluyen linfocitos T y B, células NK, monocitos y células dendríticas. Su accesibilidad y su rol como puente entre la inmunidad periférica y la inflamación sinovial hacen que sean un recurso valioso en investigación traslacional. Diversos estudios han demostrado que las PBMC reflejan procesos inmunológicos clave en la AR, incluyendo el desequilibrio Th17/Treg, la activación monocítica y la sobreexpresión de genes de interferón tipo I. Además, el RNA-seq en PBMC permite obtener biomarcadores menos invasivos que las biopsias sinoviales, lo que favorece su aplicación en cohortes clínicas.

4.4.3 Enfoque innovador.

La diversidad genética y ancestral de la población colombiana representa un escenario particularmente relevante para los estudios moleculares de la AR. La mayoría de los estudios transcriptómicos sobre esta enfermedad se han realizado en cohortes europeas o asiáticas, lo que limita la comprensión de los mecanismos moleculares de la AR en poblaciones latinoamericanas. En este contexto, la población colombiana, caracterizada por un complejo mestizaje entre ancestría indígena, africana y europea, constituye un modelo de gran interés para explorar cómo la variabilidad genética puede influir en los perfiles epigenéticos y

transcriptómicos asociados a la enfermedad. El estudio de estas poblaciones podría aportar información clave para comprender la heterogeneidad molecular de la AR y avanzar hacia estrategias de medicina de precisión en contextos poblacionales diversos. Estudios de asociación genética en AR han demostrado que polimorfismos del HLA y genes reguladores de la respuesta inmune muestran frecuencias diferentes en poblaciones latinoamericanas frente a europeas. Por lo tanto, el análisis de expresión génica en una cohorte colombiana puede revelar patrones específicos no identificables en otras poblaciones (61–63).

De igual manera, la heterogeneidad clínica y serológica de los pacientes en Colombia demanda un abordaje más personalizado. La estratificación de pacientes por perfiles serológicos (FR, ACPA, ANA) no siempre coincide con la progresión clínica o con la respuesta a fármacos modificadores de la enfermedad (FARME) (64). En este sentido, la transcriptómica basada en RNA-seq ofrece una oportunidad para identificar diferencias moleculares entre estos subgrupos, generando información aplicable a la medicina personalizada. Este enfoque resulta particularmente relevante en Latinoamérica, donde las barreras de acceso a terapias biológicas y las diferencias en adherencia terapéutica aumentan el impacto de la enfermedad sobre la calidad de vida y la productividad laboral.

De esta manera, el análisis transcriptómico mediante RNA-seq en una cohorte colombiana representa una oportunidad para profundizar en la caracterización molecular de la AR en el contexto local y, simultáneamente, enriquecer el conocimiento global de la enfermedad mediante la incorporación de datos provenientes de poblaciones ancestralmente diversas y subrepresentadas en los estudios transcriptómicos.

5. Estado del arte

Estudios recientes han probado PBMCs para explorar las alteraciones transcriptomas asociadas con la AR, tanto en comparación con individuos sanos como en relación con la actividad de la enfermedad. Binvignat et al. aplicaron secuenciación de RNA de célula única (scRNA-seq) en PBMCs de 18 pacientes con AR y 18 controles sanos, identificando 18 subpoblaciones celulares distintas. Entre los hallazgos más relevantes se observó la presencia de monocitos activados por interferón (IFITM3+), un aumento de linfocitos CD4+ de memoria efectora en pacientes con actividad moderada a alta, y una disminución de monocitos no clásicos en pacientes en remisión o con baja actividad de la enfermedad. Asimismo, se identificaron genes diferencialmente expresados, incluyendo TNF, JUN, EGR1, IFIT2, MAFB y G0S2, junto con una disminución en la expresión de genes relacionados con la presentación antigénica como HLA-DQB1, HLA-DRB5 y TNFSF13B (23).

De manera complementaria, un estudio que integró datos de scRNA-seq provenientes de PBMCs y líquido sinovial con cohortes transcriptómicas de tipo bulk reveló que poblaciones celulares como las peripheral helper T cells (TPHs), células dendríticas, monocitos CD16+ y linfocitos CD8+ se encuentran estrechamente asociadas con el número de articulaciones afectadas. Estas poblaciones mostraron ser más representativas del estado inflamatorio local que algunos índices clínicos tradicionales de actividad de la enfermedad, lo que sugiere que

determinados perfiles transcriptómicos en PBMC podrían reflejar procesos inflamatorios sinoviales y potencialmente servir como herramientas de monitoreo clínico (65).

En la misma línea, Liu et al. identificaron posibles biomarcadores diagnósticos mediante scRNA-seq en PBMCs, centrándose en subpoblaciones de células T. Entre los genes destacados se encuentran CD8A, CCL5, GZMB, NKG7, PRF1, GZMH, CCR7, GZMK y GZMA, lo que respalda la hipótesis de que distintos subtipos serológicos o clínicos de AR podrían asociarse con firmas transcriptómicas específicas (66).

Estudios funcionales han explorado el potencial de los PBMCs para modelar procesos relacionados con la destrucción ósea característica de la AR. Un estudio reciente de diferenciación osteoclástica inducida a partir de PBMCs mostró que la estimulación con RANKL y M-CSF activa múltiples rutas de señalización, incluyendo señalización de quimiocinas, interacción citocina-receptor, formación de NETs y vías mediadas por TNF. Estos resultados indican que los PBMCs no solo reflejan alteraciones transcriptómicas en su estado basal, sino que también pueden revelar mecanismos moleculares asociados con inflamación y destrucción ósea cuando se evalúan en condiciones funcionales específicas (67).

6. Metodología.

6.1 Tipo de muestras.

Las células mononucleares de sangre periférica (PBMCs) utilizadas en este estudio fueron obtenidas del banco maestro de células perteneciente al macroproyecto llevado a cabo entre el Instituto Distrital de Ciencia, Biotecnología e Innovación en Salud (IDCBIS) y la Universidad Nacional de Colombia (UNAL) llamado “*Caracterización genética de vesículas extracelulares en Artritis Reumatoide: análisis del contenido de ARN no codificante en una cohorte de pacientes colombianos*”. La inclusión de los pacientes fue autorizada mediante consentimiento informado general del macroproyecto, cumpliendo los lineamientos éticos nacionales e internacionales para investigación biomédica. La participación de los pacientes fue aprobado por el Comité Corporativo de Ética en Investigación del Hospital Universitario Fundación Santa Fe de Bogotá.

Las PBMCs fueron previamente aisladas por investigadores de la Unidad de Terapias Avanzadas (UTA) del IDCBIS a partir de muestras de sangre periférica mediante centrifugación por gradiente de densidad con Ficoll-Paque™ PLUS (GE Healthcare, Suecia), siguiendo las recomendaciones estándar del fabricante. Posteriormente, las células fueron resuspendidas en medio crioprotector compuesto por 90 % de suero fetal bovino (FBS) (Gibco™, Estados Unidos) y 10 % Dimetilsulfóxido (DMSO) (Sigma-Aldrich, Estados Unidos); se alicuotaron en crioviales de 1 mL y fueron almacenadas en tanque de nitrógeno líquido (−196 °C) en las instalaciones del IDCBIS hasta su uso.

6.2 Clasificación de las muestras.

Se analizaron PBMCs provenientes de 39 pacientes con diagnóstico confirmado de AR, de acuerdo con los criterios del Colegio Americano de Reumatología (ACR) y Liga Europea

contra el Reumatismo (EULAR). Las muestras de los pacientes contaban con información clínica y serológica completa, incluyendo títulos de factor reumatoide (FR), anticuerpos anti-péptidos cíclicos citrulinados (ACPA), anticuerpos antinucleares (ANA) y puntuación de actividad clínica (CDAI). Las muestras fueron clasificadas en ocho grupos serológicos definidos por la combinación de los tres marcadores inmunológicos (**Tabla 1**):

<i>Endofenotipo</i>	Factor Reumatoide (FR)	Anticuerpos anti-péptidos cíclicos citrulinados (ACPA)	anticuerpos antinucleares (ANA)
1	Negativo	Negativo	Negativo
2	Positivo	Negativo	Negativo
3	Positivo	Positivo	Negativo
4	Positivo	Positivo	Positivo
5	Negativo	Positivo	Positivo
6	Negativo	Negativo	Positivo
7	Positivo	Negativo	Positivo
8	Negativo	Positivo	Negativo

Tabla 1. Clasificación por endofenotipos. Serología de los endofenotipos clasificatorios de los pacientes en estudio. FR: Factor reumatoide. ACPA: anticuerpos anti-péptido citrulinado. ANA: Anticuerpos antinucleares.

6.3 Extracción de ARN.

La extracción de ARN total se realizó a partir de PBMCs criopreservadas en tanque de nitrógeno líquido. Tras la descongelación, las células se centrifugaron y se resuspendieron en medio X-VIVO® 15 Serum-free Hematopoietic Cell Medium (Lonza, Suiza). La concentración celular se estandarizó a 1.5×10^6 células utilizando el contador automático Countess 3 (Invitrogen, Estados Unidos). Tras la estandarización, se sembraron en placas de 24 pozos con medio X-VIVO® 15 Serum-free Hematopoietic Cell Medium (Lonza, Suiza), suplementado con IL-2, Glutamina y Penicilina-Estreptomicina, con un volumen final de 2 mL en cada pozo. Esto se llevó a incubación a 37 °C por 4 horas; esto con el fin de lograr de la activación de las células antes de la extracción de ARN. Para la extracción de ARN se procesaron 400 µL de las células activadas en la plataforma MagCore® Automated Nucleic Acid Extractor Plus II (RBC Bioscience, Taiwán) empleando el TriXact RNA Kit, siguiendo las especificaciones del fabricante. El procedimiento permitió la obtención de ARN total apto para análisis transcriptómicos posteriores.

6.4 Cuantificación y análisis de la integridad del ARN.

La concentración de ARN total se determinó por fluorometría en el Qubit™ 3 Fluorometer (Invitrogen™, Estados Unidos) empleando el kit Qubit™ RNA HS Assay Kit. La integridad del ARN se evaluó mediante el sistema 4150 TapeStation (Agilent, Estados Unidos) con el cartucho RNA ScreenTape. El ARN purificado se almacenó a -20 °C hasta la preparación de las librerías. Los resultados de las extracciones se presentan en los anexos (anexo 1).

6.5. Preparación de las librerías.

Se utilizaron 60 ng de ARN total por muestra. Las librerías se prepararon con el kit Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus (Illumina, Estados Unidos), siguiendo el protocolo estándar del fabricante. El procedimiento incluyó la depleción de ARN ribosomal y mitocondrial, seguida de fragmentación enzimática del ARN, síntesis de ADNc, adenilación del extremo 3', ligación de adaptadores, purificación y amplificación de las librerías. La integridad y el tamaño promedio de los fragmentos generados durante la preparación de la librería se evaluaron mediante el sistema 4150 TapeStation (Agilent, Estados Unidos), mientras que la cuantificación final se realizó utilizando el Qubit™ 3 Fluorometer (Invitrogen™, Estados Unidos). Los fragmentos de la librería final presentaron un tamaño entre 375 y 475 pares de bases (pb), rango recomendado por el fabricante para el inicio del proceso de secuenciación.

6.6. Secuenciación.

El proceso de secuenciación se llevó a cabo en la plataforma NextSeq 2000 (Illumina, Estados Unidos) ubicada en el Laboratorio de Ciencias Omicas y Medicina Personalizada (OmiMed) del IDC BIS. Se secuenciaron 39 muestras utilizando un cartucho P2 de 300 ciclos, con lecturas paired-end (2×150 pb), garantizando una adecuada cobertura y profundidad para análisis transcriptómicos posteriores. Durante el montaje se incluye el control interno recomendado por el fabricante (PhiX) para la verificación de calidad y calibración de señales ópticas. El proceso de demultiplexado se realizó de forma manual mediante la plataforma BaseSpace (Illumina). Los archivos de salida se generaron en formato FASTQ comprimido, y los datos crudos fueron almacenados en el servidor del laboratorio para su posterior análisis bioinformático. El laboratorio contaba con un montaje de validación previo para RNAseq, lo que asegura la correcta calibración y desempeño del sistema antes de la ejecución experimental.

6.7. Análisis bioinformático.

El flujo bioinformático se desarrolló a partir de los archivos FASTQ generados por la plataforma BaseSpace (Illumina). La depuración de las lecturas se realizó con fastp (v0.26.0), incluyendo recorte de adaptadores y filtrado por calidad, cuyos resultados fueron integrados mediante MultiQC (v1.29). Posteriormente, las lecturas filtradas se alinearon contra el genoma humano de referencia GRCh38.p14 (Ensembl Release 115) utilizando STAR (v2.7.11b), considerando exclusivamente genes codificantes de ARNm. La cuantificación de lecturas se realizó mediante featureCounts (v2.1.1), generando una matriz de conteos crudos por gen y muestra. Los datos se analizaron en R (v4.5.1) con el paquete DESeq2 (v1.40.2) para la estimación de cambios en la expresión génica diferencial.

La normalización de los conteos se realizó mediante el método de size factors implementado en DESeq2, ajustando las diferencias en la profundidad de secuenciación entre muestras. Previo al análisis diferencial, se aplicó un filtrado de baja abundancia basado en el conteo total por gen, conservando aquellos con una suma de conteos ≥ 2000 ($\text{rowSums} \geq 2000$), con el fin de reducir el impacto de genes con baja señal transcriptómica y mejorar la estabilidad del análisis.

La expresión diferencial se evaluó utilizando un modelo de distribución binomial negativa, considerando el endofenotipo serológico como único factor de diseño. Los valores de p se ajustaron mediante el método de Benjamini–Hochberg para controlar la tasa de falsos descubrimientos (FDR), estableciendo un umbral de p -ajustado < 0.05 . Adicionalmente, se aplicó un criterio de relevancia biológica basado en la magnitud del cambio de expresión ($|\log_2FC| \geq 1$).

El análisis exploratorio incluyó un análisis de componentes principales (PCA) a partir de los conteos normalizados, con el objetivo de evaluar la variabilidad global del transcriptoma e identificar patrones de agrupamiento entre las muestras. De manera complementaria, se analizaron matrices de correlación y mapas de calor (heatmaps) para evaluar la similitud entre muestras y la distribución de la expresión génica.

Los resultados del análisis diferencial se representaron mediante volcano plot y conteo de DEGs permitiendo visualizar la magnitud del cambio de expresión y su significancia estadística. Asimismo, los genes diferencialmente expresados se agruparon mediante clustering jerárquico, generando mapas de calor que ilustran patrones de regulación entre los endofenotipos.

Finalmente, se realizó un análisis de enriquecimiento funcional utilizando Gene Ontology (GO) y rutas KEGG, con el objetivo de identificar procesos biológicos, funciones moleculares y vías metabólicas sobre-representadas. Los resultados se integraron mediante análisis de redes funcionales para explorar la organización de los procesos biológicos asociados a la heterogeneidad transcriptómica observada.

6.8. Análisis estadístico

Se realizó un análisis exploratorio con el objetivo de evaluar la asociación entre variables clínicas y el índice de actividad de la enfermedad Clinical Disease Activity Index (CDAI) en una cohorte de pacientes con artritis reumatoide (AR). El CDAI se definió como la variable dependiente principal, utilizada como indicador de la severidad de la enfermedad.

Las variables independientes incluyeron características demográficas y serológicas, entre ellas sexo, edad, endofenotipo serológico, presencia de factor reumatoide, anticuerpos anti-proteínas citrulinadas (ACPA), anticuerpos antinucleares (ANA) junto con su patrón inmunofluorescente, y la temporalidad de la enfermedad.

El procesamiento y análisis estadístico se realizaron en el entorno R (versión 4.3.2). Para evaluar las diferencias en la actividad clínica de la enfermedad entre los distintos endofenotipos, definidos como una variable categórica nominal, se empleó la prueba no paramétrica de Kruskal–Wallis.

Esta prueba permite comparar la distribución de una variable continua, como el índice CDAI, entre dos o más grupos independientes sin asumir normalidad en los datos. Su uso resulta adecuado considerando la naturaleza no paramétrica de las variables clínicas analizadas y el tamaño de muestra limitado por grupo. Se utilizaron pruebas bilaterales de hipótesis con un nivel de significancia de $\alpha = 0.05$.

7. Resultados

7.1. Caracterización clínica y serológica de la cohorte de pacientes.

La cohorte incluida en este estudio estuvo conformada por 39 pacientes con diagnóstico clínico y serológico de AR. Las características clínicas y serológicas basales de los pacientes se presentan a continuación (tabla 2).

Característica	Valor
Total de pacientes	39
Mujeres, n (%)	30 (76.9%)
Hombres, n (%)	9 (23.1%)
Edad (años), media $\pm$ DE	49.8 $\pm$ 14.4
Factor Reumatoide (decimal), media $\pm$ DE	4.346 $\pm$ 7.251
anti-CCP (decimal), media $\pm$ DE	16.984 $\pm$ 32.079
ANAs $\geq 1/80$, n (%)	21
ANAs $\geq 1/160$, n (%)	20
ANAs $\geq 1/320$, n (%)	17
Patrón ANAs	Moteado fino: 17 Homogéneo: 3 Granular fino: 1
Temporalidad (ARE / ART / ARMT), n	14 / 12 / 12
CDAI total, media $\pm$ DE	11.4 $\pm$ 11.1
Tabaquismo (Niega / Exfumador / Activo), n	28 / 7 / 1

FR y anti-CCP expresados como cociente (valor/corte del laboratorio).
ANAs se muestran como títulos $\geq 1/80$, $\geq 1/160$ y $\geq 1/320$.

Tabla 2. Características clínicas basales de los pacientes en el estudio.

Se observó un predominio de mujeres, representando el 76,9% del total de los participantes (30/39 participantes), con una edad media de 49.8 ± 14.4 años. Los niveles de FR y ACPA fueron normalizados respecto al punto de corte del laboratorio, expresándose como la razón entre el valor observado y el umbral de positividad. Los valores promedio fueron de $4,346 \pm 7,251$ para FR y $16,984 \pm 32,079$ para ACPA, evidenciando una amplia dispersión, especialmente en ACPA.

Esta variabilidad en los niveles de FR y ACPA indica la presencia de distintos estados de activación inmunológica dentro de la cohorte, lo cual es consistente con la heterogeneidad fisiopatológica de la AR. Estos resultados evidencian que la serología convencional podría no ser suficiente para caracterizar completamente a los pacientes, lo que respalda la necesidad de enfoques de estratificación más robustos.

7.2. Asociación entre los marcadores serológicos y la actividad clínica de la enfermedad.

Para evaluar la relación entre los marcadores serológicos y el CDAI en esta cohorte, se analizó la correlación entre los niveles de FR, ACPA y ANA con el CDAI mediante el coeficiente de Spearman (Figura 7).

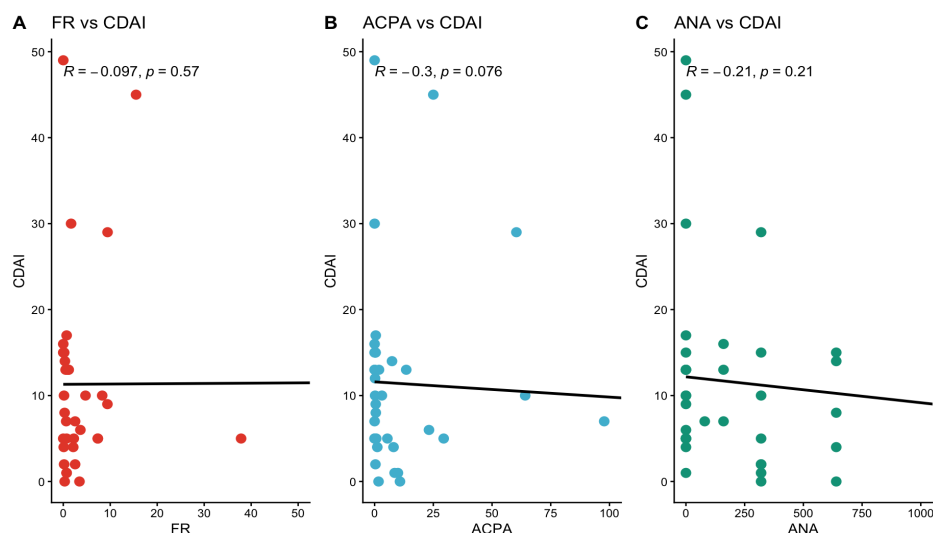


Figura 7. Correlación entre los marcadores serológicos FR, ACPA y ANA con la actividad clínica de la enfermedad medida por CDAI. Se evaluó la asociación entre los niveles de FR, ACPA y ANAs con el CDAI mediante el coeficiente de correlación de Spearman. Cada punto representa un paciente individual. La línea negra indica la tendencia general de los datos. Los valores de R y p correspondientes a cada análisis se muestran en cada panel.

En los tres paneles se observa un patrón consistente caracterizado por alta dispersión y ausencia de tendencias definidas. En el caso de FR, la mayoría de los pacientes presentan valores bajos con una amplia distribución del CDAI, sin evidencia de relación proporcional ($R = -0.097$; $p = 0.57$). De manera similar, ACPA muestra mayor variabilidad en sus niveles, pero sin una asociación consistente con la actividad clínica ($R = -0.30$; $p = 0.076$). Para ANA, se mantiene este comportamiento, donde valores comparables del marcador se asocian con rangos amplios de CDAI ($R = -0.21$; $p = 0.21$).

Estos resultados indican que ninguno de los marcadores serológicos evaluados se asocia de manera significativa con la actividad clínica, evidenciando que la variabilidad del CDAI no es explicada por FR, ACPA o ANA. Por tanto, esto sugiere que, aunque la serología convencional es útil para la clasificación inmunológica, su capacidad para reflejar la actividad clínica y la heterogeneidad de la enfermedad es limitada, por lo que se requieren nuevas estrategias de estratificación orientadas a la identificación de endofenotipos clínico-biológicos.

7.3. Actividad clínica de la enfermedad según endofenotipo.

Considerando la limitada capacidad de la clasificación serológica para explicar la variabilidad clínica observada, se planteó evaluar si una estratificación más basada en endofenotipos permitía capturar de manera más precisa las diferencias inmunológicas dentro de la cohorte. Para ello, se analizó la distribución de los tres marcadores serológicos definidos para la clasificación de esta cohorte entre los endofenotipos definidos.

La distribución de los niveles de FR entre los endofenotipos evidenció una separación consistente entre grupos seropositivos y seronegativos, con valores elevados en los endofenotipos 2, 3, 4 y 7, y niveles cercanos a cero en los endofenotipos 1, 5, 6 y 8 (Figura 8). Se identificaron diferencias significativas entre los grupos (Kruskal-Wallis: $p = 0.000132$), aunque estas se concentraron principalmente en comparaciones específicas,

particularmente entre el endofenotipo 1 y los endofenotipos seropositivos para este marcador serológico, endofenotipo 2, 3, 4 y 7 ($p\text{-aj} \approx 0.015$). A pesar de esta diferenciación, la variabilidad intragrupo fue considerable, lo que limita la capacidad del FR para reflejar niveles homogéneos dentro de cada endofenotipo.

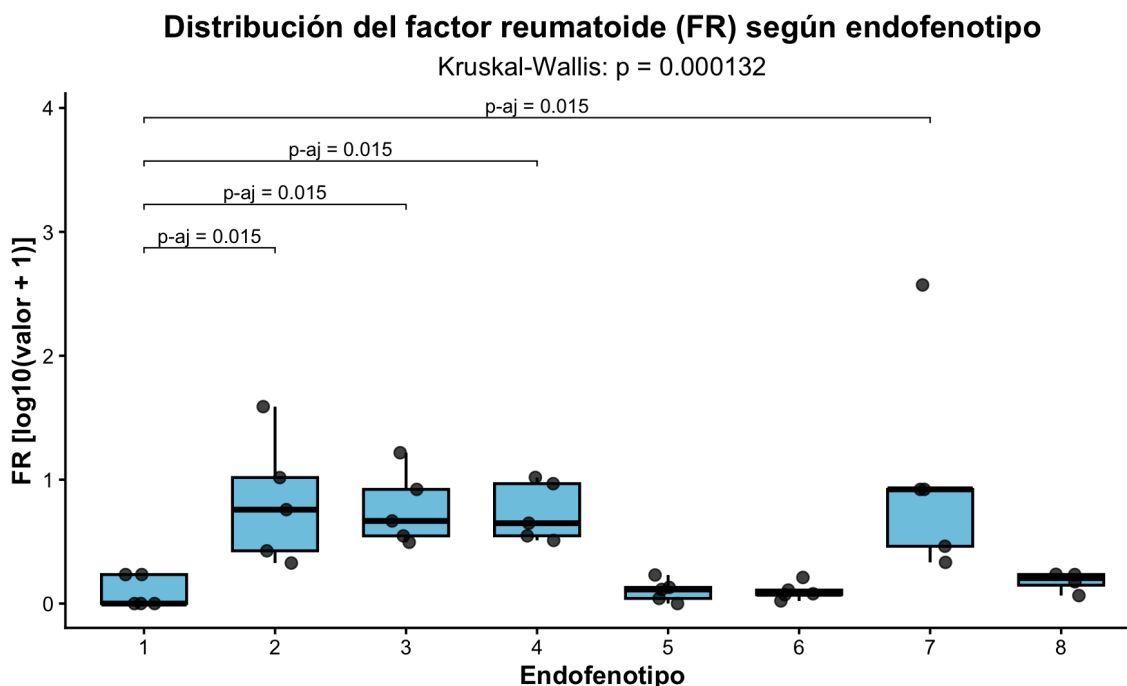


Figura 8. Distribución del factor reumatoide (FR) según endofenotipo clínico. Se muestra la distribución de los niveles de FR en los ocho endofenotipos clínicos definidos en la cohorte. Los valores fueron transformados a escala logarítmica ($\log_{10}(\text{valor} + 1)$) para mejorar la visualización de la dispersión. Cada punto representa una muestra individual, las cajas corresponden al rango intercuartílico (IQR) y la línea central indica la mediana. La comparación global entre grupos se realizó mediante la prueba de Kruskal–Wallis, seguida de un análisis post-hoc de Dunn con corrección de Benjamini-Hochberg ($p\text{-aj} < 0.05$).

En el caso de ACPA, se identificó un patrón de mayor capacidad discriminante, con valores elevados en los endofenotipos 3, 4, 5 y 8, destacándose el endofenotipo 4, mientras que los endofenotipos 1, 2, 6 y 7 mostraron niveles bajos (Figura 9). Se identificaron diferencias significativas entre grupos ($p = 0.000171$), nuevamente restringidas a comparaciones específicas, principalmente entre el endofenotipo 1 y los endofenotipos 3 ($p\text{-aj} = 0.006$) y 4 ($p\text{-aj} = 0.013$). A pesar de esta diferenciación, se mantuvo una variabilidad intragrupo relevante, indicando que la positividad no implica una magnitud uniforme de respuesta autoinmune.

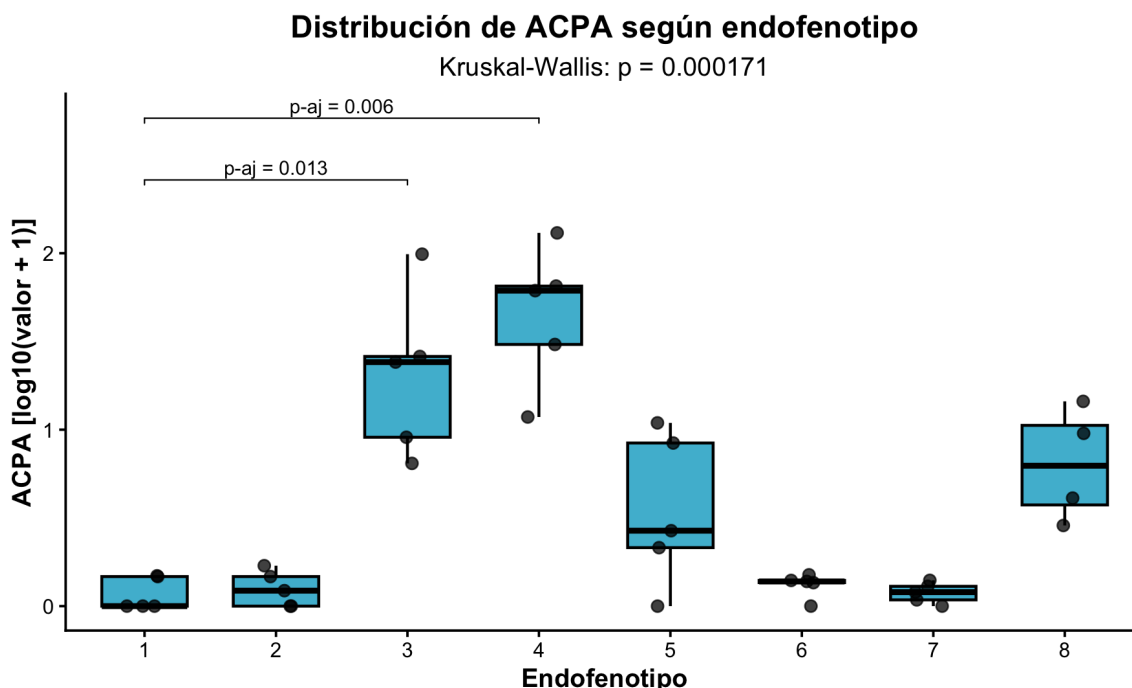


Figura 9. Distribución de los niveles de ACPA según endofenotipo clínico. Se muestra la distribución de los valores de ACPA en los ocho endofenotipos clínicos definidos en la cohorte. Los datos fueron transformados a escala logarítmica ($\log_{10}(\text{valor} + 1)$) para facilitar la visualización de la dispersión. Cada punto representa una muestra individual, las cajas corresponden al rango intercuartílico (IQR) y la línea central indica la mediana. La comparación global entre grupos se realizó mediante la prueba de Kruskal–Wallis, seguida de un análisis post-hoc de Dunn con corrección de Benjamini-Hochberg ($p\text{-aj} < 0.05$).

Ampliando este análisis, se evaluó la distribución de los ANAs como posible eje inmunológico adicional. Este marcador evidenció el patrón de separación más definido entre endofenotipos (Figura 10), con niveles elevados en los endofenotipos 4, 5, 6 y 7, y valores bajos en los restantes. Se observaron diferencias altamente significativas ($p = 2.81 \times 10^{-5}$), con múltiples comparaciones significativas en el análisis post-hoc, particularmente entre el endofenotipo 1 y los endofenotipos 4, 5, 6 y 7 ($p\text{-aj}$ entre 0.012 y 0.021), lo que confirma la existencia de una separación robusta entre grupos ANA-negativos y ANA-positivos.

A diferencia de FR y ACPA, los ANAs mostraron menor variabilidad intragrupo, sugiriendo un eje inmunológico bien definido asociado a la reactividad nuclear, capaz de discriminar con mayor consistencia subgrupos de pacientes. No obstante, la disponibilidad limitada de datos para este marcador debe considerarse en la interpretación. Estos resultados posicionan a los ANAs como el marcador serológico con mayor capacidad discriminante entre los endofenotipos evaluados, sugiriendo que este eje de respuesta autoinmune podría capturar una dimensión biológica relevante que no es completamente reflejada por FR o ACPA.

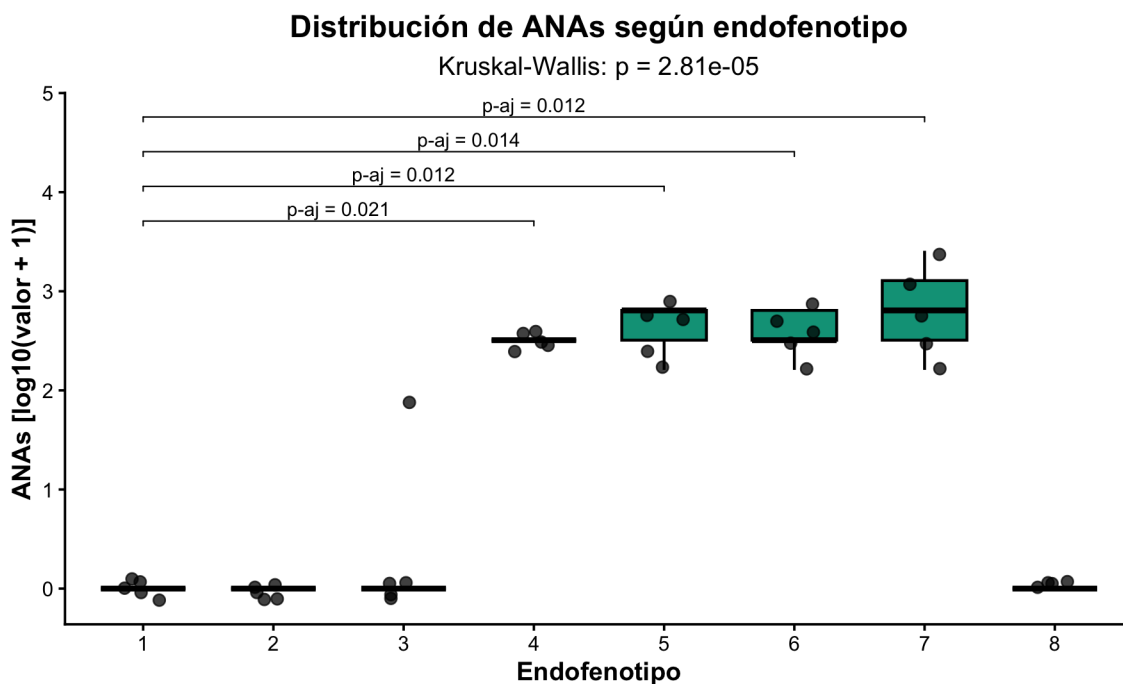


Figura 10. Distribución de los niveles de ANAs según endofenotipo clínico. Se muestra la distribución de los niveles de ANAs en los endofenotipos clínicos definidos en la cohorte. Los valores fueron transformados a escala logarítmica ($\log_{10}(\text{valor} + 1)$) para facilitar la visualización de la dispersión. Cada punto representa una muestra individual, las cajas corresponden al rango intercuartílico (IQR) y la línea central indica la mediana. La comparación global entre grupos se realizó mediante la prueba de Kruskal–Wallis, seguida de un análisis post-hoc de Dunn con corrección de Benjamini–Hochberg ($p\text{-aj} < 0.05$).

A pesar de las diferencias en los marcadores serológicos, resulta necesario determinar si estas variaciones se traducen en cambios en la actividad clínica de la enfermedad. Con este propósito, se evaluó la distribución del CDAI entre los endofenotipos clínicos (figura 11).

La distribución del CDAI evidenció una marcada heterogeneidad intragrupo, sin un patrón claro de diferenciación entre los grupos definidos. A diferencia de los marcadores serológicos, los valores de CDAI mostraron un amplio solapamiento entre endofenotipos, sin agrupaciones consistentes que permitan discriminar niveles de actividad clínica. La prueba de Kruskal–Wallis no evidenció diferencias significativas entre los endofenotipos ($p = 0.813$), Indicando que pacientes con perfiles serológicos distintos pueden presentar niveles similares de actividad clínica. Este comportamiento evidencia una disociación entre los mecanismos inmunológicos detectados por la serología y la expresión clínica de la enfermedad.

Adicionalmente, se observan valores extremos en múltiples endofenotipos, lo que refuerza la existencia de una variabilidad clínica considerable dentro de cada grupo. Esta dispersión indica que la severidad de la enfermedad no es homogénea incluso entre pacientes que comparten un mismo perfil serológico. En este contexto, aunque los marcadores serológicos permiten identificar diferencias inmunológicas entre endofenotipos, no explican la variabilidad en la actividad clínica de la AR. Esta discordancia evidencia la necesidad de integrar aproximaciones que capturen dimensiones biológicas más profundas de la AR.

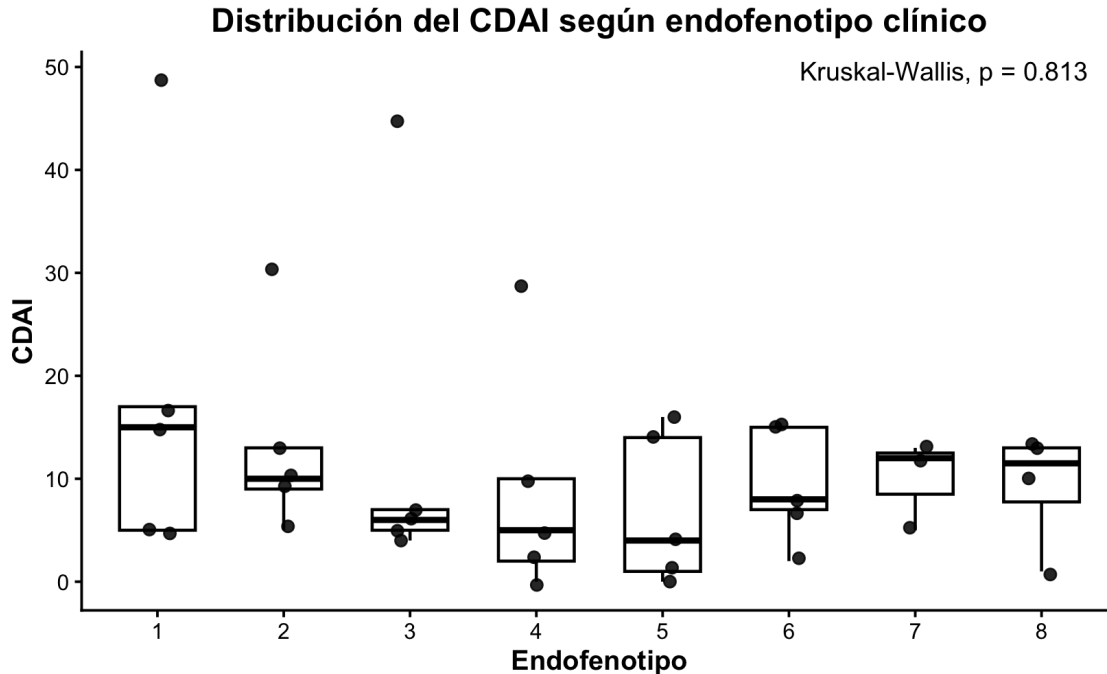


Figura 11. Distribución del CDAI según endofenotipo clínico. Se muestra la distribución de los valores de CDAI en los ocho endofenotipos clínicos definidos en la cohorte. Cada punto representa una muestra individual, las cajas corresponden al rango intercuartílico (IQR) y la línea central indica la mediana. La comparación global entre los grupos se realizó mediante la prueba de Kruskal–Wallis.

Dado que estas diferencias no se traducen de manera consistente en el CDAI, se integraron los resultados en una matriz comparativa para identificar las comparaciones con diferencias estadísticamente significativas y priorizar aquellos contrastes con mayor respaldo biológico para los análisis posteriores (figura 12). El análisis evidenció que la capacidad discriminante varía según el marcador evaluado, con un patrón no uniforme entre endofenotipos, destacándose el endofenotipo 1 por presentar diferencias con múltiples grupos.

En FR, las diferencias se concentran principalmente entre el endofenotipo 1 y los endofenotipos 2, 3, 4 y 7, lo que sugiere una capacidad de discriminación parcial. En ACPA, estas se restringen a comparaciones más específicas, particularmente con los endofenotipos 3 y 4. Por su parte, los ANAs muestran un patrón más consistente, con diferencias significativas entre el endofenotipo 1 y los endofenotipos 4, 5, 6 y 7, reforzando su papel como eje inmunológico distintivo. Algunas comparaciones, como las que involucran al endofenotipo 8, no presentan diferencias significativas en ninguno de los marcadores, lo que sugiere una menor resolución discriminativa en estos contrastes.

Estos resultados indican que la estratificación serológica permite identificar subgrupos con diferencias inmunológicas específicas, pero no captura de manera uniforme todas las relaciones entre endofenotipos. Esta variabilidad en el soporte serológico indica que la relevancia biológica de las comparaciones no es homogénea entre todos los grupos.

Bajo este panorama, se adoptó un enfoque orientado a priorizar aquellas comparaciones con mayor evidencia serológica, sin perder la visión global de la cohorte. Esto permitió enfocar la interpretación transcriptómica en contrastes con mayor sustento biológico, facilitando la identificación de patrones de expresión génica más robustos y biológicamente interpretables.

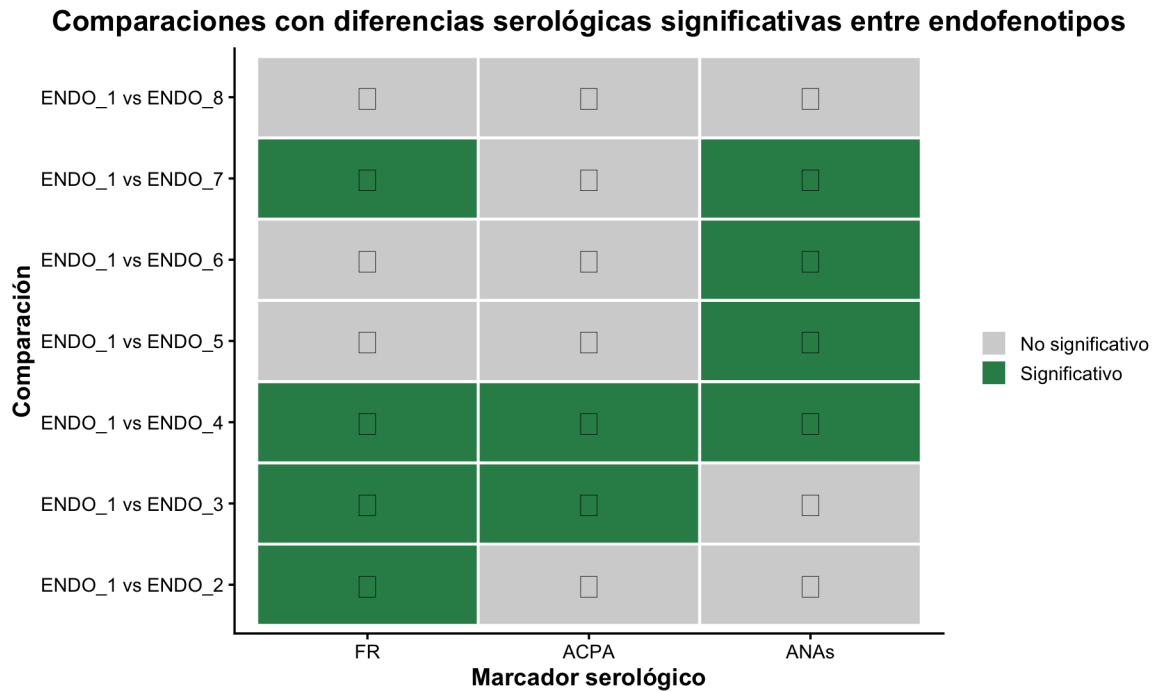


Figura 12. Comparaciones con diferencias serológicas significativas entre endofenotipos clínicos. Se muestra una matriz resumen de las comparaciones entre el endofenotipo 1 y los endofenotipos 2 a 8, indicando la presencia (verde) o ausencia (gris) de diferencias estadísticamente significativas para cada marcador serológico (FR, ACPA y ANAs). Las diferencias fueron evaluadas mediante prueba post-hoc de Dunn con corrección de Benjamini-Hochberg ($p\text{-aj} < 0.05$). Esta representación permite identificar de manera integrada las comparaciones con respaldo serológico significativo.

7.4. Estructura global del perfil transcriptómico de las muestras según endofenotipo clínico.

Con el fin de responder le primer objetivo propuesto, se exploró la estructura global del perfil transcriptómico de los PBMC mediante un análisis de componentes principales (PCA) a partir de la matriz de expresión transformada mediante VST, utilizando los 2000 genes más variables para capturar las principales fuentes de variación del conjunto de datos (figura 13).

El PCA mostró una separación clara de dos grupos principales a lo largo del PC1, componente que explicó el 94% de la variabilidad total. Un subconjunto reducido de muestras se ubicó hacia valores negativos de este eje, claramente separado del grupo mayoritario, concentrado en valores positivos de PC1. Además, dentro de este segundo grupo se observó heterogeneidad adicional a lo largo del PC2 que, aunque explicó solo el 2% de la variación, permitió evidenciar dispersión interna y sugiere la presencia de subestructura transcriptómica dentro del bloque principal.

Los endofenotipos no conforman agrupaciones completamente discretas, sino que presentan solapamiento en el espacio transcriptómico, indicando que la clasificación serológica no explica por sí sola la variabilidad biológica observada. La dispersión intragrupo y la presencia de muestras con comportamiento diferencial evidencian heterogeneidad molecular tanto entre como dentro de los endofenotipos.

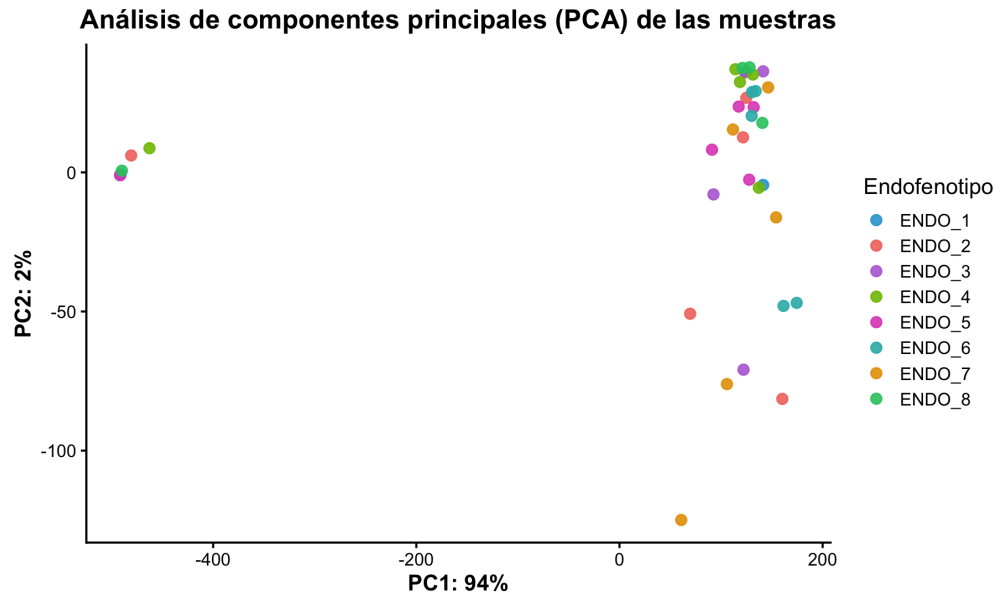


Figura 13. Análisis de componentes principales (PCA) de las 39 muestras incluidas en el estudio. El PCA se calculó a partir de la matriz de expresión génica transformada mediante Variance Stabilizing Transformation (VST). Cada punto representa una muestra individual y el color indica el endofenotipo clínico correspondiente. Los ejes muestran el porcentaje de variabilidad explicado por los dos primeros componentes principales.

Con el propósito de complementar estos hallazgos, se construyó un heatmap a partir de la matriz de expresión génica transformada mediante VST, considerando los 2000 genes con mayor variabilidad. El heatmap evidenció una organización estructurada de las muestras según sus perfiles transcriptómicos, con la formación de clústeres definidos mediante agrupamiento jerárquico (figura 14). En concordancia con el PCA, se identifica un subconjunto de muestras con un patrón de expresión claramente diferenciado, separado del resto, lo que confirma la presencia de perfiles transcriptómicos distintos dentro de la cohorte.

La mayoría de las muestras se agrupan en cuatro bloques principales, con variaciones internas que evidencian heterogeneidad intragrupo. A nivel de los endofenotipos, no se observan agrupaciones discretas, sino un solapamiento entre muestras, lo que confirma que la clasificación serológica no explica completamente la organización transcriptómica global. También, se identifican patrones de expresión compartidos entre distintos endofenotipos y variabilidad dentro de un mismo endofenotipo, respaldando la existencia de subestructuras transcriptómicas que trascienden la clasificación serológica.

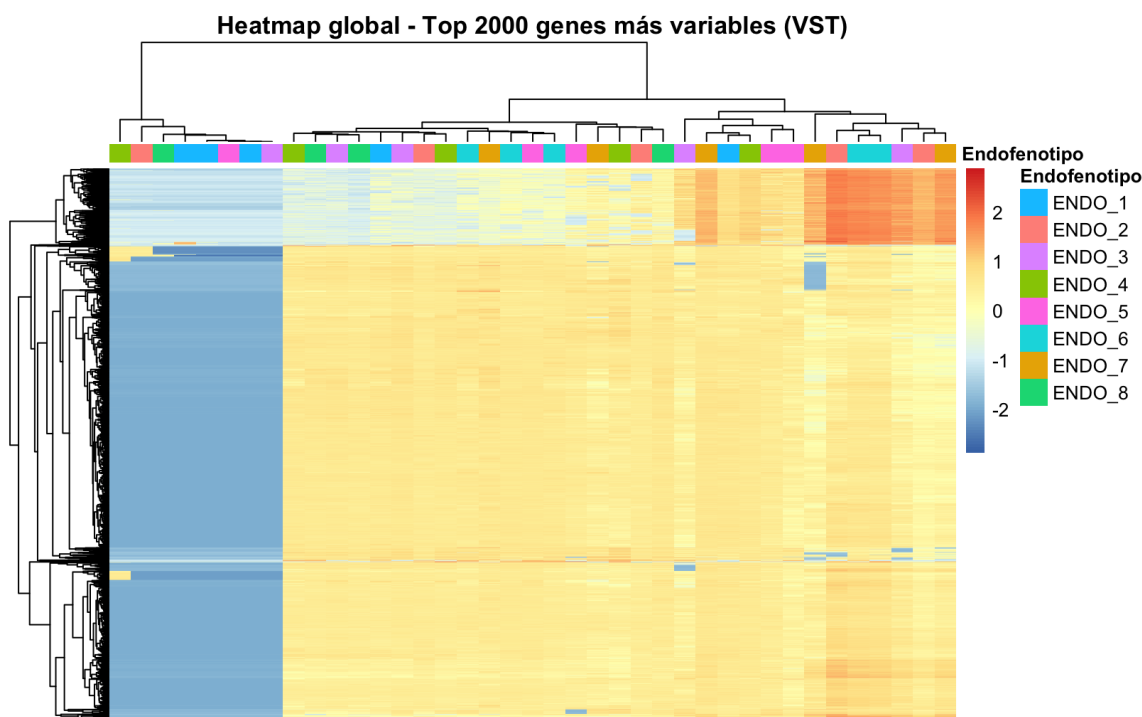


Figura 14. Heatmap global de expresión génica basado en los 2000 genes más variables (VST). El heatmap se construyó a partir de la matriz de expresión transformada mediante VST. Cada columna representa una muestra y cada fila un gen. La escala de color indica la expresión relativa (z-score por gen). El agrupamiento jerárquico se realizó tanto para genes como para muestras, y la anotación superior corresponde al endofenotipo clínico de cada muestra.

Con el objetivo de cuantificar los patrones observados en el heatmap, se construyó una matriz de correlación a partir de la expresión génica transformada mediante VST. Este análisis permitió evaluar de manera cuantitativa la similitud transcriptómica entre las muestras y evidenció una estructura organizada, caracterizada por cuatro bloques definidos de alta y baja correlación (figura 15). De forma consistente con los análisis previos, se observa un subconjunto de muestras con alta correlación entre sí, compatible con la presencia de perfiles transcriptómicos similares dentro de este grupo. Sin embargo, estos subgrupos no corresponden de manera exclusiva a los endofenotipos serológicos, sino que muestran un solapamiento entre muestras de distintos grupos, indicando que la clasificación serológica no define de forma estricta la similitud transcriptómica. Del mismo modo, se observa variabilidad intragrupo, evidenciando heterogeneidad dentro de los propios endofenotipos.

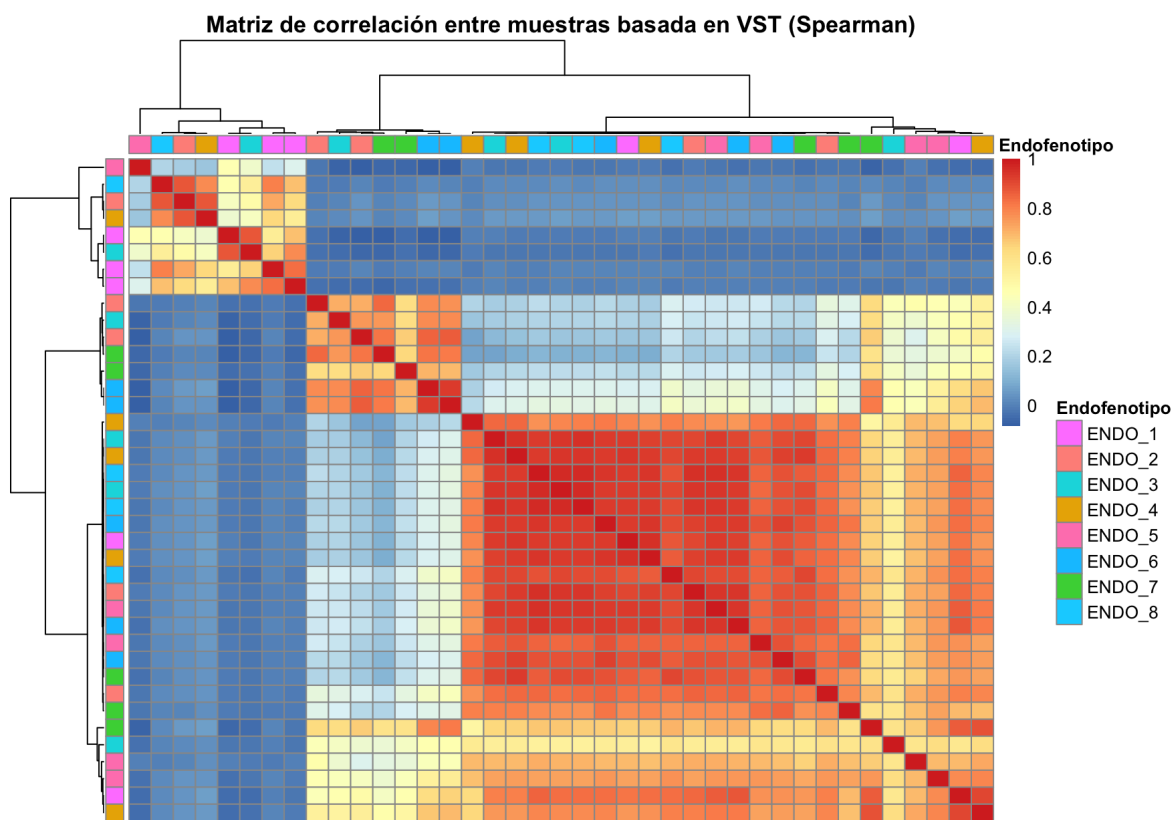


Figura 15. Matriz de correlación entre muestras basada en la matriz de expresión transformada mediante VST. La correlación entre muestras se calculó utilizando el coeficiente de Spearman a partir de la matriz de expresión génica transformada mediante VST. Cada celda representa el grado de similitud transcriptómica entre pares de muestras. La escala de color indica el valor de correlación, donde valores cercanos a 1 representan alta similitud. El agrupamiento jerárquico se realizó utilizando distancia basada en correlación, y la anotación corresponde al endofenotipo clínico de cada muestra.

Estos resultados muestran que la cohorte presenta subestructuras transcriptómicas bien definidas que no son capturadas por la clasificación serológica inicial. La integración de los análisis de PCA, heatmap y correlación demuestra que la variabilidad transcriptómica en PBMC no se distribuye de manera aleatoria, sino que se organiza en estados biológicos diferenciados dentro de la cohorte. La consistencia entre estos enfoques refuerza la robustez de los patrones identificados y sustenta la necesidad de abordar la caracterización de la AR desde un enfoque transcriptómico, capaz de capturar dimensiones funcionales del sistema inmune que no son evidentes mediante clasificaciones clínicas tradicionales.

7.5. Patrones de expresión génica intragrupo según endofenotipo clínico.

Para evaluar la organización transcriptómica dentro de cada endofenotipo serológico, se construyeron heatmaps intragrupo utilizando los 2000 genes más variables de cada conjunto de muestras (figura 16). Este análisis permitió examinar la consistencia y variabilidad de los perfiles de expresión génica al interior de cada endofenotipo.

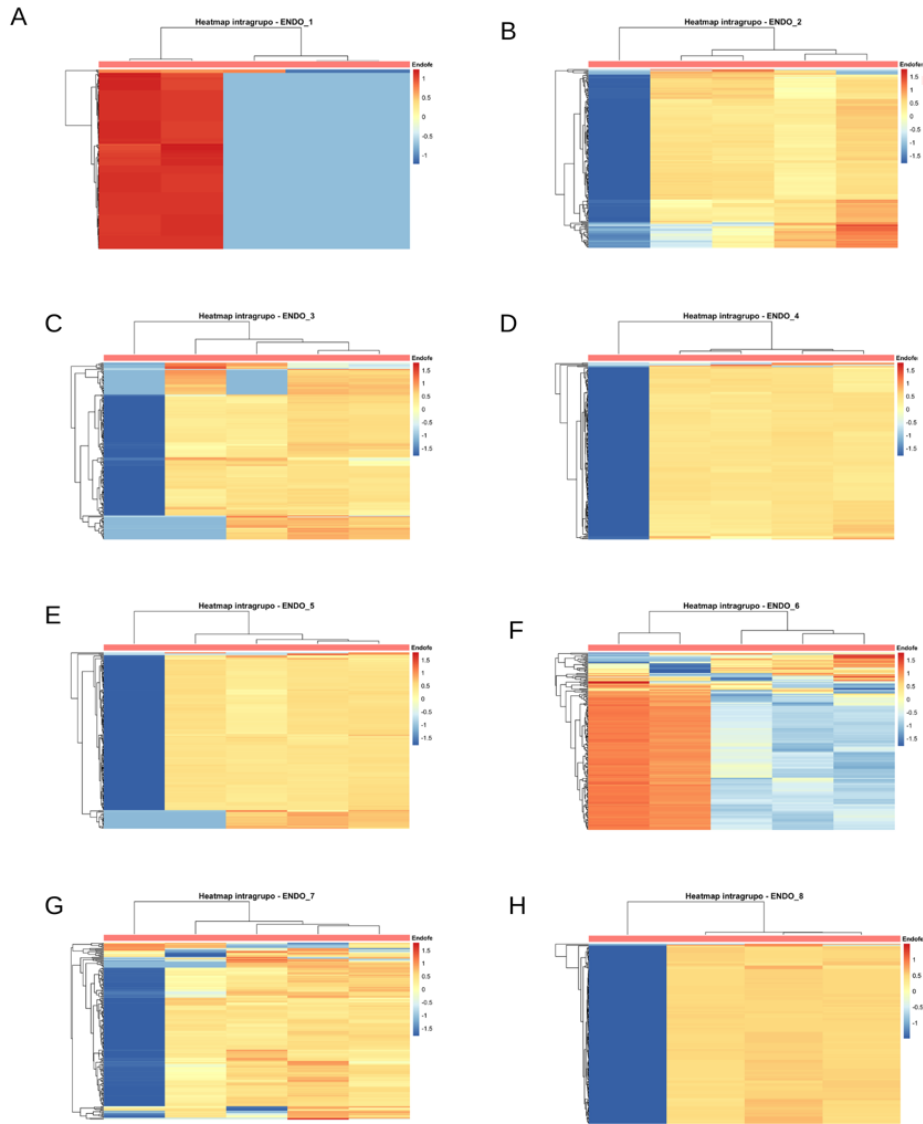


Figura 16. Heatmaps intragrupo de expresión génica en PBMC según endofenotipo clínico. Los heatmaps (A–H) representan los patrones de expresión génica para cada endofenotipo clínico (ENDO_1 a ENDO_8), contruidos a partir de la matriz de expresión transformada mediante VST y considerando los genes con mayor variabilidad dentro de cada grupo. Cada columna corresponde a una muestra y cada fila a un gen. La escala de color indica la expresión relativa (z-score por gen). El agrupamiento jerárquico se realizó tanto para genes como para muestras, permitiendo evaluar la organización transcriptómica intragrupo.

Los resultados evidencian que los endofenotipos serológicos no presentan un comportamiento homogéneo en términos de expresión génica. Aunque algunos grupos (2, 4, 5 y 8) muestran patrones relativamente consistentes, se observa de forma recurrente la presencia de muestras con perfiles transcriptómicos discordantes, introduciendo variabilidad intragrupo.

El endofenotipo 1 presenta un perfil más uniforme, aunque con dos subestructuras definidas que reflejan heterogeneidad interna. Por su parte, los endofenotipos 3 y 7 muestran mayor dispersión, mientras que el endofenotipo 6, pese de su variabilidad, comparte similitudes globales con el endofenotipo 1.

Por su parte, los endofenotipos 3 y 7 presentan mayor variabilidad intragrupo y comparten la presencia de muestras con perfiles transcriptómicos atípicos, similar a lo observado en los endofenotipos 2, 4, 5 y 8, aunque sin el mismo nivel de homogeneidad. Este comportamiento refuerza la existencia de una organización transcriptómica más compleja y menos definida entre los grupos.

Los hallazgos observados revelan una organización transcriptómica compleja, caracterizada por la coexistencia de patrones compartidos y perfiles diferenciados tanto dentro como entre grupos. Esta heterogeneidad refuerza la hipótesis de que la clasificación serológica no captura completamente la diversidad molecular de la cohorte y justifica avanzar hacia análisis comparativos de expresión génica entre endofenotipos para identificar las bases funcionales de estas diferencias.

Con el fin de precisar qué muestras contribuían a esta variabilidad intragrupo, se identificaron aquellas con perfiles de expresión génica discordantes respecto al patrón predominante de su endofenotipo. Estas muestras fueron determinadas a partir de su comportamiento en los heatmaps intragrupo, donde presentaban agrupamientos separados o patrones de expresión claramente diferenciados frente al resto de muestras del mismo grupo (Tabla 4).

Endofenotipo	Muestras con perfiles de expresión distintos al de su endofenotipo serológico
1	G1P95T1 G1P77T1
2	G1P44T1
3	G1P58T1
4	G1P42T1
5	G1P64T1
8	G1P48T1

Tabla 3. Muestras con perfiles de expresión génica discordantes respecto a su endofenotipo serológico. Se listan las muestras que, dentro de cada endofenotipo, presentan patrones transcriptómicos atípicos en comparación con el resto de las muestras del mismo grupo, identificadas a partir del análisis de agrupamiento jerárquico en los heatmaps intragrupo. Estas muestras muestran perfiles de expresión diferencial que no siguen la tendencia general de su endofenotipo.

7.6. Análisis comparativo de la expresión génica entre endofenotipos clínicos y optimización del conjunto de datos.

Con base en los resultados del análisis de los marcadores serológico y su distribución entre endofenotipos, se identificaron diferencias significativas concentradas en comparaciones que involucran al endofenotipo 1 (figura 12), evidenciando su comportamiento diferencial dentro de la cohorte.

A partir de ello, se realizó un análisis comparativo de expresión génica mediante 28 contrastes pareados entre endofenotipos, lo que permitió explorar de manera exhaustiva las posibles variaciones transcriptómicas entre los distintos perfiles serológicos definidos. Considerando la evidencia estadística previa, el análisis se orientó principalmente hacia las comparaciones que incluyen al endofenotipo 1, dado su mayor respaldo serológico, mientras que las restantes se incluyen en los anexos como información complementaria.

Con el objetivo de caracterizar la organización transcriptómica y evaluar el grado de separación entre los grupos comparados, se realizó un PCA para cada una de las comparaciones seleccionadas, utilizando la matriz de expresión génica transformada mediante VST (Figura 14).

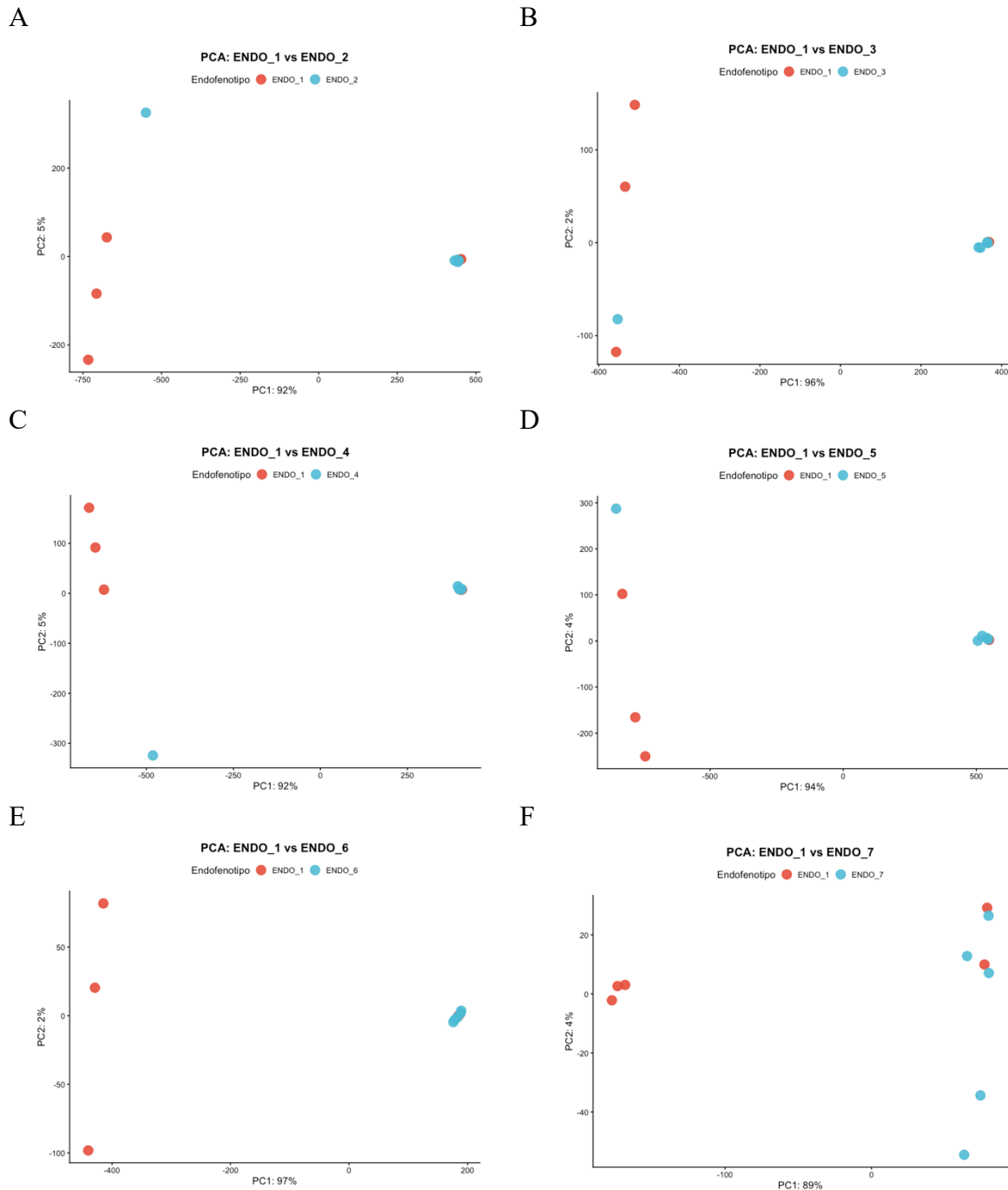
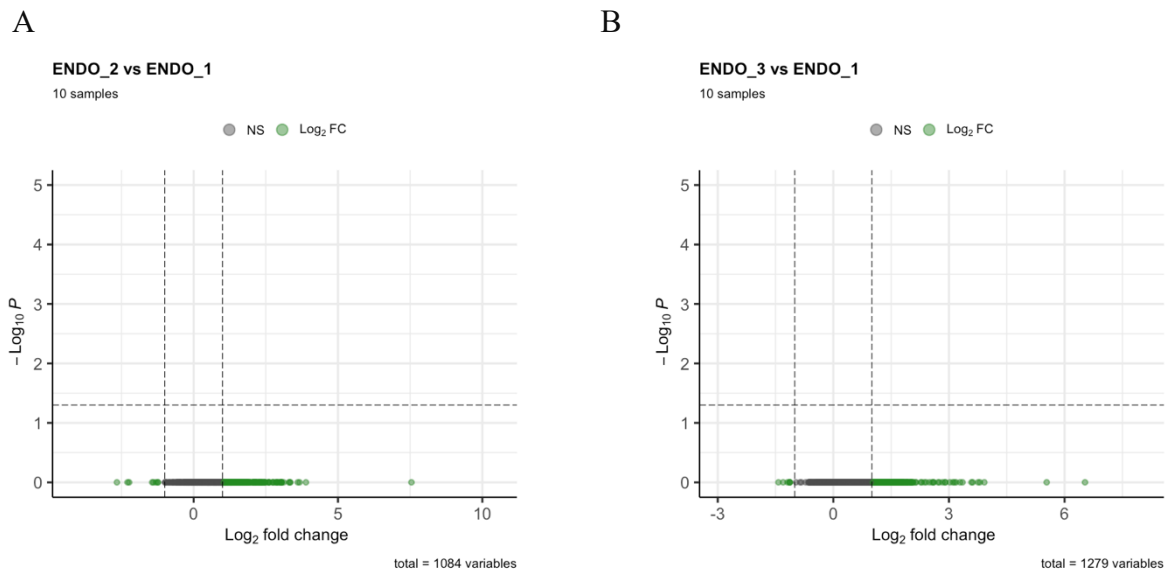


Figura 14. PCA de las comparaciones entre el endofenotipo 1 y los demás endofenotipos clínicos. Cada panel (A–F) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 7, respectivamente. El PCA se construyó a partir de la matriz de expresión génica transformada mediante VST, utilizando los 2000 genes con mayor variabilidad. Cada punto representa una muestra individual y el color indica el endofenotipo correspondiente.

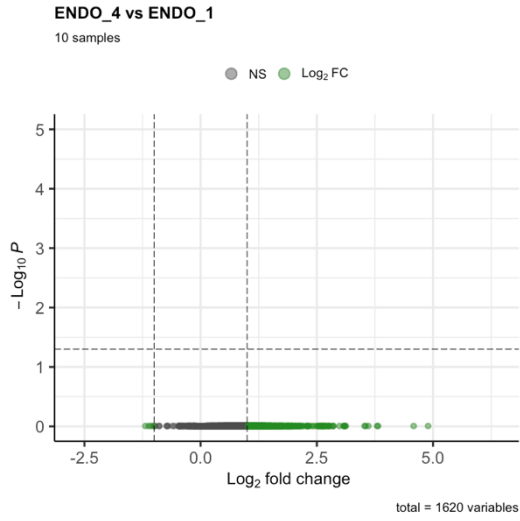
Se observó un solapamiento parcial entre muestras de distintos endofenotipos, evidenciando perfiles transcriptómicos no concordantes con la clasificación serológica, en línea con lo descrito previamente (sección 7.4). En particular, las muestras G1P95T1 y G1P77T1 del endofenotipo 1 mostraron un patrón consistente de agrupamiento con todos los grupos comparados (paneles A–G), indicando un perfil de expresión más cercano al de otros endofenotipos que al de su propio grupo. Este comportamiento, también observado en el heatmap, no pudo ser explicado por variables clínicas disponibles, lo que sugiere la presencia de variabilidad biológica no capturada por la serología. De forma similar, se identificaron muestras de otros grupos que se aproximan al endofenotipo 1 (G1P44T1, G1P58T1, G1P42T1, G1P64T1 y G1P48T1), lo que indica la presencia de perfiles transcriptómicos transversales a la clasificación serológica.

Este patrón de solapamiento bidireccional evidencia heterogeneidad intragrupo, donde algunas muestras no comparten el perfil transcriptómico de su endofenotipo. Estos resultados confirman que la clasificación serológica no define de manera estricta el perfil transcripcional de los pacientes, sugiriendo que la variabilidad transcriptómica se organiza en ejes biológicos transversales.

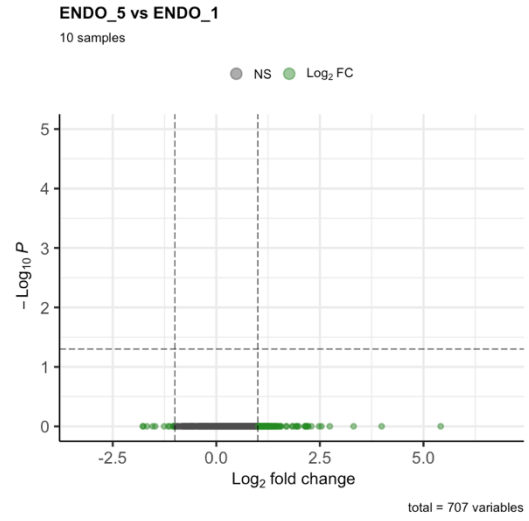
Para evaluar las diferencias en la expresión génica entre los endofenotipos priorizados, se realizó un análisis de expresión diferencial para cada una de las comparaciones que involucran al endofenotipo 1, cuyos resultados se visualizaron mediante volcano plots (Figura 15).



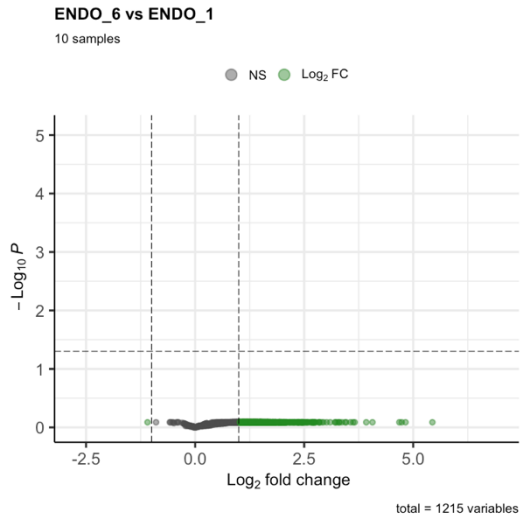
C



D



E



F

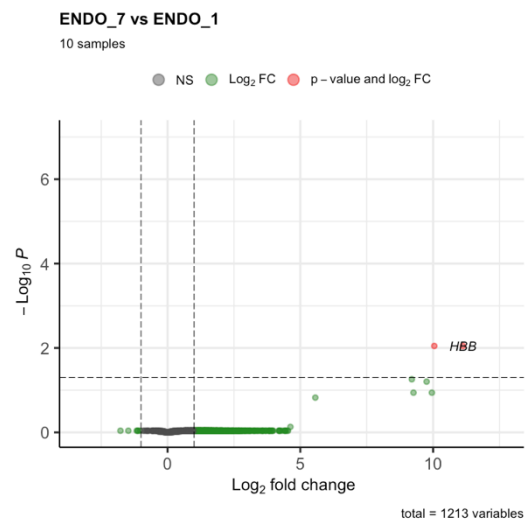


Figura 15. Análisis de expresión diferencial entre el endofenotipo 1 y los demás endofenotipos clínicos. Cada panel (A–F) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 7, respectivamente. Los volcano plots muestran la relación entre el cambio de expresión (log₂ fold change) y la significancia estadística (-log₁₀ p-ajustada). Las líneas punteadas indican los umbrales de significancia (p-ajustada < 0.05) y magnitud de cambio ($|\log_2 FC| \geq 1$). No se identificaron genes diferencialmente expresados bajo estos criterios en ninguna de las comparaciones.

En todas las comparaciones (paneles A–F), no se identificaron genes diferencialmente expresados bajo los criterios establecidos (p-ajustada < 0.05 y $|\log_2 FC| \geq 1$), observándose una distribución de los genes cercana al eje central y con baja significancia estadística. Este resultado indica la ausencia de diferencias transcriptómicas robustas en esta etapa del análisis.

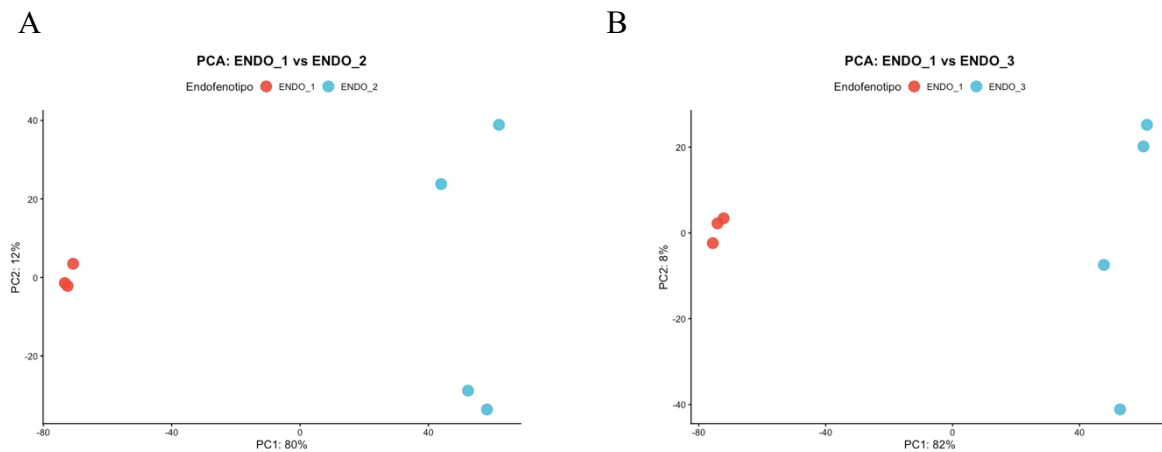
Este patrón sugiere que la variabilidad intragrupo y la presencia de muestras con perfiles atípicos, previamente identificadas, podrían estar enmascarando las diferencias reales entre endofenotipos. En este contexto, la ausencia de DEGs no refleja homogeneidad biológica, sino limitaciones en la resolución del análisis asociadas a la heterogeneidad del conjunto de datos.

La presencia de muestras con perfiles transcriptómicos discordantes indica que parte de la variabilidad de la cohorte está influenciada por subconjuntos específicos y no se distribuye de manera homogénea, reforzando la interpretación anterior.

Para evaluar el impacto de estas muestras discordante sobre la expresión génica, se realizó un proceso de depuración del conjunto de datos mediante su exclusión (tabla 4), con el objetivo de reducir la variabilidad no estructurada y mejorar la detección de patrones de expresión génica. Este enfoque permitió plantear una reevaluación de la organización transcriptómica bajo condiciones de menor heterogeneidad, orientada a identificar señales biológicas previamente enmascaradas.

7.7. Reevaluación de la estructura transcriptómica tras la depuración del conjunto de datos.

Tras la depuración del conjunto de datos, se recalcularon los PCA con el fin de evaluar el impacto de la exclusión de las muestras discordantes sobre la organización transcriptómica global. Como se observa en la figura 16, las comparaciones entre el endofenotipo 1 y los demás grupos muestran una separación clara y consistente a lo largo del PC1 en todos los casos. Este componente explica entre el 73% y 88% de la variabilidad total, evidenciando un eje dominante que, tras la depuración, permite discriminar de manera robusta entre los endofenotipos.



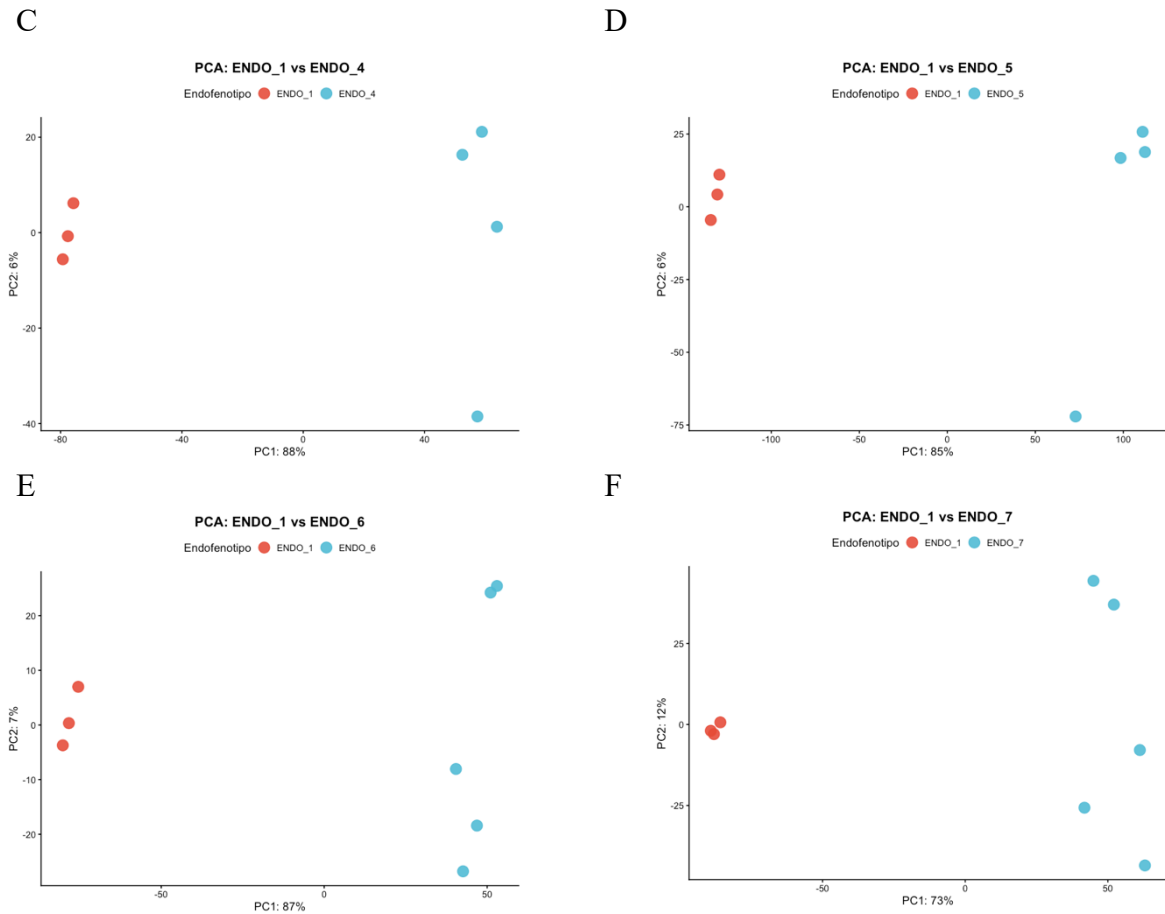


Figura 16. PCA de las comparaciones entre el endofenotipo 1 y los demás endofenotipos clínicos tras la depuración del conjunto de datos. Cada panel (A–G) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 7, respectivamente. El PCA se construyó a partir de la matriz de expresión génica transformada mediante VST, utilizando los 2000 genes con mayor variabilidad. Cada punto representa una muestra individual y el color indica el endofenotipo correspondiente.

Si los comparamos con los PCA iniciales, donde se observaba solapamiento entre muestras de distintos endofenotipos, la depuración del conjunto de datos reduce significativa este fenómeno, evidenciando una organización más definida en el espacio transcriptómico. La exclusión de las muestras discordantes elimina fuentes de variabilidad que interferían con la separación entre grupos.

A nivel intragrupo, se observa un patrón de agrupación más compacto y consistente dentro de cada endofenotipo, reflejando una disminución de la heterogeneidad interna. El endofenotipo 1 mantiene un perfil transcriptómico claramente diferenciado, reforzando su carácter distintivo.

Para cuantificar estas diferencias transcriptómicas, se realizó un análisis de expresión diferencial comparando el endofenotipo 1 frente a los endofenotipos 2 a 7, determinando el número de DEGs en cada comparación, diferenciando entre el número de genes sobreexpresados y subexpresados.

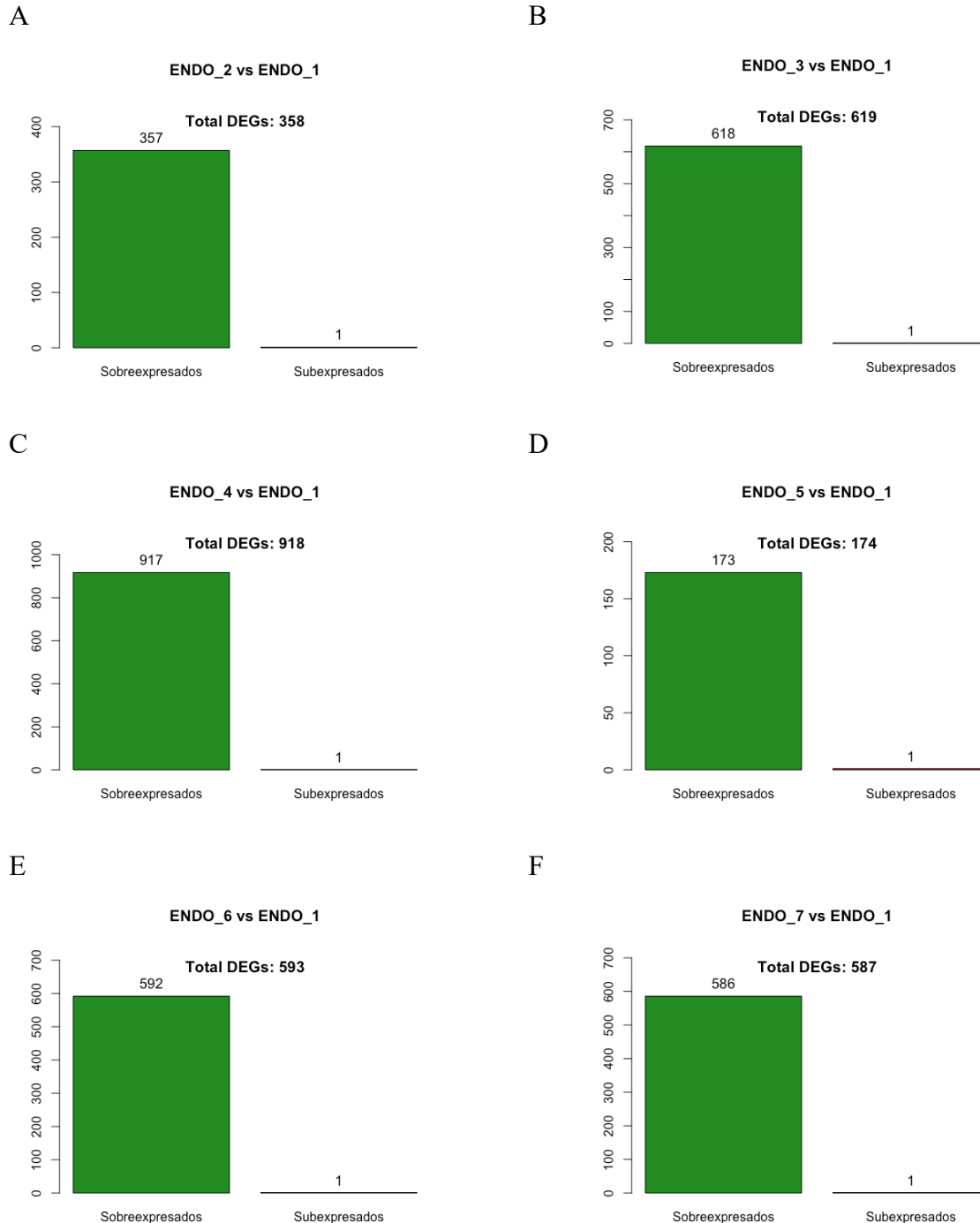
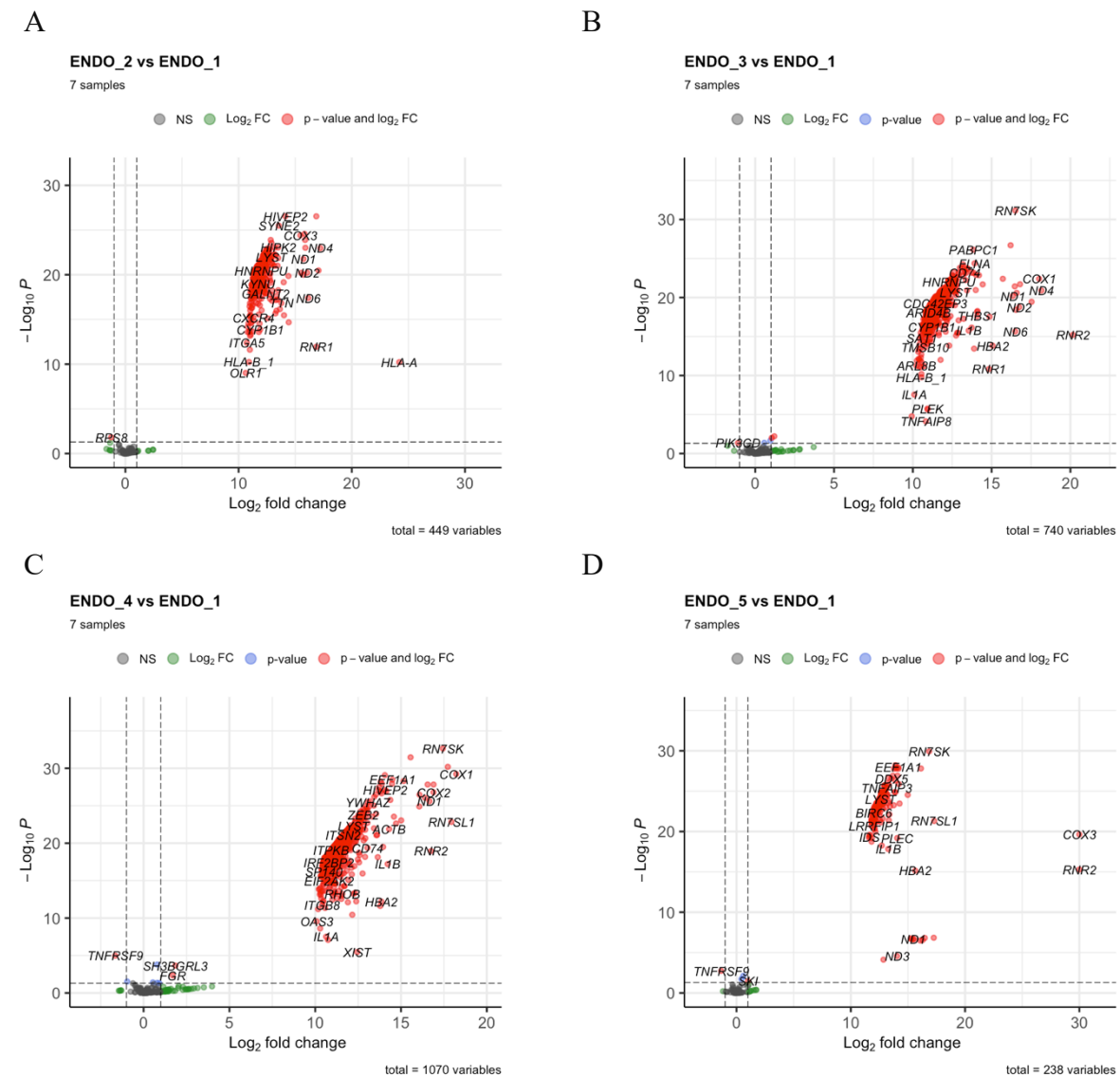


Figura 17. Número de genes diferencialmente expresados entre el endofenotipo 1 y los endofenotipos 2 a 8. Cada panel (A–F) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 8, respectivamente. Las barras muestran el número de genes sobreexpresados y subexpresados identificados en cada contraste, de acuerdo con los criterios de significancia estadística y magnitud de cambio establecidos para el análisis diferencial. El número total de DEGs se indica en la parte superior de cada panel.

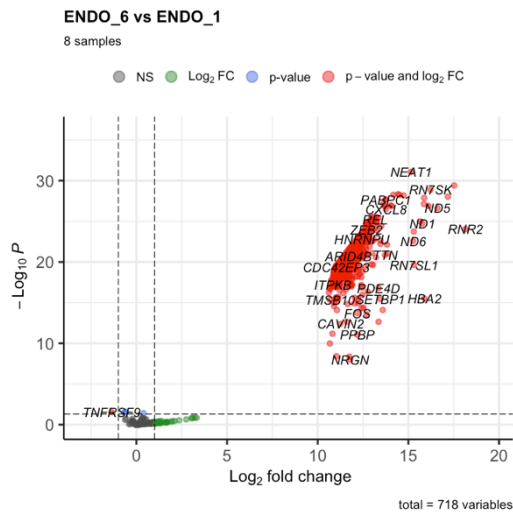
Todas las comparaciones evidenciaron un número considerable de DEGs, indicando que la reducción de la heterogeneidad intragrupo permitió revelar diferencias transcriptómicas previamente enmascaradas. Este comportamiento confirma que el endofenotipo 1 presenta un perfil de expresión consistentemente diferenciado frente a los demás grupos.

El número de DEGs varió entre comparaciones, evidenciando distintos grados de divergencia transcriptómica. La comparación con el endofenotipo 4 presentó el mayor número de DEGs ($n = 918$), mientras que el endofenotipo 5 mostró el menor ($n = 174$), lo que indica que las diferencias no son uniformes entre los grupos. Un patrón consistente fue el predominio de genes sobreexpresados. La mayoría de los DEGs correspondieron a incrementos en la expresión en los endofenotipos comparados frente al endofenotipo 1, mientras que los genes subexpresados fueron escasos o ausentes, sugiriendo desplazamientos transcriptómicos sistemáticos más que cambios bidireccionales.

Para integrar la magnitud y significancia de los cambios, se analizaron los volcano plots correspondientes componentes (figura 18).



E



F

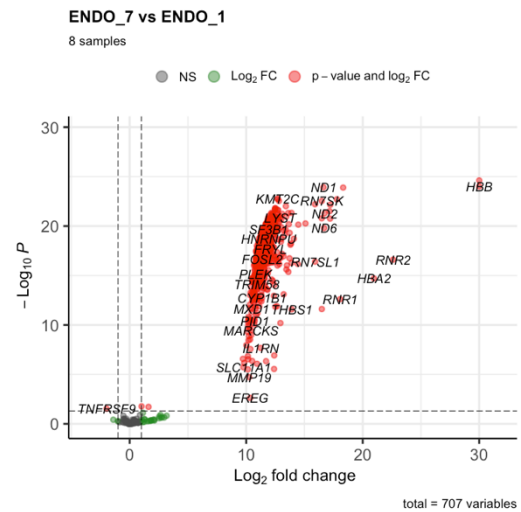


Figura 18. Volcano plots de la expresión génica diferencial entre el endofenotipo 1 y los endofenotipos 2 a 7. Cada panel (A–F) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 8, respectivamente. El eje X representa el cambio en la expresión génica ($\log_2$ fold change) y el eje Y la significancia estadística ($-\log_{10}$ p-value ajustado). Los puntos en rojo corresponden a genes diferencialmente expresados que cumplen simultáneamente los criterios de significancia estadística ($\text{padj} < 0.05$) y magnitud de cambio ($|\log_2\text{FC}| \geq 1$). Se resaltan algunos de los genes más relevantes por su magnitud de cambio y nivel de significancia.

Los volcano plots muestran una distribución asimétrica de los DEGs, con desplazamiento hacia valores positivos de $\log_2$ fold change, consistente con el predominio de genes sobreexpresados. Este patrón indica que los endofenotipos 2 al 7 presentan perfiles sistemáticamente desplazados respecto al endofenotipo 1.

Entre los genes recurrentes se identifican aquellos asociados a funciones mitocondriales y a procesos inmunológicos e inflamatorios, lo que indica que las diferencias transcriptómicas están impulsadas por alteraciones en el metabolismo energético y la activación inmune. Para profundizar en esta interpretación, se construyó un ranking de los 10 genes diferencialmente expresados más relevantes en cada comparación (Figura 19), ordenados según su significancia estadística (padj) y magnitud de cambio ($\log_2\text{FC}$).

El análisis evidencia un patrón recurrente de genes mitocondriales (COX1, COX2, COX3, ND1–ND5, CYTB), reguladores transcripcionales (NEAT1, RN7SK, MALAT1) y moduladores de la respuesta inmune (HIVEP2, HIPK2, CD44, SOD2), lo que sugiere tres ejes biológicos principales: reprogramación metabólica mitocondrial, reorganización del control transcripcional y activación de rutas inmuno-inflamatorias.

A nivel comparativo, se identifican agrupaciones entre comparaciones con similitud transcriptómica. ENDO_2 vs ENDO_1 y ENDO_5 vs ENDO_1 comparten un perfil dominado por genes mitocondriales y reguladores transcripcionales, mientras que ENDO_3, ENDO_4 y ENDO_6 integran componentes metabólicos, estructurales y de señalización inmune. Por su parte, ENDO_7 vs ENDO_1 presenta un perfil más heterogéneo y menos recurrente, lo que sugiere un estado molecular diferenciado.

Top 10 genes diferencialmente expresados por comparación						
Genes ordenados por padj y log2FC						
	ENDO_2 vs ENDO_1	ENDO_3 vs ENDO_1	ENDO_4 vs ENDO_1	ENDO_5 vs ENDO_1	ENDO_6 vs ENDO_1	ENDO_7 vs ENDO_1
1	COX1	RN7SK	RN7SK	NEAT1	NEAT1	HBB
2	ND4	NEAT1	NEAT1	RN7SK	HIVEP2	ND4
3	COX2	KDM6B	ND4	AHNAK	COX1	ND5
4	HIVEP2	PABPC1	COX1	PABPC1	CD44	ND1
5	CYTB	SYNE2	ND5	DDX5	CYTB	C9orf163
6	COX3	HIPK2	COX3	RNF213	ND4	ND2
7	HIPK2	HUWE1	AHNAK	FLNA	SOD2	LYST
8	HUWE1	MSN	PABPC1	SRRM2	SYNE2	NEAT1
9	SYNE2	SIPA1L1	EEF1A1	SYNE2	RN7SK	HIPK2
10	ND1	TLN1	MALAT1	EEF1A1	THBS1	CTNNB1

Figura 19. Ranking de los 10 genes diferencialmente expresados más significativos en cada comparación entre el endofenotipo 1 y los endofenotipos 2 a 7. Los genes se ordenaron según su significancia estadística (padj) y magnitud de cambio en la expresión (log2 fold change). Las celdas resaltadas indican genes compartidos entre múltiples comparaciones, evidenciando patrones transcriptómicos recurrentes.

Aunque existe un núcleo común de genes compartidos, la presencia de componentes específicos en cada comparación indica que los endofenotipos combinan firmas transcriptómicas compartidas con elementos particulares. Esta organización, en concordancia con la separación observada en el PCA y respaldada por los análisis de DEGs y volcano plots, evidencia que las diferencias transcriptómicas entre endofenotipos son estructuradas y reproducibles, donde el endofenotipo 1 emerge como un estado molecular claramente diferenciado, mientras que los demás endofenotipos comparten patrones parcialmente comunes, pero conservan componentes distintivos.

Este comportamiento sugiere que las diferencias no responden a cambios aislados en genes individuales, sino a la activación de programas biológicos que involucran múltiples niveles de regulación transcripcional. Con base en estos hallazgos, el análisis funcional de los genes diferencialmente expresados resulta fundamental para comprender la organización de estas alteraciones a nivel de rutas y redes biológicas.

7.8. Análisis de enriquecimiento funcional de los genes diferencialmente expresados.

Para profundizar en la interpretación funcional de los DEGs, se empleó la Ontología Génica, priorizando la categoría de Procesos Biológicos (BP) como eje central de este análisis, dada su capacidad para integrar los cambios transcriptómicos en términos de programas biológicos coherentes entre endofenotipos. A partir de estos resultados, el análisis de redes permitió observar que los procesos enriquecidos no actúan de forma aislada, sino que se organizan en módulos funcionales interconectados, reflejando patrones de cohesión, centralidad y jerarquía que estructuran la respuesta biológica. Por su parte, las categorías de Funciones Moleculares (MF) y Componentes Celulares (CC) aportaron información más específica pero

menos integradora frente a los objetivos de este estudio, por lo que se presentan como material complementario en los anexos (anexo 2).

Los resultados de enriquecimiento en el apartado BP evidencian que las comparaciones no generan perfiles independientes, sino que comparten ejes funcionales definidos, sugiriendo una organización biológica común entre subconjuntos de endofenotipos. Particularmente las comparaciones, ENDO_2 vs ENDO_1 y ENDO_5 vs ENDO_1 convergen en un núcleo funcional dominado por la señalización de NF- κ B, la transducción intracelular y la respuesta a estímulos hormonales, configurando un estado inflamatorio compartido (Figura 20).

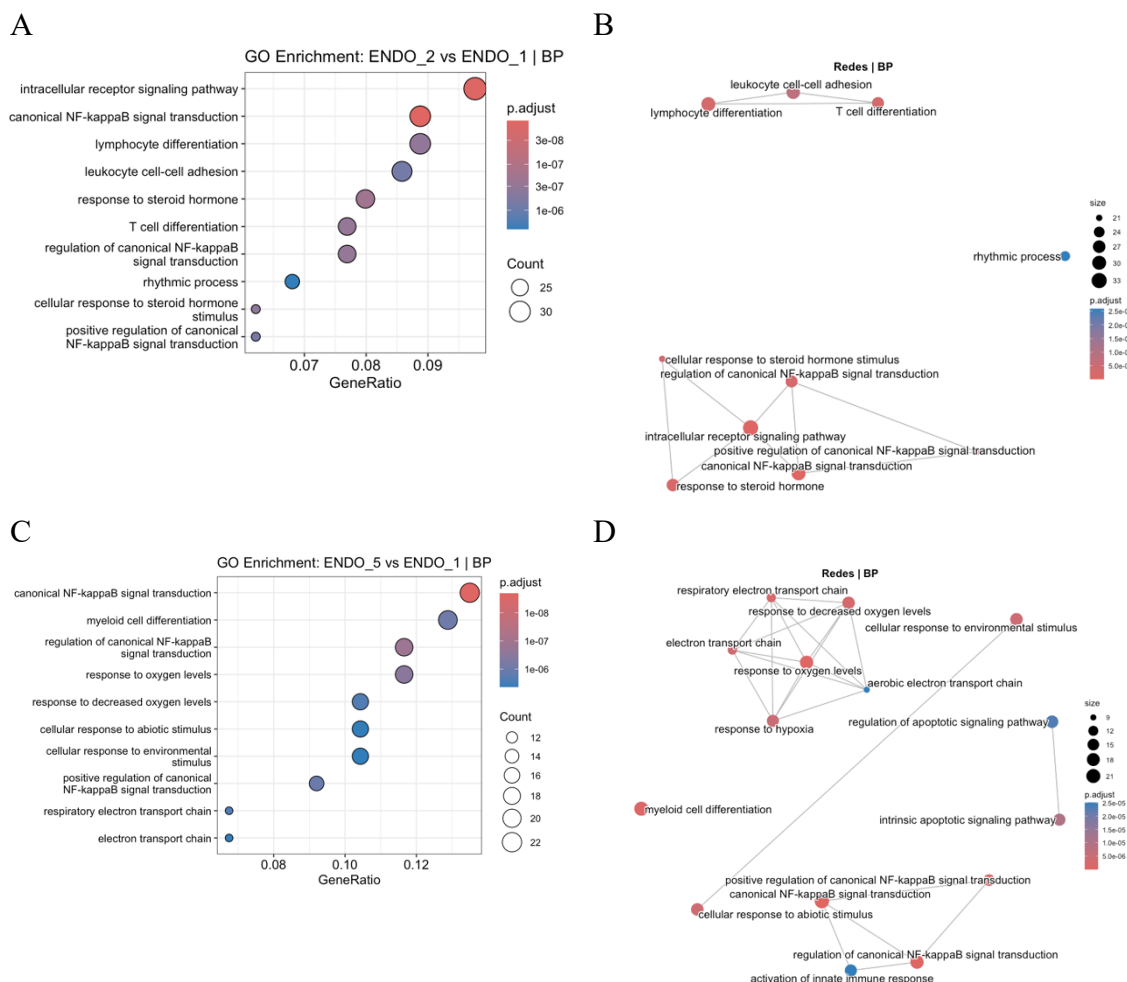


Figura 20. Enriquecimiento funcional de procesos biológicos (GO-BP) en comparaciones entre endofenotipos clínicos. (A, C) Análisis de enriquecimiento de Gene Ontology (Biological Process) para los genes diferencialmente expresados en las comparaciones ENDO_2 vs ENDO_1 y ENDO_5 vs ENDO_1, respectivamente. El eje X representa el GeneRatio, el tamaño de los puntos indica el número de genes asociados (Count) y el color corresponde al valor de significancia ajustada (p.adjust). (B, D) Redes de enriquecimiento funcional que muestran la relación entre los términos biológicos enriquecidos en cada comparación. Los nodos representan procesos biológicos y su tamaño refleja el número de genes involucrados, mientras que el color indica la significancia estadística.

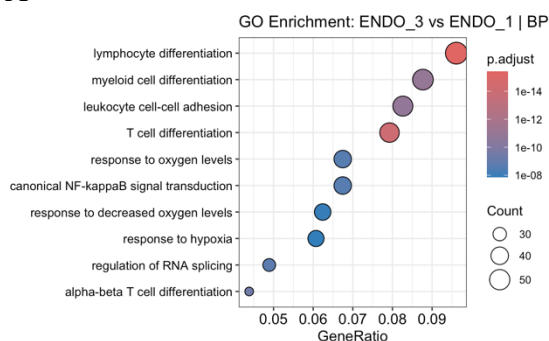
A partir de este eje común, se observan diferencias en la organización funcional de cada endofenotipo. En ENDO_2, este núcleo inflamatorio se articula con procesos de diferenciación linfocitaria, diferenciación de células T y adhesión leucocitaria, sugiriendo una expansión hacia la activación del componente inmunológico adaptativo. Por su parte, ENDO_5 incorpora procesos asociados a hipoxia, cadena de transporte de electrones y

respuesta a estímulos ambientales, sugiriendo una adaptación metabólica frente a condiciones de estrés inflamatorio.

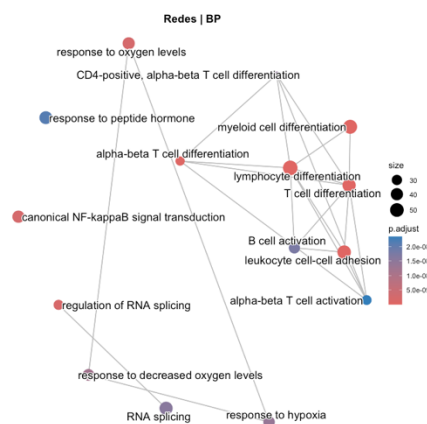
El análisis de redes funcionales respalda este patrón. En ENDO_2 vs ENDO_1, se observa un núcleo compacto dominado por la señalización de NF- κ B y la transducción intracelular, con extensiones hacia diferenciación linfocitaria y adhesión leucocitaria, indicando una organización inmunológica más estructurada. En ENDO_5 vs ENDO_1, aunque se conserva este eje inflamatorio, la red incorpora un módulo adicional asociado a hipoxia y metabolismo energético, menos integrado al núcleo principal. La unión de estos resultados refleja que ambas comparaciones comparten una base inflamatoria común, pero se diferencian en su organización funcional: ENDO_2 orientado a la activación inmune y ENDO_5 a la integración de componentes metabólicos.

En un segundo subgrupo, las realizadas entre ENDO_3 vs ENDO_1, ENDO_4 vs ENDO_1 y ENDO_6 vs ENDO_1 muestran un patrón funcional dominado por procesos de diferenciación linfocitaria y mieloide, activación de células T y adhesión leucocitaria, lo que sugiere un estado de respuesta inmune estructurada más allá de la inflamación basal. Sobre esta base común emergen variaciones específicas: ENDO_3 incorpora señales asociadas a hipoxia, ENDO_4 refuerza componentes de activación leucocitaria y hematopoyesis, mientras que ENDO_6 introduce de manera consistente mecanismos de procesamiento del ARN.

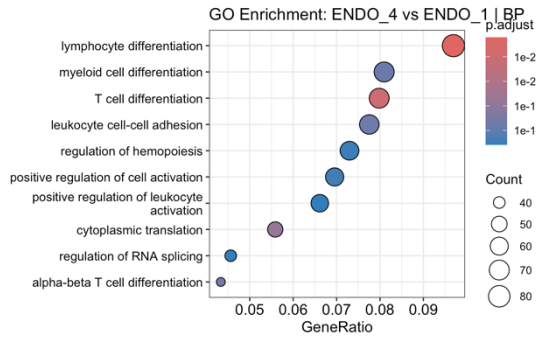
A



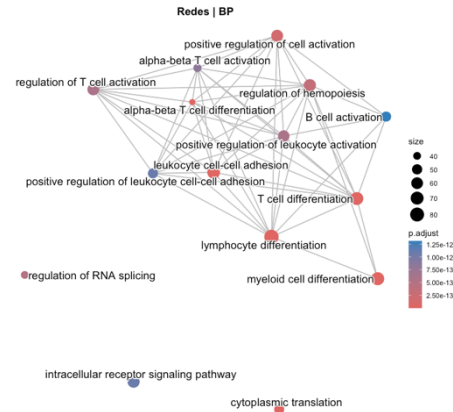
B



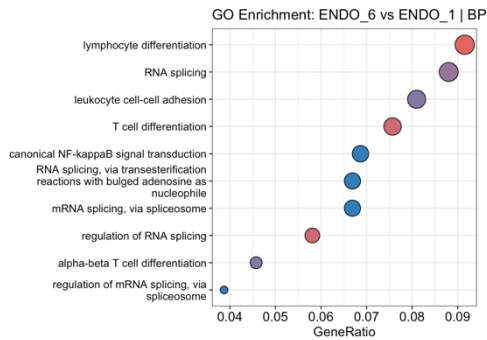
C



D



E



F

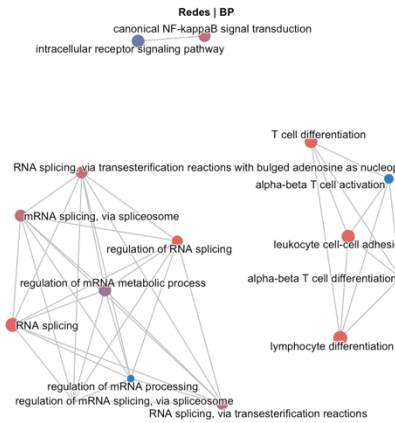


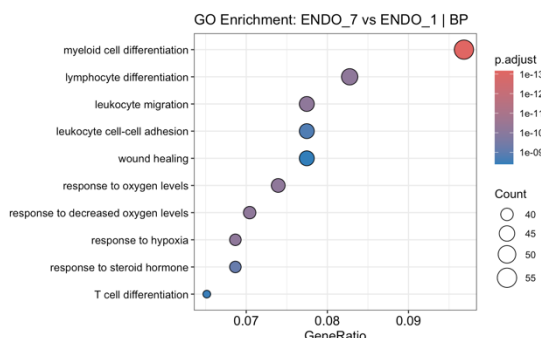
Figura 21. Enriquecimiento funcional de procesos biológicos (GO-BP) en comparaciones entre endofenotipos clínicos. (A, C, E) Análisis de enriquecimiento de Gene Ontology (Biological Process) para los genes diferencialmente expresados en las comparaciones ENDO_3 vs ENDO_1, ENDO_4 vs ENDO_1 y ENDO_6 vs ENDO_1, respectivamente. El eje X representa el GeneRatio, el tamaño de los puntos indica el número de genes asociados (Count) y el color corresponde al valor de significancia ajustada (p.adjust). (B, D, F) Redes de enriquecimiento funcional que muestran la relación entre los términos biológicos enriquecidos en cada comparación. Los nodos representan procesos biológicos y su tamaño refleja el número de genes involucrados, mientras que el color indica la significancia estadística.

A nivel de red, estos procesos se organizan en una estructura densa con un núcleo central conectado, del cual se desprenden módulos periféricos que definen la especificidad de cada endofenotipo. En ENDO_3, los términos asociados a hipoxia se integran de forma periférica; en ENDO_4, la red se expande incorporando funciones hematopoyéticas; y en ENDO_6, se distingue un módulo de procesamiento del ARN parcialmente desacoplado del núcleo principal. Esta organización indica que un mismo programa inmunológico central se despliega de manera diferencial entre endofenotipos, integrando componentes metabólicos y postranscripcionales que modulan su expresión funcional (Figura 21).

Por último, la comparación ENDO_7 vs ENDO_1 presenta un perfil funcional diferenciado de los demás subgrupos. Este subgrupo está dominado por procesos de migración leucocitaria, fagocitosis, organización del citoesqueleto y respuesta a hipoxia. La asociación a estos procesos sugiere un estado inmunológico orientado a la actividad efectora y a la interacción con el tejido, más que a programas de regulación o diferenciación celular,

evidenciando un microambiente inflamatorio dinámico en el que coexisten mecanismos de daño y reparación.

A



B

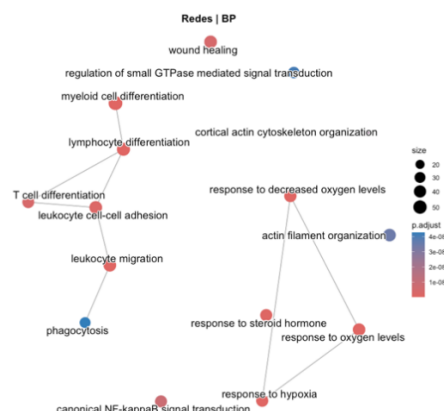


Figura 22. Enriquecimiento funcional de los genes diferencialmente expresados basado en términos de Gene Ontology (Proceso biológico) para las comparaciones de los endofenotipos ENDO_1 vs ENDO_7. (A) Análisis de enriquecimiento de Gene Ontology (Biological Process) para los genes diferencialmente expresados en las comparaciones ENDO_7 vs ENDO_1. El eje X representa el GeneRatio, el tamaño de los puntos indica el número de genes asociados (Count) y el color corresponde al valor de significancia ajustada (p.adjust). (B) Redes de enriquecimiento funcional que muestran la relación entre los términos biológicos enriquecidos en cada comparación. Los nodos representan procesos biológicos y su tamaño refleja el número de genes involucrados, mientras que el color indica la significancia estadística.

En esta comparación, la red adopta una estructura menos densa y más fragmentada, sin un núcleo funcional dominante, donde los módulos se distribuyen de forma dispersa. Los procesos asociados a migración, reorganización estructural, hipoxia y fagocitosis muestran una conectividad limitada, indicando la coexistencia de múltiples respuestas paralelas más que la activación de un programa central unificado. Esta organización sugiere un estado funcional heterogéneo, caracterizado por la convergencia de mecanismos efectores y de remodelación tisular en un entorno biológico complejo (Figura 22).

Estos resultados derivados de Gene Ontology y redes funcionales sugieren que las diferencias transcriptómicas entre endofenotipos se organizan en procesos biológicos coherentes, estructurados en torno a ejes inflamatorios, metabólicos y de regulación génica. Aunque estos enfoques permiten identificar procesos y su interconexión, no delimita de manera específica las rutas celulares en las que convergen estos cambios. Por esta razón, el análisis de enriquecimiento mediante KEGG permite integrar estos resultados dentro de vías moleculares definidas, aportando un marco funcional más concreto para interpretar cómo las alteraciones observadas se traducen en procesos biológicos específicos.

El análisis basado en rutas KEGG confirma la organización descrita previamente, estructurada en tres programas biológicos, y aporta resolución a nivel de mecanismos moleculares (figura 23).

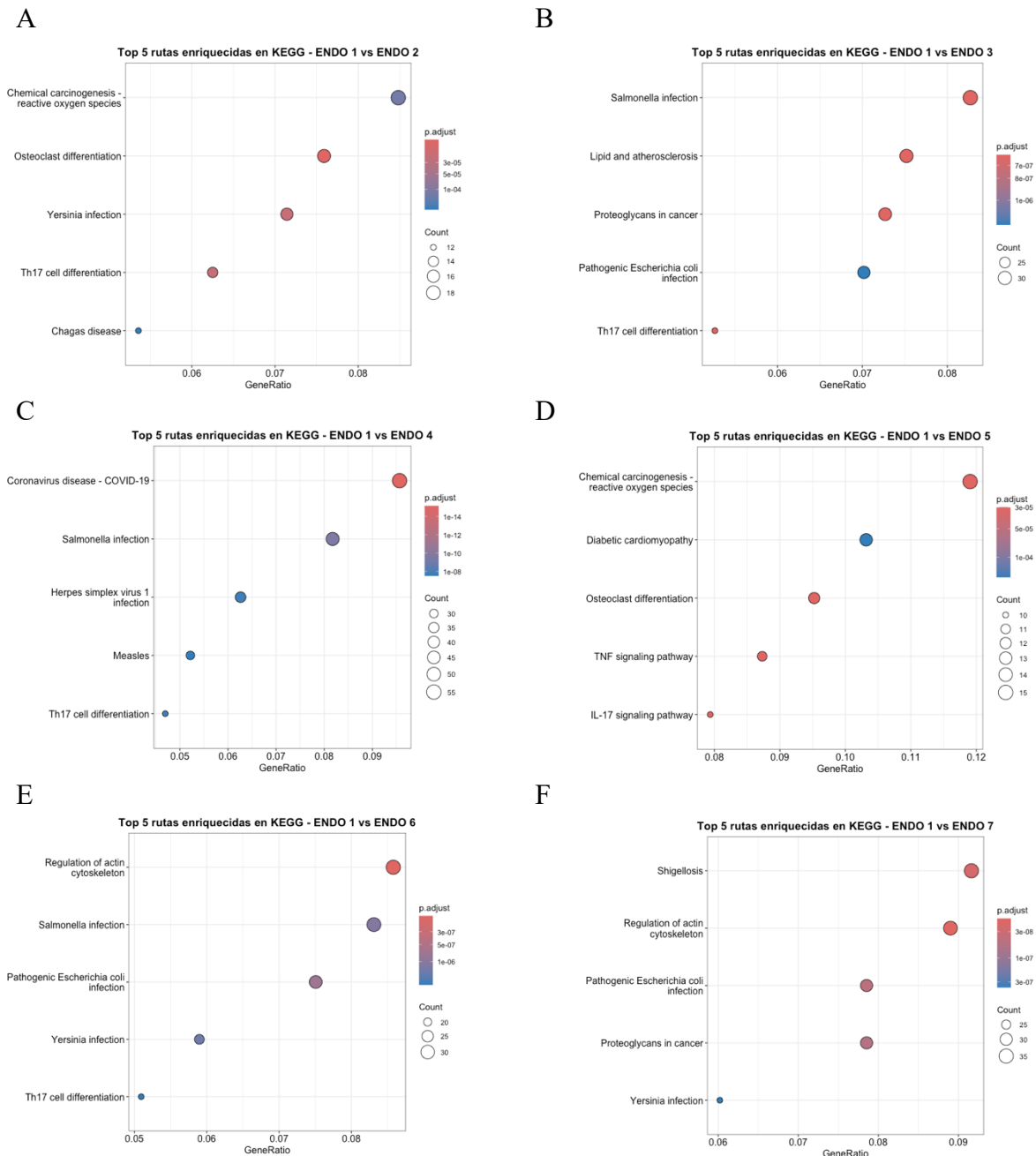


Figura 23. Enriquecimiento de rutas KEGG de los genes diferencialmente expresados en las comparaciones ENDO_2-ENDO_7 frente a ENDO_1. Cada gráfico muestra las principales rutas enriquecidas, donde el eje X representa el GeneRatio, el tamaño de los puntos indica el número de genes asociados (Count) y el color corresponde a la significancia ajustada (p.adjust).

En el primer subgrupo (ENDO_2 y ENDO_5), el eje común este asociado a estrés oxidativo y metabolismo redox, definido por el enriquecimiento en rutas relacionadas con especies reactivas de oxígeno. Individualmente, ENDO_5 presenta una mayor activación inflamatoria mediada por citocinas, mientras que ENDO_2 se orienta hacia procesos de diferenciación inmunológica, sugiriendo un estado menos efector.

El segundo subgrupo (ENDO_3, ENDO_4 y ENDO_6) se caracteriza por una activación inmunológica estructurada, dominada por la diferenciación de células Th17 y la activación

de módulos inmunes conservados. Dentro de este eje, ENDO_4 exhibe una respuesta más amplia, ENDO_6 incorpora componentes de reorganización del citoesqueleto, y ENDO_3 mantiene un perfil más centrado en la respuesta adaptativa.

A su vez, ENDO_7 conserva un perfil independiente, dominado por procesos de migración celular, interacción con el microambiente y reorganización estructural, consistente con un estado efector.

De manera global, el enriquecimiento funcional basados en Gene Ontology, redes y KEGG evidencian que las diferencias transcriptómicas entre endofenotipos no responden a un patrón aleatorio ni aislado, sino que reflejan programas biológicos coherentes y reproducibles. La coincidencia entre estos enfoques confirma la existencia de tres ejes funcionales principales —inflamatorio, metabólico y de regulación génica— que estructuran la heterogeneidad observada en esta cohorte. La consistencia entre comparaciones y niveles de análisis respalda la robustez de los hallazgos y sugiere que los perfiles transcriptómicos en PBMC reflejan estados biológicos definidos, capaces de capturar dimensiones funcionales de la enfermedad que no son evidentes mediante la clasificación serológica convencional.

En consecuencia, la clasificación basada exclusivamente en marcadores serológicos no captura de manera completa la diversidad molecular de la cohorte. La diversidad de patrones funcionales identificados plantea la necesidad de explorar una reclasificación de las muestras basada en criterios transcriptómicos, que permita agrupar a los pacientes según patrones biológicos compartidos y no únicamente por su perfil serológico, con el objetivo de mejorar la interpretación de la heterogeneidad y su posible relevancia clínica.

7.9. Reclasificación transcriptómica de las muestras basada en perfiles de expresión génica.

Los resultados de expresión diferencial y enriquecimiento funcional evidenciaron patrones biológicos compartidos entre subconjuntos de endofenotipos, lo que motivó la exploración de la estructura global del transcriptoma mediante un enfoque no supervisado. Para ello, se analizaron las relaciones de similitud entre muestras a partir de sus perfiles de expresión génica, permitiendo identificar agrupaciones basadas en patrones transcriptómicos no completamente explicados por la clasificación serológica. A partir de esta organización, se definieron clústeres transcriptómicos que reflejan la estructura biológica subyacente de la cohorte, dando lugar a una reclasificación de las muestras independiente de los endofenotipos iniciales (Tabla 4). La descripción detallada de las muestras reasignadas a cada endofenotipo se describe en los anexos (anexo 3).

Clúster	ENDO_1	ENDO_2	ENDO_3	ENDO_4	ENDO_5	ENDO_6	ENDO_7	ENDO_8
CLUS_1	3	1	1	1	1	0	0	1
CLUS_2	0	2	1	0	0	2	2	0
CLUS_3	1	2	2	3	2	3	1	3
CLUS_4	1	0	1	1	2	0	2	0

Tabla 4. Distribución de las muestras por clúster transcriptómico (CLUS_1–CLUS_4) y endofenotipo serológico. Los valores indican el número de muestras compartidas entre cada clúster y endofenotipo. Los valores indican el número de muestras asignadas a cada combinación clúster–endofenotipo.

La distribución de las muestras en estos clústeres evidencia una reorganización sustancial respecto a la clasificación serológica, donde un mismo endofenotipo se fragmenta en múltiples clústeres y, a su vez, cada clúster integra muestras de distintos endofenotipos. En particular, CLUS_3 concentra el mayor número de muestras (n=17) e incluye representantes de todos los endofenotipos, lo que sugiere la presencia de un programa transcriptómico ampliamente compartido.

Para evaluar la coherencia de esta reclasificación, se realizó un PCA utilizando los clústeres transcriptómicos definidos (CLUS_1–CLUS_4). Como se observa en la Figura 24, la distribución de las muestras evidencia una organización más estructurada en el espacio transcriptómico, con una separación más consistente entre grupos en comparación con el PCA graficado con la caracterización inicial.

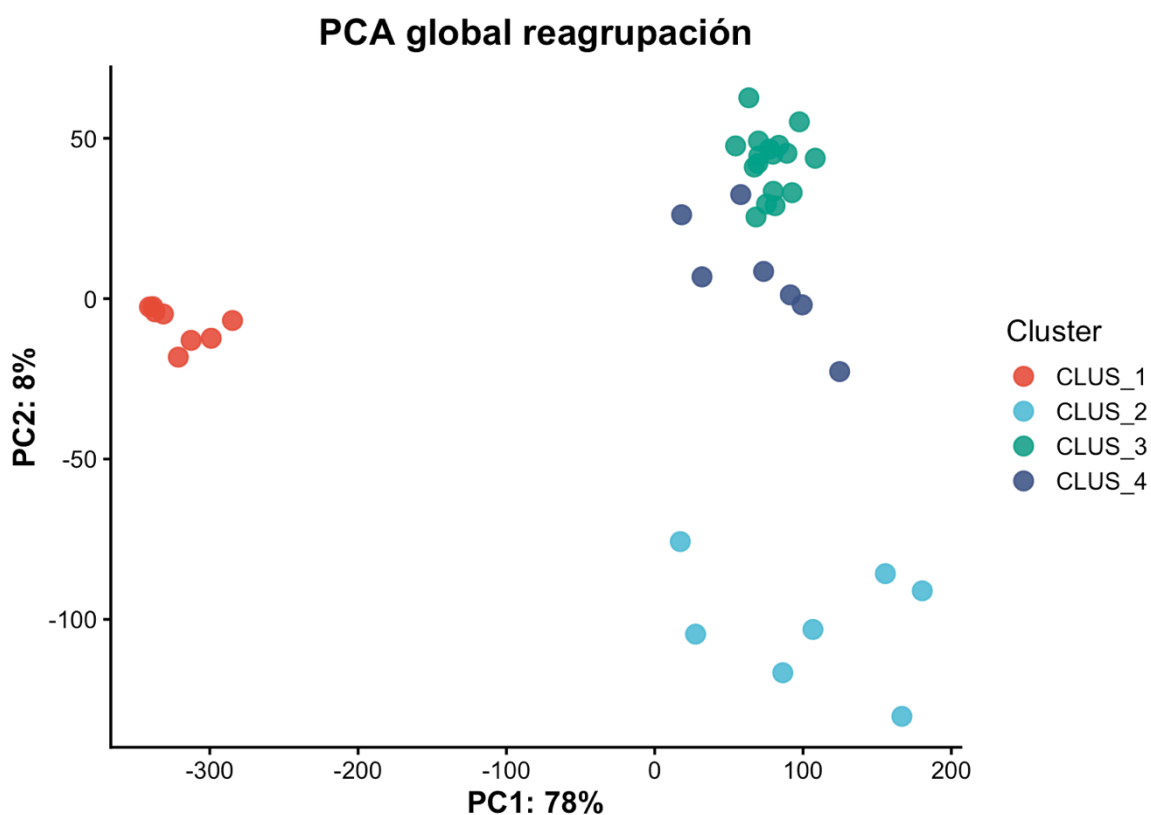


Figura 24. Análisis de componentes principales (PCA) de las muestras basado en datos de expresión génica transformados mediante VST, tras la reclasificación en clústeres transcriptómicos (CLUS_1–CLUS_4). Cada punto representa una muestra individual, coloreada según su asignación a clúster.

CLUS_1 se encuentra completamente segregado hacia valores negativos de PC1, reflejando un perfil transcriptómico claramente diferenciado. Por su parte, CLUS_2, CLUS_3 y CLUS_4 se agrupan en el extremo positivo de este componente, separándose a lo largo de PC2 (8%). CLUS_3 presenta una distribución compacta, indicativa de alta homogeneidad, CLUS_2 muestra mayor dispersión y CLUS_4 ocupa una posición intermedia con cierto solapamiento. Este patrón indica que el PC1 está dominado por un programa transcriptómico específico que distingue a CLUS_1 del resto, mientras que las diferencias entre CLUS_2,

CLUS_3 y CLUS_4 responden a variaciones más sutiles capturadas por PC2, lo que sugiere la existencia de subestados dentro de un mismo marco biológico general.

Tras evidenciar la separación global mediante el PCA y la redistribución de las muestras entre clústeres, se generó un heatmap basado en los 2000 genes más variables utilizando datos transformados mediante VST (Figura 25), con el fin de evaluar la consistencia de la agrupación desde una perspectiva no supervisada y a nivel global de expresión génica.

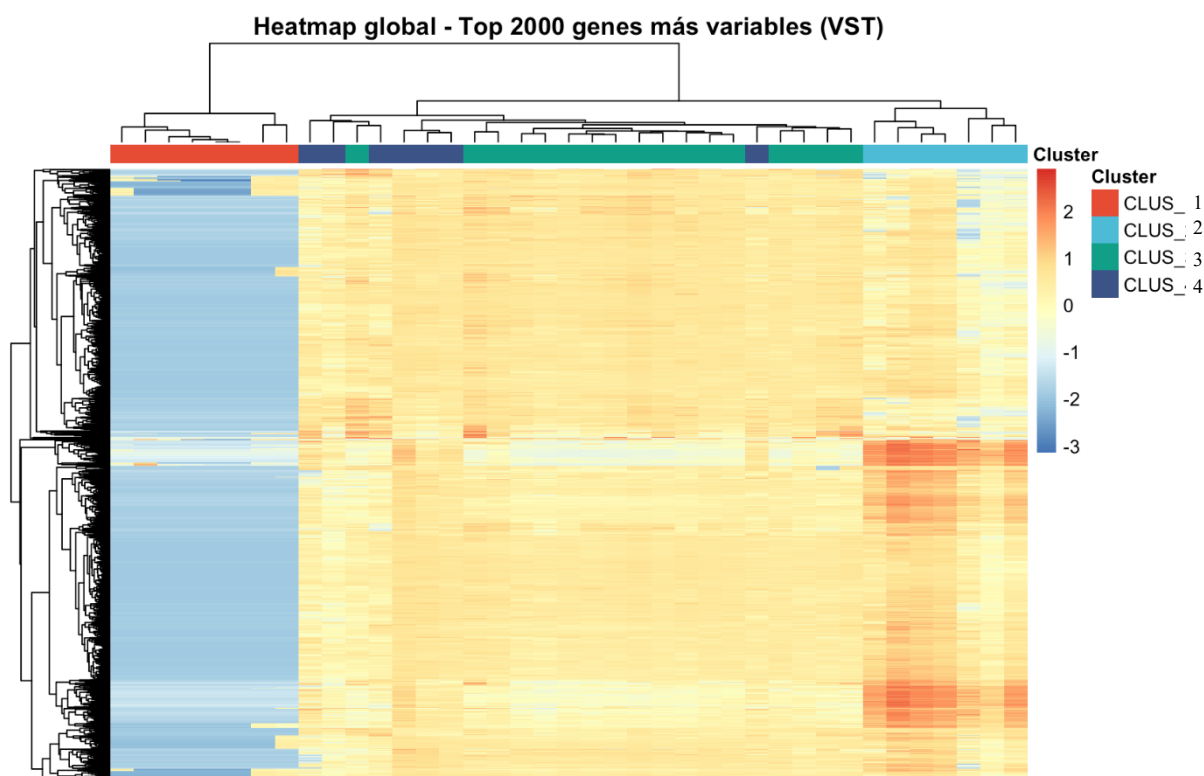


Figura 25. Heatmap global de expresión génica basado en los 2000 genes más variables tras transformación VST de los datos de RNA-Seq. Cada columna representa una muestra y cada fila un gen. Los valores están escalados por fila (z-score), donde los colores indican niveles relativos de expresión (alto en rojo, bajo en azul). La anotación superior corresponde a la asignación de las muestras en clústeres transcriptómicos (CLUS_1-CLUS_4). El dendrograma refleja la similitud jerárquica entre muestras, evidenciando la organización estructural de los perfiles transcriptómicos y la coherencia interna de los clústeres definidos.

El heatmap muestra una organización coherente con la reclasificación propuesta, evidenciada por bloques definidos de expresión génica. CLUS_1 presenta un patrón claramente diferenciado, en concordancia con su separación en el PCA, mientras que CLUS_2, CLUS_3 y CLUS_4 mantienen mayor proximidad entre sí, aunque conservan perfiles internos distinguibles. Además, se identifican módulos de coexpresión específicos por clúster, lo que sugiere la activación de programas biológicos diferenciados y refuerza la hipótesis de que la variabilidad observada responde a estados transcriptómicos organizados.

Con el fin de evaluar de manera más precisa la separación entre los grupos definidos, se realizaron PCAs por pares entre los clústeres (CLUS_1-CLUS_4), considerando todas las combinaciones posibles sin repetición (Figura 26).

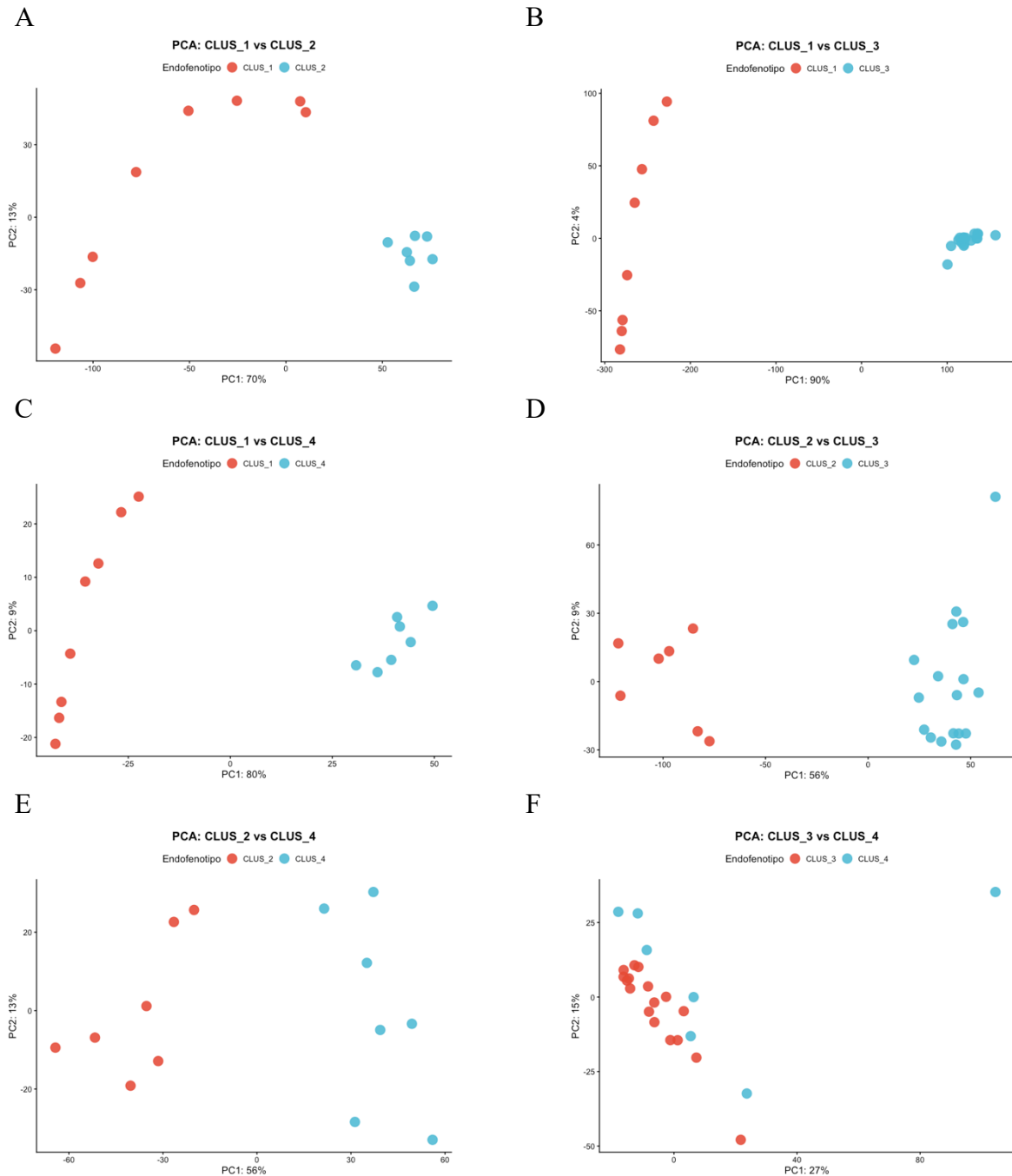


Figura 26. Análisis de componentes principales (PCA) por pares entre los clusters transcriptómicos reclasificados (CLUS_1–CLUS_4). Cada gráfico representa una comparación entre dos clusters, considerando todas las combinaciones posibles sin repetición. Cada punto corresponde a una muestra individual, coloreada según su asignación de cluster. Los ejes PC1 y PC2 indican el porcentaje de varianza explicada en cada comparación, evidenciando el grado de separación transcriptómica entre los grupos analizados.

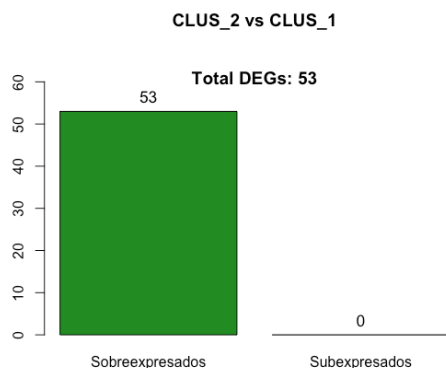
Las comparaciones por pares evidencian una separación consistente entre los clusters. En particular, las comparaciones que incluyen a CLUS_1 muestran una separación robusta frente a todos los demás clusters, explicada principalmente por el PC1, lo que sugiere la presencia de un perfil transcriptómico diferencial que lo distingue claramente del resto.

Por su parte, las comparaciones entre CLUS_2, CLUS_3 y CLUS_4 muestran una menor distancia relativa, con separaciones más moderadas, lo que indica una mayor proximidad biológica entre estos grupos. De forma específica, CLUS_3 y CLUS_4 se encuentran notablemente cercanos entre sí, lo cual es coherente con el PCA global y con una alta similitud en sus perfiles transcriptómicos. Esta proximidad sugiere que ambos clústeres comparten un componente biológico común, diferenciándose por variaciones más sutiles.

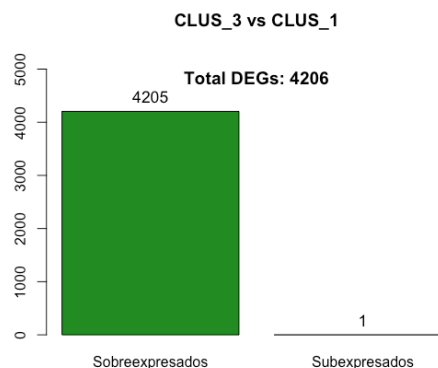
Estos resultados respaldan la reclasificación propuesta, evidenciando una estructura jerárquica en la organización transcriptómica, con un clúster claramente diferenciado (CLUS_1) y un bloque de clústeres relacionados pero distinguibles (CLUS_2–CLUS_4). A partir de esta organización, se realizó el análisis de expresión diferencial para identificar los genes que sustentan estas diferencias y aportar una base molecular a la separación observada.

El análisis de expresión diferencial entre clústeres mostro un patrón altamente heterogéneo en la magnitud y dirección de las diferencias transcriptómicas (Figura 27). Las comparaciones que involucran a CLUS_1 evidencian resultados asimétricos: CLUS_2 vs CLUS_1 presenta un número reducido de DEGs (n=53), mientras que CLUS_3 vs CLUS_1 muestra una diferencia muy marcada, con un alto número de DEGs (n=4206), lo que sugiere que CLUS_3 representa el estado transcriptómico distintivo respecto a CLUS_1. En una posición intermedia se ubica CLUS_4 vs CLUS_1, con 310 DEGs, lo que indica una separación clara pero menos pronunciada.

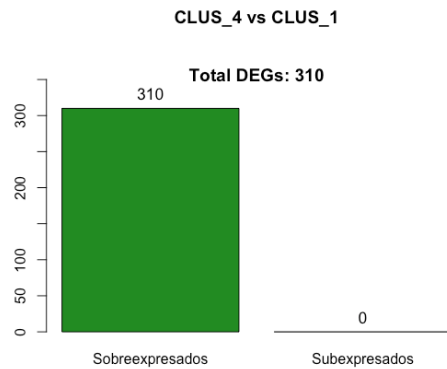
A



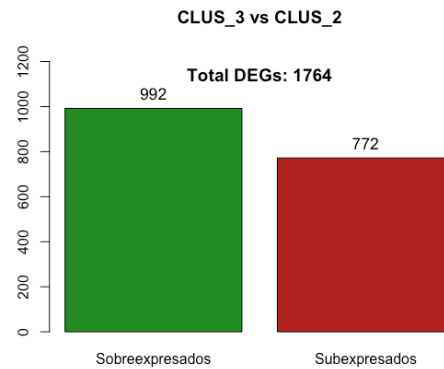
B



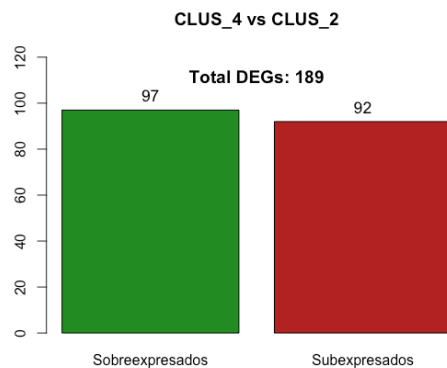
C



D



E



F

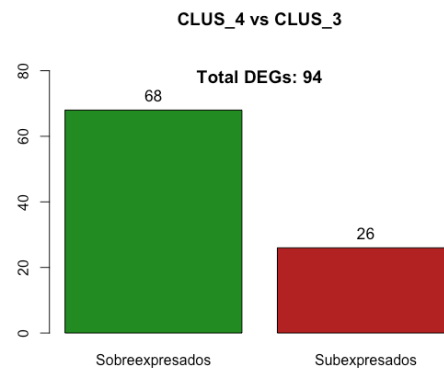


Figura 27. Número de genes diferencialmente expresados (DEGs) en las comparaciones por pares entre los clusters transcriptómicos (CLUS_1–CLUS_4). Cada gráfico muestra el total de genes significativamente diferencialmente expresados, diferenciando entre genes sobreexpresados (verde) y subexpresados (rojo) en la primera condición de cada comparación. Estos resultados reflejan la magnitud y dirección de los cambios transcriptómicos entre clusters.

Un aspecto relevante es el predominio casi absoluto de genes sobreexpresados en todas las comparaciones que incluyen CLUS_1, con ausencia o mínima presencia de genes subexpresados. Este patrón en los DEGs sugiere que CLUS_1 no solo es distinto, sino que presenta un perfil transcriptómico más restringido en comparación con los demás clústeres.

Por otro lado, las comparaciones entre CLUS_2, CLUS_3 y CLUS_4 muestran un comportamiento más balanceado. CLUS_3 vs CLUS_2 ($n=1764$) y CLUS_4 vs CLUS_2 ($n=189$) presentan proporciones más equilibradas entre genes sobreexpresados y subexpresados, lo que sugiere relaciones transcriptómicas más complejas y bidireccionales. Finalmente, CLUS_4 vs CLUS_3 ($n=94$) muestra una menor magnitud de cambio, consistente con la proximidad observada entre estos clústeres en el PCA.

Con este propósito, se generaron volcano plots para cada comparación entre clústeres (Figura 28), con el fin de visualizar de manera integrada la magnitud y la significancia de los cambios en la expresión génica.

Figura 28. Volcano plots de expresión génica diferencial para las comparaciones entre clusters (CLUS_1–CLUS_4). Cada punto representa un gen, donde el eje X corresponde al log2 fold change (log2FC) y el eje Y a la significancia estadística expresada como $-\log_{10}$ del valor p ajustado. Los puntos en rojo indican genes diferencialmente expresados con significancia estadística y cambio de expresión relevante, mientras que los puntos en gris representan genes no significativos. Las líneas punteadas delimitan los umbrales de significancia y magnitud de cambio. Estos gráficos permiten visualizar la magnitud y dirección de los cambios transcriptómicos entre los clusters evaluados.

Los volcano plots evidencian contrastes claros entre clústeres en términos de magnitud y significancia de los cambios transcriptómicos. Las comparaciones frente a CLUS_1, particularmente con CLUS_3 y CLUS_4, muestran un desplazamiento marcado hacia valores positivos de log2FC acompañado de alta significancia, indicando una activación transcriptómica global en estos grupos. Por su parte, CLUS_2 vs CLUS_1 presenta un perfil restringido, con un número reducido de genes modulados, consistente con su proximidad transcriptómica.

Entre CLUS_2, CLUS_3 y CLUS_4, los patrones son más equilibrados y heterogéneos. Particularmente, CLUS_3 vs CLUS_2 concentra un número considerable de genes significativos distribuidos en ambos sentidos del cambio, lo que evidencia diferencias relevantes a pesar de su cercanía en el espacio transcriptómico. Finalmente, CLUS_4 vs CLUS_3 muestra una menor magnitud global de cambio, en concordancia con su alta similitud.

Para identificar los genes que sustentan estas diferencias y evaluar si existe recurrencia entre comparaciones, se construyó un ranking de los 10 genes diferencialmente expresados más significativos en cada contraste (Figura 29). Este análisis permite trascender la magnitud global de los DEGs e identificar los genes específicos que estructuran la heterogeneidad entre clústeres.

Top 10 genes diferencialmente expresados por comparación (clusters)						
Ordenados por padj y log2FC						
	CLUS_2 vs CLUS_1	CLUS_2 vs CLUS_4	CLUS_3 vs CLUS_1	CLUS_3 vs CLUS_2	CLUS_3 vs CLUS_4	CLUS_4 vs CLUS_1
1	NEAT1	RPL3	RN7SK	ESR1	LOC10537936	RN7SK
2	SETBP1	ARHGDI	NEAT1	RFC3	IGSF5	NEAT1
3	RN7SK	FTL	AHNAK	LOC10537936	LOC10272363	AHNAK
4	IMMP2L	ESR1	FLNA	GNA14	LOC105379064	SRRM2
5	PDE4D	LOC105371378	RN7SL1	LOC105370504	SLC30A8	RNF213
6	KCNQ1OT1	CEP63	SRRM2	LOC112268061	LOC102723670	PABPC1
7	TRMT11	ADAMTS17	EEF1A1	TESMIN	MYH2	SYNE2
8	SNTB1	TTC28	RNF213	MYH2	SCN11A	DDX5
9	CHD6	RPL13A	PABPC1	IGSF5	CTNNA2	FLNA
10	HIPK2	HLA-DRA_1	HIVEP2	LOC10272363	ARGFX	KMT2D

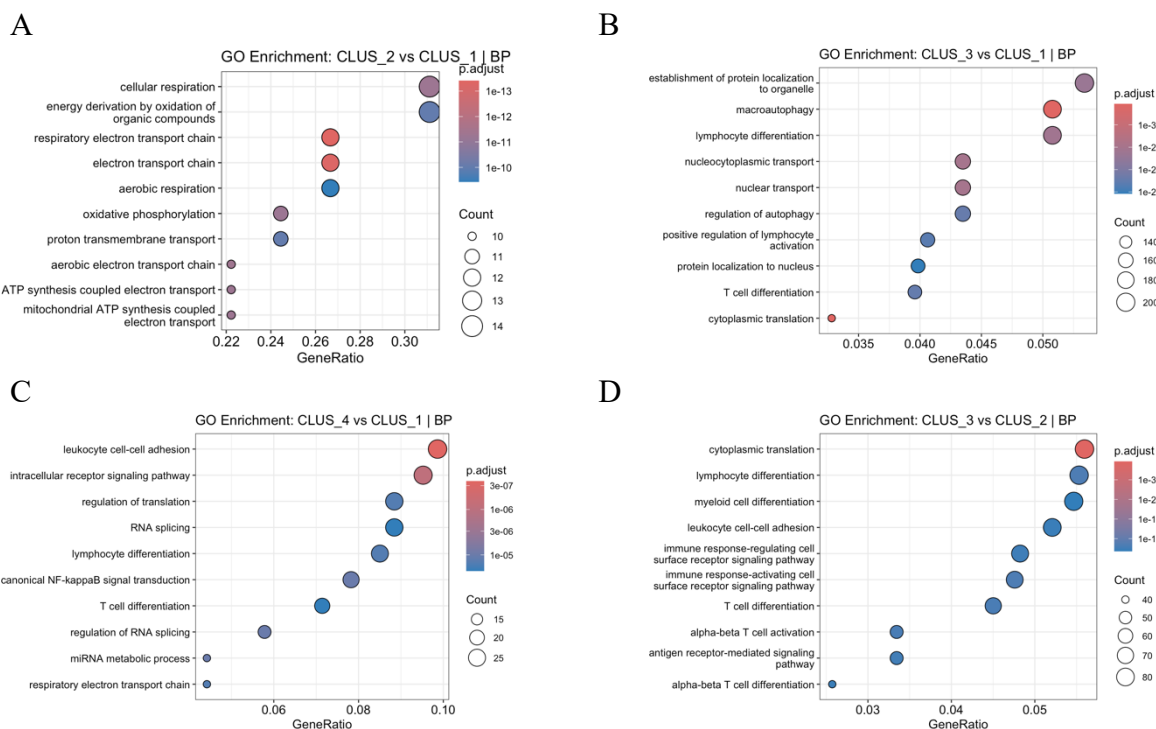
Figura 29. Ranking de los 10 genes diferencialmente expresados más significativos en las comparaciones por pares entre los clústeres transcriptómicos (CLUS_1–CLUS_4). Los genes se ordenaron según su significancia estadística (padj) y magnitud de cambio

(log2 fold change). Las celdas resaltadas indican genes compartidos entre múltiples comparaciones, evidenciando patrones transcriptómicos recurrentes.

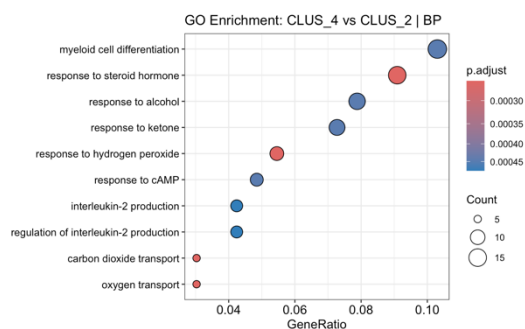
Este ranking de genes revela la recurrencia de genes asociados a regulación transcripcional y organización nuclear como NEAT1 y RN7SK, así como elementos implicados en procesamiento de ARN (SRRM2) y arquitectura celular (FLNA, PABPC1). Estos genes aparecen de forma consistente en múltiples comparaciones, especialmente en aquellas que involucran CLUS_3, lo que sugiere que las muestras agrupadas en este clúster tienen una mayor reorganización del control transcripcional. Las comparaciones con menor divergencia, como CLUS_4 vs CLUS_3, presentan menor solapamiento y menor magnitud de cambio, en concordancia con los resultados previos.

Los genes con mayor magnitud de cambio incluyen, además, reguladores del control transcripcional y organización nuclear, como SETBP1 y ESR1, así como genes asociados a señalización celular y respuesta inflamatoria, como PDE4D, HIVEP2 e IGSF5; componentes relacionados con metabolismo y estrés celular, como IMMP2L y FTL; y genes implicados en organización estructural y arquitectura celular, como SYNE2 y MYH2. Este patrón sugiere que las diferencias entre clústeres responden a la activación diferencial de módulos funcionales que integran regulación génica, adaptación metabólica y reorganización celular.

La recurrencia de estas funciones a nivel génico indica que las diferencias entre clústeres responden a programas biológicos organizados. Para evaluar de manera sistemática, se realizó un análisis de enriquecimiento en Ontología Génica para la categoría de procesos biológicos (Figura 30).



E



F

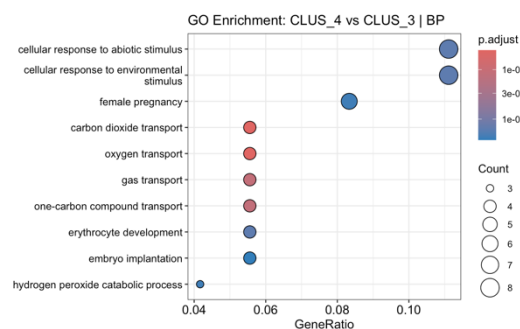


Figura 30. Enriquecimiento funcional basado en Gene Ontology (proceso biológico, BP) para las comparaciones entre clústers (CLUS_1–CLUS_4). Cada gráfico muestra los principales procesos biológicos enriquecidos, donde el eje X representa el GeneRatio, el tamaño de los puntos corresponde al número de genes asociados (Count) y el color indica la significancia estadística ajustada (p.adjust). Este análisis permite identificar los programas biológicos predominantes que subyacen a las diferencias transcriptómicas entre los clústers evaluados.

Las comparaciones evidencian una organización funcional compartida entre clústeres, más que perfiles independientes. La comparación CLUS_2 vs CLUS_1 está dominado por procesos mitocondriales, lo que sugiere un estado centrado en el metabolismo energético. Por su parte CLUS_3 muestra un desplazamiento hacia la activación inmunológica, con enriquecimiento en diferenciación linfocitaria, activación de células T y señalización mediada por receptores. A ello se suman procesos de transporte intracelular y autofagia, lo que sugiere un estado funcional más activo. La comparación CLUS_3 vs CLUS_2 refuerza esta transición desde un estado metabólico hacia uno inmunológicamente estructurado.

El CLUS_4 presenta un perfil intermedio que integra ambos ejes, con procesos de adhesión leucocitaria, señalización intracelular y activación de NF- κ B, junto con mecanismos de procesamiento de ARN y regulación transcripcional. Frente al CLUS_2, incorpora además respuestas a estímulos químicos y estrés oxidativo, lo que refleja una mayor complejidad funcional.

Estos resultados sugieren una progresión funcional que va desde un estado dominado por metabolismo energético (CLUS_2), una activación inmunológica estructurada (CLUS_3), hasta un estado más complejo que integra regulación génica y adaptación a estímulos (CLUS_4), sugiriendo distintos niveles de especialización funcional entre los clústeres. A partir de esta caracterización, se evaluó la organización de los procesos biológicos mediante redes funcionales (Figura 31), con el fin de analizar su conectividad y modularidad

Las redes funcionales muestran diferencias claras en la organización de los procesos biológicos entre comparaciones. CLUS_2 vs CLUS_1 presenta una estructura compacta y altamente interconectada, dominada por un único módulo metabólico cohesivo centrado en funciones mitocondriales. CLUS_3 vs CLUS_1 muestra una organización más modular, donde coexisten procesos de activación inmunológica, autofagia y transporte intracelular distribuidos en bloques parcialmente conectados, reflejando un estado funcional más diversificado. Esta transición se refuerza en CLUS_3 vs CLUS_2, donde emerge un núcleo inmunológico bien definido, dominado por diferenciación y activación de células T, acompañado de procesos periféricos menos integrados. CLUS_4 vs CLUS_1 presenta una arquitectura intermedia que integra módulos inmunológicos, de señalización y regulación

transcripcional. En contraste, las comparaciones CLUS_4 vs CLUS_2 y CLUS_4 vs CLUS_3 muestran redes menos densas y más fragmentadas, con módulos de baja interconectividad, lo que sugiere respuestas más heterogéneas. Los resultados de CC y MF son presentados en anexos (anexo 4).

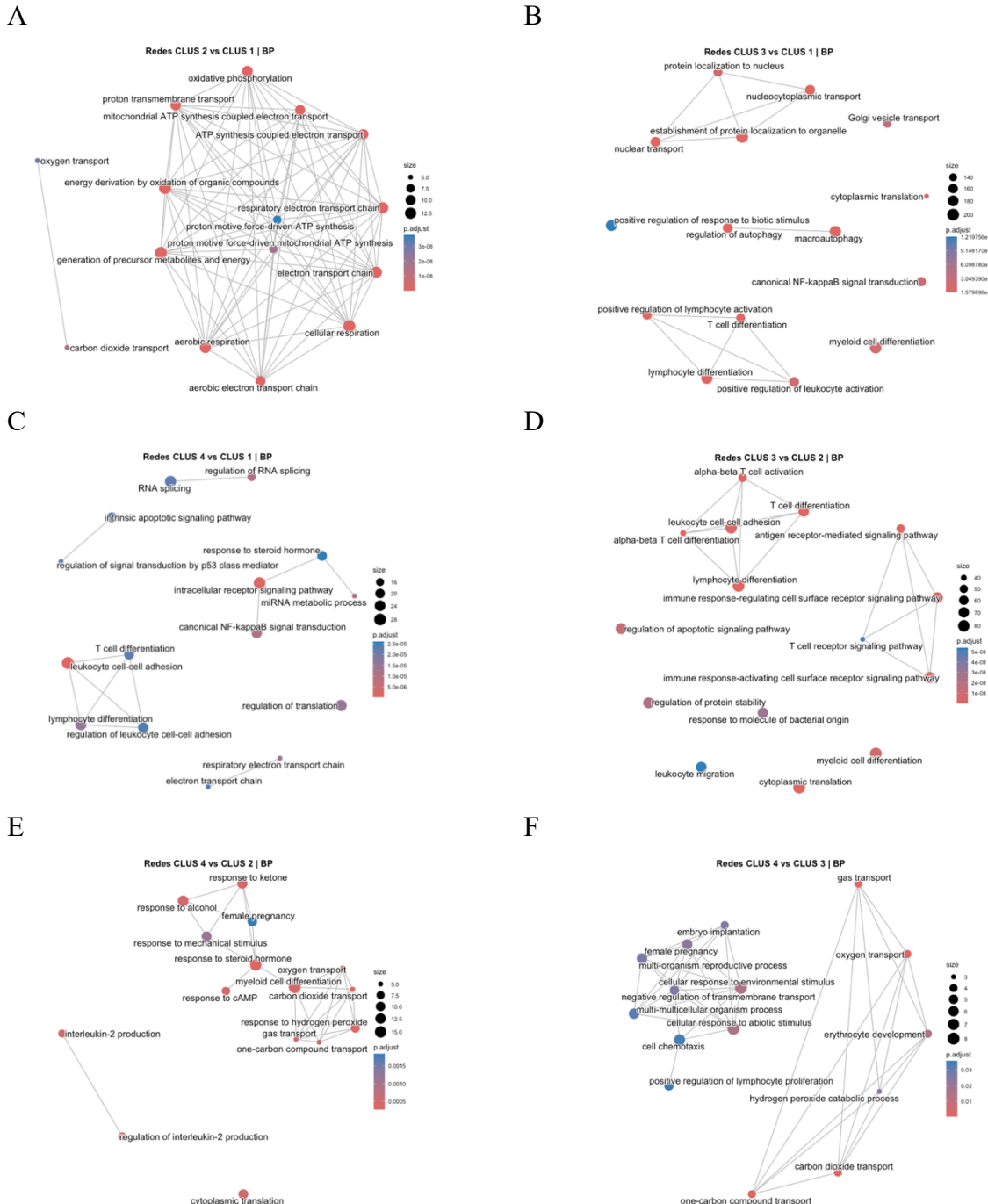


Figura 31. Redes funcionales de enriquecimiento basadas en Gene Ontology para procesos biológicos (Biological Process, BP) en las comparaciones entre clústers transcritómicos (CLUS_1–CLUS_4). Cada nodo representa un proceso biológico enriquecido, cuyo tamaño corresponde al número de genes asociados y cuyo color indica la significancia ajustada (p.adjust). Las conexiones entre nodos reflejan relaciones funcionales entre los términos enriquecidos. Estas redes permiten visualizar la organización estructural de los programas biológicos diferenciales entre clústers, evidenciando módulos funcionales metabólicos, inmunológicos y de respuesta a estímulos sistémicos.

Las diferencias entre clústeres no solo dependen de los procesos implicados, sino de su grado de organización, distinguiendo estados dominados por módulos altamente cohesivos frente a otros más dispersos y multifuncionales. Esta organización sugiere que los cambios transcriptómicos observados se estructuran en programas funcionales definidos, que pueden ser delimitados a nivel de rutas biológicas específicas. Es por esto por lo que el análisis de rutas KEGG permite integrar estos hallazgos dentro de vías moleculares concretas (Figura 32).

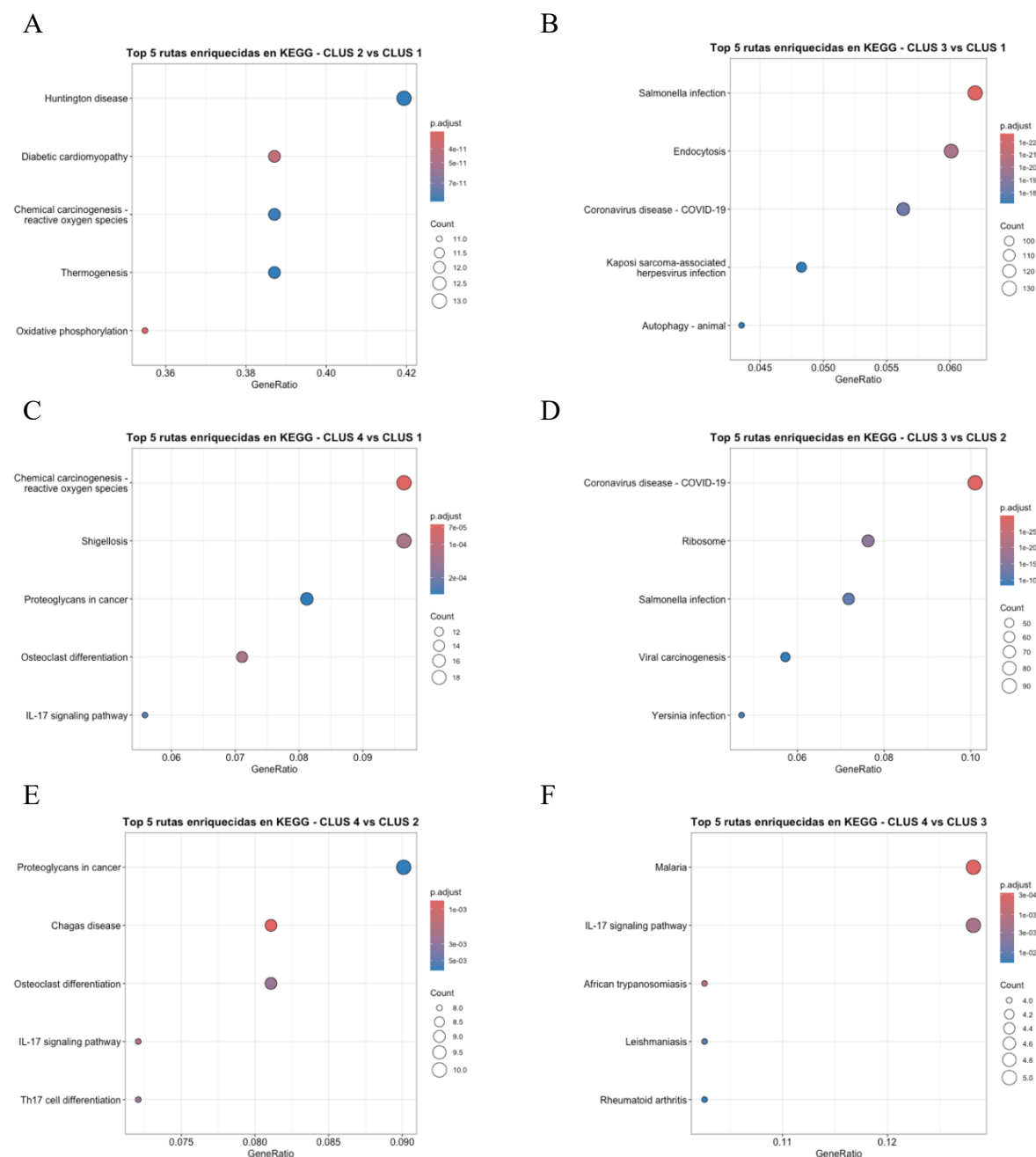


Figura 32. Enriquecimiento de rutas KEGG de los genes diferencialmente expresados en las comparaciones entre clusters transcriptómicos (CLUS_1–CLUS_4). Cada gráfico muestra las principales rutas enriquecidas, donde el eje X representa el GeneRatio, el tamaño de los puntos indica el número de genes asociados (Count) y el color corresponde a la significancia ajustada (p.adjust). Las rutas asociadas a infecciones representan la activación de módulos inmunológicos conservados y no procesos infecciosos activos.

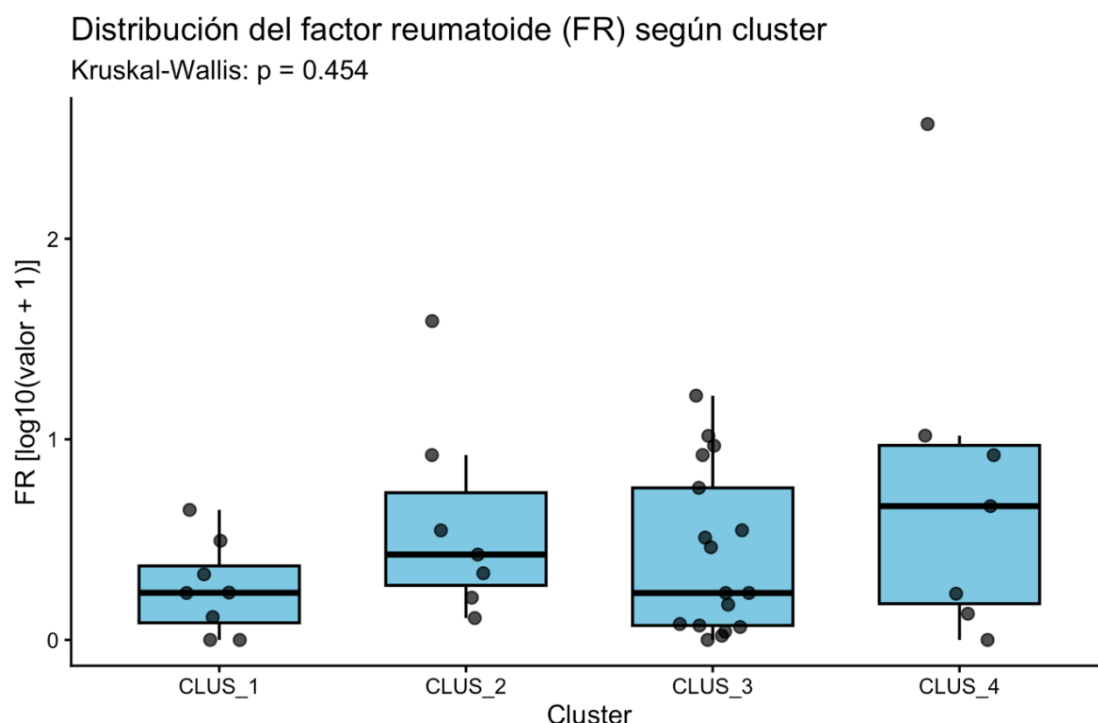
El análisis de rutas KEGG permite precisar la base molecular de cada estado. CLUS_2 vs CLUS_1 mantiene un perfil dominado por metabolismo energético y estrés oxidativo. Este patrón sugiere un estado metabólicamente activo, pero funcionalmente restringido, con predominio de procesos bioenergéticos sobre mecanismos efectores.

Las comparaciones que involucran CLUS_3 reflejan una activación inmunológica estructurada. El enriquecimiento en rutas asociadas a infecciones virales y bacterianas (p. ej., Infección por Salmonella, enfermedad por Coronavirus) y procesos como endocitosis y autofagia indica la activación de módulos inmunes conservados, con capacidad de respuesta y regulación a nivel celular.

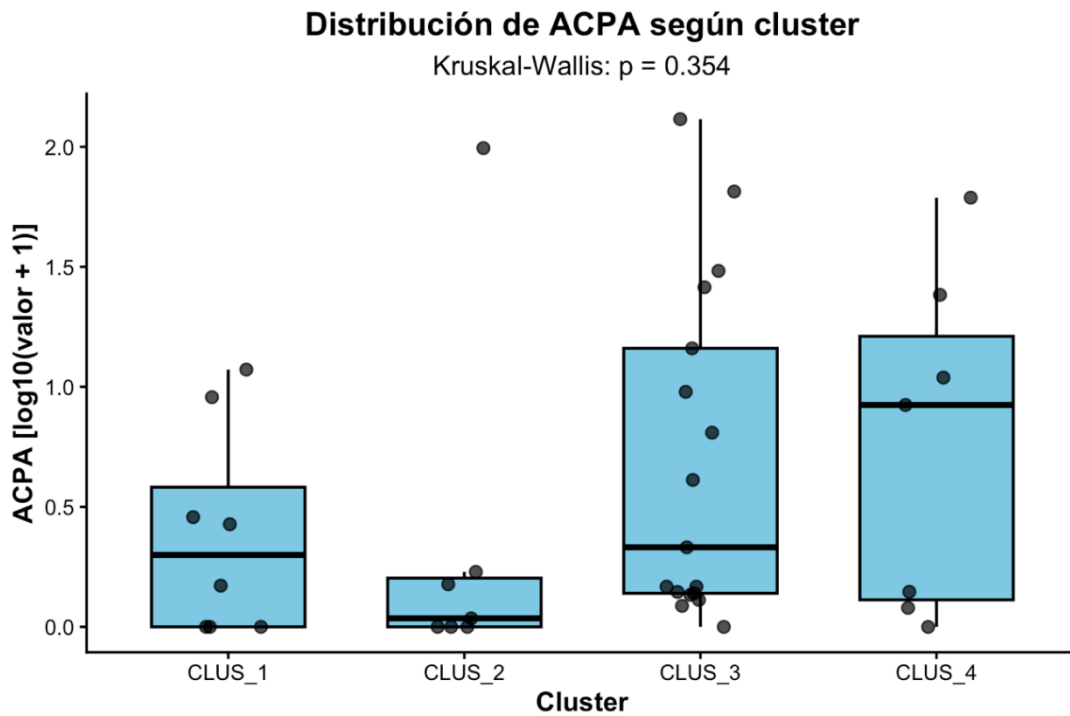
Por otro lado, el CLUS_4 integra señalización inflamatoria más específica, destacando rutas como vía de señalización de la IL-17 y diferenciación de osteoclastos, junto con procesos asociados a estrés oxidativo y remodelación tisular. Este patrón indica una transición hacia un estado más complejo, donde convergen activación inmune, regulación inflamatoria y adaptación al microambiente.

Estos resultados plantean la necesidad de evaluar la correspondencia entre la estratificación transcriptómica y los marcadores clínicos de la cohorte. Con este propósito, se examinó la relación entre los clústeres transcriptómicos y los factores serológicos (FR, ACPA y ANAs), con el objetivo de determinar si la heterogeneidad molecular observada se refleja en la serología o constituye una dimensión biológica independiente de estos indicadores (Figura 33).

A



B



C

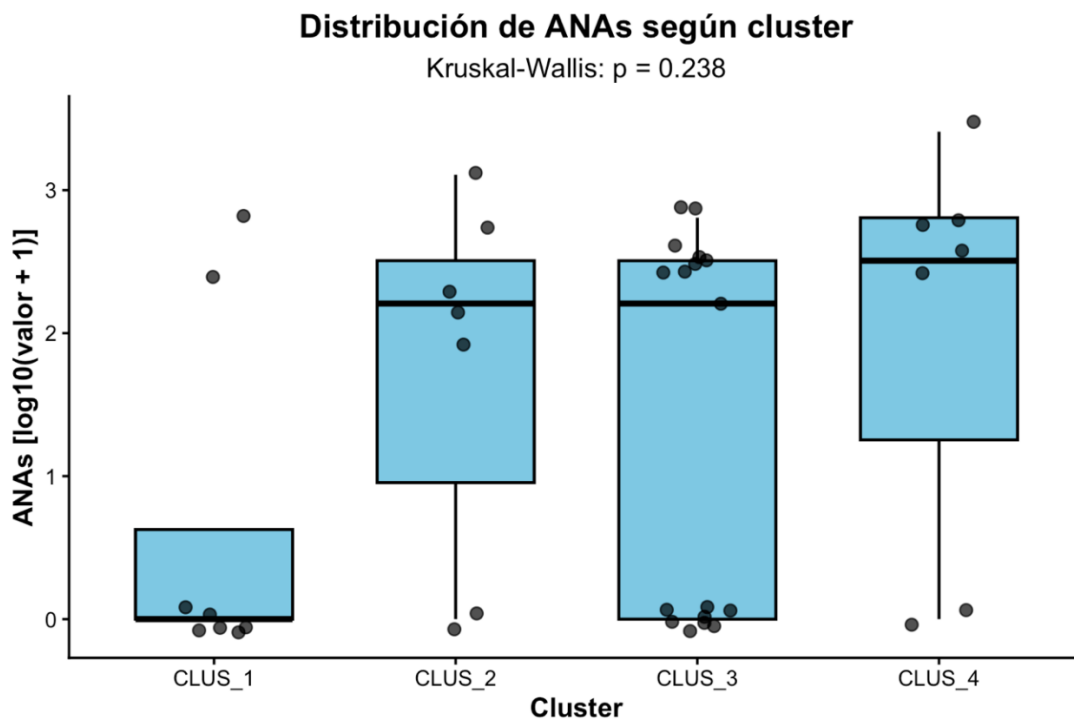


Figura 33. Distribución de FR, ACPA y ANAs según clúster. Se presenta la distribución de los niveles de factor reumatoide (FR), anticuerpos anti-péptidos citrulinados (ACPA) y anticuerpos antinucleares (ANAs) entre los clústeres transcriptómicos. Los valores fueron transformados a $\log_{10}(\text{valor} + 1)$. No se observaron diferencias estadísticamente significativas entre los grupos (prueba de Kruskal-Wallis; FR: $p = 0.454$, ACPA: $p = 0.354$, ANAs: $p = 0.238$), aunque se evidencian tendencias de mayor dispersión en clústeres con perfiles transcriptómicos más complejos.

La integración de las variables serológicas no evidenció diferencias significativas entre los clústeres (FR: $p = 0.454$; ACPA: $p = 0.354$; ANAs: $p = 0.238$), lo que indica que estos marcadores no capturan la variabilidad funcional observada a nivel transcriptómico. Aunque se identifican tendencias —como una mayor dispersión de ACPA en CLUS_3 y CLUS_4, y niveles más elevados de ANAs en estos mismos grupos—, existe una cercanía considerable entre clústeres que limita su capacidad discriminativa.

Estos resultados sugieren que la serología convencional captura solo una dimensión parcial de la respuesta inmune, centrada en la presencia de autoanticuerpos, sin reflejar los estados funcionales activos de las células inmunes. En contraste, la transcriptómica permite identificar programas biológicos dinámicos asociados a activación, regulación y adaptación celular, revelando una heterogeneidad no evidente en la caracterización clínica tradicional. La ausencia de correlación entre ambos niveles indica que los clústeres transcriptómicos representan estados biológicos independientes de los marcadores serológicos clásicos, lo que resalta la necesidad de integrar enfoques ómicos en la estratificación de pacientes con AR.

Resulta entonces pertinente evaluar si esta organización transcriptómica se relaciona con el CDAI. Para ello, se analizó la asociación entre los clústeres definidos y el CDAI, con el objetivo de determinar si los estados moleculares identificados se traducen en diferencias en la severidad y actividad inflamatoria de los pacientes (Figura 34).

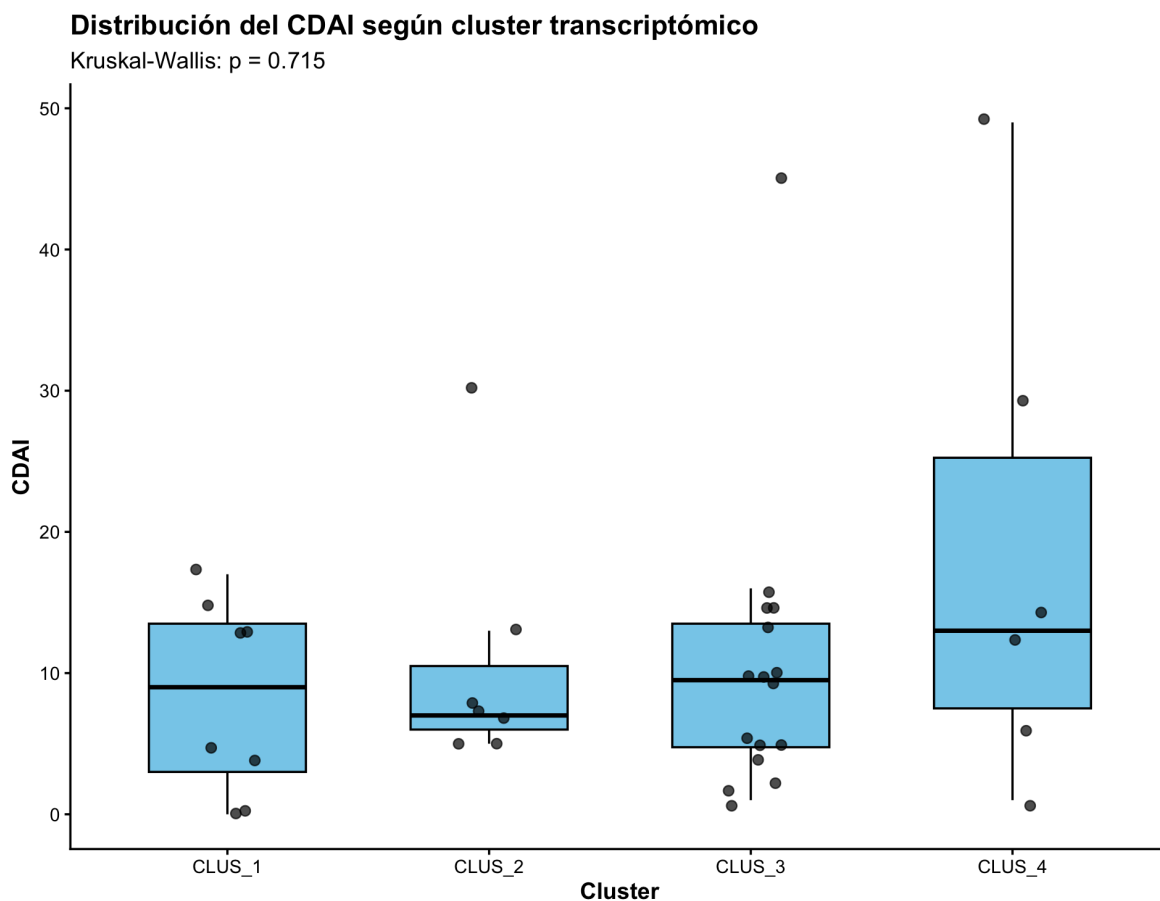


Figura 34. Distribución del índice de actividad clínica de la enfermedad (CDAI) según los clústeres transcriptómicos (CLUS_1–CLUS_4). Cada punto representa un paciente individual y las cajas muestran la mediana y el rango intercuartílico. No se observaron diferencias estadísticamente significativas entre los grupos (Kruskal-Wallis, $p = 0.715$).

La distribución del CDAI no evidenció diferencias significativas entre los clústeres transcriptómicos (Kruskal-Wallis, $p = 0.715$), lo que indica que la actividad clínica global de la enfermedad no se asocia directamente con la estratificación molecular identificada. Sin embargo, a nivel descriptivo se observan patrones relevantes. CLUS_4 presenta la mayor dispersión y los valores más altos de CDAI, incluyendo los casos con mayor actividad clínica, lo que sugiere una tendencia hacia un estado más severo. Por su parte, CLUS_2 muestra una distribución más compacta y con menor variabilidad, consistente con un perfil clínico más homogéneo. CLUS_1 y CLUS_3 presentan distribuciones intermedias, con similitud considerable entre sí.

Este comportamiento indica que, aunque los estados transcriptómicos reflejan diferencias funcionales a nivel molecular, estas no se traducen de manera lineal en la actividad clínica medida por CDAI. Esto sugiere que la actividad de la enfermedad y la transcriptómica capturan dimensiones complementarias, donde la primera refleja la manifestación clínica y la segunda los estados biológicos subyacentes.

Finalmente, la reclasificación transcriptómica permitió identificar una organización estructurada de la cohorte basada en estados funcionales diferenciados, que no se corresponden con la clasificación serológica tradicional ni con la actividad clínica medida por CDAI. Los clústeres definidos reflejan programas biológicos específicos asociados a metabolismo, activación inmunológica y adaptación al microambiente, evidenciando que la heterogeneidad de la AR se manifiesta a múltiples niveles y no puede ser capturada por un único eje de caracterización. Estos hallazgos respaldan la utilidad de la transcriptómica como herramienta para refinar la estratificación de pacientes, aportando una visión más precisa de la biología subyacente de la enfermedad.

8. Discusión

La clasificación serológica de la artritis reumatoide (AR) se ha sustentado históricamente en la medición de factor reumatoide (FR), anticuerpos antiproteínas citrulinadas (ACPA) y, más recientemente, anticuerpos antinucleares (ANAs). Estos marcadores constituyen un pilar en la práctica clínica para el diagnóstico, la estratificación inicial y el seguimiento de los pacientes, debido a su asociación con mecanismos autoinmunes y su valor pronóstico en determinados contextos. Sin embargo, su uso parte de la premisa de que reflejan el estado biológico de la enfermedad, lo cual ha sido cuestionado en estudios recientes que evidencian limitaciones en su capacidad para capturar la heterogeneidad molecular de la AR (67,68).

En esta cohorte, los niveles de FR, ACPA y ANAs no mostraron asociación significativa con la actividad clínica medida mediante CDAI, lo que indica presuntamente que la serología convencional no logra capturar de manera directa la expresión clínica de la enfermedad. Este hallazgo es consistente con reportes que describen una débil correlación entre biomarcadores tradicionales y la actividad inflamatoria o progresión de la AR, sugiriendo que estos marcadores representan únicamente una dimensión parcial del estado inmunológico (68). Lo cual nos indica que la presencia de autoanticuerpos no se traduce necesariamente en una

manifestación clínica proporcional de la enfermedad ni permite discriminar o predecir la severidad de la enfermedad.

Esta limitación se hace más evidente al contrastar la clasificación serológica con el perfil transcriptómico de las PBMC. Aunque los endofenotipos serológicos permitieron una diferenciación parcial entre grupos, esta no se correspondió con la organización global de la expresión génica. El análisis transcriptómico reveló una estructura definida, dominada por un eje principal de variación y subgrupos moleculares independientes de la clasificación clínica. Estudios recientes como el de Chang et al. han demostrado que el análisis transcriptómico permite identificar subtipos inmunológicos en AR que no son capturados por los criterios clínicos convencionales, evidenciando que la organización biológica de la enfermedad trasciende la serología (69).

El solapamiento observado entre los endofenotipos en el espacio transcriptómico refuerza esta idea, mostrando que pacientes con perfiles serológicos distintos pueden compartir estados biológicos similares. Esta discordancia entre la clasificación clínica y la organización molecular ha sido descrita previamente en estudios multi-ómicos, donde se evidencia que la actividad clínica o incluso la respuesta terapéutica no siempre se corresponde con los estados moleculares subyacentes (70). Con estos resultados, sugerimos que la clasificación basada en autoanticuerpos pierde capacidad explicativa de la condición del paciente cuando se compara frente a datos de alta resolución molecular.

Un hallazgo central en esta cohorte fue la heterogeneidad intragrupo, evidenciada por la presencia de muestras con perfiles transcriptómicos discordantes respecto a su endofenotipo. Esta variabilidad no corresponde a ruido experimental, sino a una característica estructural de la cohorte. En sistemas inmunológicos, la variabilidad en la expresión génica es una propiedad inherente que refleja diferencias en estados de activación celular y regulación dinámica de la respuesta inmune, incluso dentro de grupos clínicamente homogéneos, lo que ha sido ampliamente descrito en estudios de transcriptómica a nivel poblacional y celular (71). Su impacto fue evidente en la ausencia inicial de genes diferencialmente expresados (DEGs), lo que podría interpretarse como homogeneidad biológica. Sin embargo, tras la depuración del conjunto de datos, emergieron diferencias transcriptómicas robustas, indicando que dichas señales estaban enmascaradas por la variabilidad interna.

La reanálisis tras retirar las muestras con perfiles transcriptómicos discordantes intragrupo permitió una separación más clara entre los endofenotipos y aumentó la detección de genes diferencialmente expresados y rutas biológicas asociadas, evidenciando que la heterogeneidad condiciona la capacidad de identificar diferencias transcriptómicas relevantes. Este fenómeno resalta una limitación frecuente en estudios de expresión génica en enfermedades complejas, donde la variabilidad intragrupo puede reducir la potencia estadística y ocultar patrones biológicos relevantes (72,73).

Tras la depuración, el endofenotipo 1 emergió como un estado molecular claramente diferenciado. Su separación consistente frente a los demás grupos sugiere la existencia de un perfil transcriptómico estable. Este comportamiento podría estar relacionado con su condición serológica triple negativa, asociada a una menor activación de mecanismos inmunes vinculados a la producción de autoanticuerpos. Este tipo de subgrupos ha sido

descrito en la literatura como entidades con características moleculares particulares, distintas de los fenotipos clásicos seropositivos (68).

Las diferencias observadas no se limitaron a genes individuales, sino que involucraron una fracción amplia del transcriptoma. El predominio de genes sobreexpresados en los grupos comparados frente al endofenotipo 1 sugiere la activación diferencial de programas biológicos completos. Este tipo de reorganización transcriptómica ha sido reportado en estudios de RNA-seq en AR, donde se identifican cambios coordinados en rutas inmunológicas y metabólicas más que alteraciones puntuales (74).

El análisis de enriquecimiento funcional permitió organizar estas diferencias en tres ejes principales: metabolismo mitocondrial, regulación transcripcional y activación inmuno-inflamatoria. Estudios realizados por Chang et al. describen la interacción entre estos procesos como un componente central en la fisiopatología de la AR, donde el metabolismo celular y la regulación génica modulan la activación del sistema inmune (69).

A partir de estos resultados, la reclasificación transcriptómica reorganizó la cohorte en clústeres basados en similitud de expresión génica. Esta aproximación evidenció que los endofenotipos serológicos no corresponden a agrupaciones moleculares homogéneas, sino que se fragmentan en múltiples estados, mientras que cada clúster integra muestras de distintos perfiles serológicos. Este tipo de reorganización ha sido reportado por Chang et al. donde se reportó que la estratificación basada en transcriptómica permite identificar subtipos biológicos más consistentes que los definidos por criterios clínicos (69).

El análisis comparativo entre clústeres evidenció estados transcriptómicos diferenciados con distintos niveles de activación funcional, sugiriendo una transición desde perfiles dominados por metabolismo energético hacia estados de activación inmunológica más complejos. Este gradiente funcional integra transcriptómica y rutas biológicas, indicando que la AR se organiza en estados dinámicos del sistema inmune (69).

Un aspecto relevante fue la ausencia de asociación entre los clústeres transcriptómicos, la serología y el CDAI. Este resultado indica que la transcriptómica captura una dimensión biológica independiente, no reflejada por los marcadores clínicos tradicionales. Estudios realizados por Tasaki et al. han demostrado que la expresión génica en sangre periférica puede revelar estados inmunológicos activos incluso en ausencia de correlación clínica directa, reforzando su valor como herramienta de caracterización funcional (70).

Un estudio realizado por Tang et al. sugiere que la clasificación serológica utilizada en la práctica clínica no se correlaciona con la organización del transcriptoma ni con los estados funcionales del sistema inmune en la AR. La heterogeneidad de la enfermedad se manifiesta en múltiples niveles y no puede ser explicada por un único eje de caracterización (75).

Finalmente, la transcriptómica no se limita a complementar la caracterización clínica, sino que redefine la comprensión de la enfermedad. La identificación de estados moleculares diferenciados evidencia que la organización biológica de la AR no está determinada por los niveles de autoanticuerpos ni por la expresión clínica observable, sino por programas funcionales del sistema inmune que permanecen ocultos en la clasificación convencional. En

este contexto, la estratificación basada en perfiles transcriptómicos permite capturar con mayor precisión la heterogeneidad de la enfermedad, desplazando el enfoque desde marcadores indirectos hacia una caracterización fundamentada en la biología subyacente (69).

9. Conclusiones

La clasificación serológica basada en FR, ACPA y ANAs, ampliamente utilizada en la práctica clínica, mostró una relación limitada con la organización del transcriptoma en PBMC en esta cohorte. Este hallazgo sugiere que dichos marcadores no capturan completamente los estados funcionales del sistema inmune, evidenciando un desacople parcial entre la serología y la dimensión transcriptómica. En la misma línea, esta clasificación en endofenotipos presentó una capacidad limitada para explicar tanto la actividad clínica como la heterogeneidad molecular observada, lo cual se refleja en la ausencia de asociación significativa con el CDAI, sugiriendo que los autoanticuerpos no representan de manera directa la dinámica clínica de la enfermedad.

A nivel molecular, el análisis transcriptómico reveló una estructura global organizada, caracterizada por la presencia de subgrupos que no corresponden a los endofenotipos serológicos. Este comportamiento evidencia que la biología de la enfermedad se estructura en dimensiones independientes de la clasificación clínica inicial. Dentro de este contexto, la variabilidad en los perfiles transcriptómicos al interior de cada grupo introdujo heterogeneidad intragrupo, lo que limitó la detección de diferencias entre endofenotipos al enmascarar señales biológicas relevantes.

La depuración del conjunto de datos permitió reducir esta variabilidad y recuperar dichas señales, evidenciando diferencias transcriptómicas robustas entre los grupos. Este resultado confirma que la ausencia inicial de genes diferencialmente expresados no reflejaba homogeneidad biológica, sino una limitación en la resolución analítica. Bajo estas condiciones, el endofenotipo 1 se consolidó como un estado molecular distintivo, con un perfil transcriptómico consistentemente diferenciado frente a los demás grupos.

Las diferencias identificadas involucraron múltiples genes diferencialmente expresados y se caracterizaron por un predominio de genes sobreexpresados, lo que sugiere alteraciones en procesos biológicos asociados con la respuesta inmune y la regulación celular. El análisis funcional permitió agrupar estas diferencias en tres ejes principales —metabólico, transcripcional e inmuno-inflamatorio— organizados en redes funcionales interconectadas.

A partir de estos hallazgos, la reclasificación transcriptómica permitió reorganizar la cohorte en clústeres que agrupan muestras provenientes de distintos endofenotipos, evidenciando que la clasificación serológica no refleja completamente los estados moleculares compartidos entre los pacientes. Los clústeres identificados representan estados funcionales del sistema inmune, mostrando una transición desde perfiles dominados por procesos metabólicos hacia estados de activación inmunológica y configuraciones funcionales más complejas. En este contexto, la ausencia de una asociación clara entre los clústeres transcriptómicos, la serología y el CDAI sugiere que la transcriptómica podría capturar componentes biológicos adicionales no completamente representados por las variables clínicas y serológicas evaluadas.

Finalmente, estos resultados evidencian que la AR presenta una heterogeneidad biológica que no puede explicarse a partir de un único eje de análisis. La transcriptómica permite definir estados funcionales del sistema inmune que permanecen ocultos en la caracterización clínica y serológica, aportando una comprensión más precisa de la organización molecular de la enfermedad.

10. Limitaciones del estudio

El tamaño de la cohorte representa una limitación importante, ya que restringe la capacidad de generalización de los resultados y puede afectar la detección de asociaciones más sutiles entre variables clínicas, serológicas y transcriptómicas. Esta limitación se ve acentuada por la distribución no uniforme de las muestras entre los endofenotipos, lo que reduce la potencia estadística en ciertas comparaciones y condiciona la interpretación de algunos hallazgos.

Adicionalmente, el uso de PBMC como fuente de ARN proporciona una visión sistémica del sistema inmune, lo cual es adecuado para capturar respuestas globales, pero no permite reflejar completamente los procesos locales que ocurren en el tejido sinovial, principal sitio de patología en la AR. En este contexto, la información obtenida debe interpretarse como una aproximación parcial de la complejidad biológica de la enfermedad.

Por otra parte, la necesidad de depuración del conjunto de datos evidencia la presencia de una alta heterogeneidad biológica dentro de la cohorte; sin embargo, este proceso también introduce un sesgo potencial al excluir muestras que podrían corresponder a estados clínicamente relevantes. Finalmente, la naturaleza transversal del estudio limita la posibilidad de evaluar la dinámica temporal de los estados transcriptómicos, así como su relación con la progresión de la enfermedad o la respuesta al tratamiento.

11. Bibliografía.

1. Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis. *JAMA*. 2018 Oct 2;320(13):1360. doi:10.1001/jama.2018.13103
2. Collison J. Extracellular vesicles in bone cell crosstalk. *Nat Rev Rheumatol*. 2018 Jan 16;14(1):2–3. doi:10.1038/nrrheum.2017.191
3. Liu F, Wang XQ, Zou JW, Li M, Pan CC, Si YQ. Association between serum antinuclear antibody and rheumatoid arthritis. *Front Immunol*. 2024 Apr 22;15. doi:10.3389/fimmu.2024.1358114
4. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016 Oct 13;538(7624):161–4. doi:10.1038/538161a
5. Binignat M, Miao BY, Wibrand C, Yang MM, Rychkov D, Flynn E, et al. Single-cell RNA-seq analysis reveals cell subsets and gene signatures associated with Rheumatoid Arthritis Disease Activity. *JCI Insight*. 2024 Jul 2. doi:10.1172/jci.insight.178499
6. Hounkpe BW, Sales LP, Ribeiro SCR, Perez MO, Caparbo VF, Domiciano DS, et al. Transcriptomic signatures of classical monocytes reveal pro-inflammatory modules and heterogeneity in polyarticular juvenile idiopathic arthritis. *Front Immunol*. 2024 May 21;15. doi:10.3389/fimmu.2024.1400036

7. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *The Lancet*. 2016 Oct;388(10055):2023–38. doi:10.1016/S0140-6736(16)30173-8
8. Crowson CS, Rollefstad S, Ik Dahl E, KITAS GD, van Riel PLCM, Gabriel SE, et al. Impact of risk factors associated with cardiovascular outcomes in patients with rheumatoid arthritis. *Ann Rheum Dis*. 2018 Jan;77(1):48–54. doi:10.1136/annrheumdis-2017-211735
9. McInnes IB, Schett G. The Pathogenesis of Rheumatoid Arthritis. *New England Journal of Medicine*. 2011 Dec 8;365(23):2205–19. doi:10.1056/NEJMra1004965
10. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *The Lancet*. 2010 Sep;376(9746):1094–108. doi:10.1016/S0140-6736(10)60826-4
11. Peng X, Wang Q, Li W, Ge G, Peng J, Xu Y, et al. Comprehensive overview of microRNA function in rheumatoid arthritis. *Bone Res*. 2023 Jan 24;11(1):8. doi:10.1038/s41413-023-00244-1
12. Safiri S, Kolahi AA, Hoy D, Smith E, Bettampadi D, Mansournia MA, et al. Global, regional and national burden of rheumatoid arthritis 1990–2017: a systematic analysis of the Global Burden of Disease study 2017. *Ann Rheum Dis*. 2019 Nov;78(11):1463–71. doi:10.1136/annrheumdis-2019-215920
13. Fernández-Ávila DG, Betancur M, Kronfly A, Jáuregui E. Exploring drug utilization patterns, healthcare resource utilization, and epidemiology of rheumatoid arthritis in Colombia: a retrospective claims database study. *BMC Rheumatol*. 2025 Feb 7;9(1):13. doi:10.1186/s41927-025-00459-1
14. Dakin SG, Coles M, Sherlock JP, Powrie F, Carr AJ, Buckley CD. Pathogenic stromal cells as therapeutic targets in joint inflammation. *Nat Rev Rheumatol*. 2018 Dec 12;14(12):714–26. doi:10.1038/s41584-018-0112-7
15. Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis. *JAMA*. 2018 Oct 2;320(13):1360. doi:10.1001/jama.2018.13103
16. Bottini N, Firestein GS. Duality of fibroblast-like synoviocytes in RA: passive responders and imprinted aggressors. *Nat Rev Rheumatol*. 2013 Jan 13;9(1):24–33. doi:10.1038/nrrheum.2012.190
17. Mousavi MJ, Karami J, Aslani S, Tahmasebi MN, Vaziri AS, Jamshidi A, et al. Transformation of fibroblast-like synoviocytes in rheumatoid arthritis; from a friend to foe. *Autoimmunity Highlights*. 2021 Dec 5;12(1):3. doi:10.1186/s13317-020-00145-x
18. Firestein GS, Echeverri F, Yeo M, Zvaifler NJ, Green DR. Somatic mutations in the p53 tumor suppressor gene in rheumatoid arthritis synovium. *Proceedings of the National Academy of Sciences*. 1997 Sep 30;94(20):10895–900. doi:10.1073/pnas.94.20.10895
19. Tu J, Huang W, Zhang W, Mei J, Zhu C. A Tale of Two Immune Cells in Rheumatoid Arthritis: The Crosstalk Between Macrophages and T Cells in the Synovium. *Front Immunol*. 2021 Jun 17;12. doi:10.3389/fimmu.2021.655477
20. Kondo Y, Yokosawa M, Kaneko S, Furuyama K, Segawa S, Tsuboi H, et al. Review: Transcriptional Regulation of CD4+ T Cell Differentiation in Experimentally Induced Arthritis and Rheumatoid Arthritis. *Arthritis & Rheumatology*. 2018 May 13;70(5):653–61. doi:10.1002/art.40398
21. Niu Q, Cai B, Huang Z chun, Shi Y ying, Wang L lan. Disturbed Th17/Treg balance in patients with rheumatoid arthritis. *Rheumatol Int*. 2012 Sep 2;32(9):2731–6. doi:10.1007/s00296-011-1984-x

22. Rönnelid J, Turesson C, Kastbom A. Autoantibodies in Rheumatoid Arthritis – Laboratory and Clinical Perspectives. *Front Immunol.* 2021 May 14;12. doi:10.3389/fimmu.2021.685312
23. Binvignat M, Miao BY, Wibrand C, Yang MM, Rychkov D, Flynn E, et al. Single-cell RNA-seq analysis reveals cell subsets and gene signatures associated with Rheumatoid Arthritis Disease Activity. *JCI Insight.* 2024 Jul 2. doi:10.1172/jci.insight.178499
24. Kennedy A, Fearon U, Veale DJ, Godson C. Macrophages in Synovial Inflammation. *Front Immunol.* 2011;2. doi:10.3389/fimmu.2011.00052
25. Huang QQ, Doyle R, Chen SY, Sheng Q, Misharin A V., Mao Q, et al. Critical role of synovial tissue–resident macrophage niche in joint homeostasis and suppression of chronic inflammation. *Sci Adv.* 2021 Jan 8;7(2). doi:10.1126/sciadv.abd0515
26. Lavin Y, Mortha A, Rahman A, Merad M. Regulation of macrophage development and function in peripheral tissues. *Nat Rev Immunol.* 2015 Dec 25;15(12):731–44. doi:10.1038/nri3920
27. Cascao R, Rosário HS, Souto-Carneiro MM, Fonseca JE. Neutrophils in rheumatoid arthritis: More than simple final effectors. *Autoimmun Rev.* 2010 Jun;9(8):531–5. doi:10.1016/j.autrev.2009.12.013
28. Chen J, Cao Y, Xiao J, Hong Y, Zhu Y. The emerging role of neutrophil extracellular traps in the progression of rheumatoid arthritis. *Front Immunol.* 2024;15:1438272. doi:10.3389/fimmu.2024.1438272 PubMed PMID: 39221253.
29. Gajendran C, Fukui S, Sadhu NM, Zainuddin M, Rajagopal S, Gosu R, et al. Alleviation of arthritis through prevention of neutrophil extracellular traps by an orally available inhibitor of protein arginine deiminase 4. *Sci Rep.* 2023 Feb 23;13(1):3189. doi:10.1038/s41598-023-30246-2
30. Song X, Zhang Y, Zhao L, Fan J, Peng T, Ma Y, et al. Analyzation of the Peripheral Blood Mononuclear Cells Atlas and Cell Communication of Rheumatoid Arthritis Patients Based on Single-Cell RNA-Seq. *J Immunol Res.* 2023;2023:6300633. doi:10.1155/2023/6300633 PubMed PMID: 37600067.
31. Hall BE, Mazhar K, Macdonald E, Cassidy M, Doty M, Judkins C, et al. Transcriptome analysis of rheumatoid arthritis uncovers genes linked to inflammation-induced pain. *Sci Rep.* 2024 Oct 29;14(1):25893. doi:10.1038/s41598-024-77212-0
32. de Brito Rocha S, Baldo DC, Andrade LEC. Clinical and pathophysiologic relevance of autoantibodies in rheumatoid arthritis. *Advances in Rheumatology.* 2019 Dec 17;59(1):2. doi:10.1186/s42358-018-0042-8
33. Iberg CA, Jones A, Hawiger D. Dendritic Cells As Inducers of Peripheral Tolerance. *Trends Immunol.* 2017 Nov;38(11):793–804. doi:10.1016/j.it.2017.07.007
34. Iwasaki A, Medzhitov R. Control of adaptive immunity by the innate immune system. *Nat Immunol.* 2015 Apr 19;16(4):343–53. doi:10.1038/ni.3123
35. Schett G, Gravallesse E. Bone erosion in rheumatoid arthritis: mechanisms, diagnosis and treatment. *Nat Rev Rheumatol.* 2012 Nov 25;8(11):656–64. doi:10.1038/nrrheum.2012.153
36. Amarasekara DS, Yun H, Kim S, Lee N, Kim H, Rho J. Regulation of Osteoclast Differentiation by Cytokine Networks. *Immune Netw.* 2018;18(1). doi:10.4110/in.2018.18.e8

37. Asagiri M, Takayanagi H. The molecular understanding of osteoclast differentiation. *Bone*. 2007 Feb;40(2):251–64. doi:10.1016/j.bone.2006.09.023
38. Baum R, Gravalles EM. Bone as a Target Organ in Rheumatic Disease: Impact on Osteoclasts and Osteoblasts. *Clin Rev Allergy Immunol*. 2016 Aug 28;51(1):1–15. doi:10.1007/s12016-015-8515-6
39. Otero M, Goldring MB. Cells of the synovium in rheumatoid arthritis. Chondrocytes. *Arthritis Res Ther*. 2007;9(5):220. doi:10.1186/ar2292
40. Li Z, Liu B, Zhao D, Wang B, Liu Y, Zhang Y, et al. Omentin-1 prevents cartilage matrix destruction by regulating matrix metalloproteinases. *Biomedicine & Pharmacotherapy*. 2017 Aug;92:265–9. doi:10.1016/j.biopha.2017.05.059
41. Goldring MB, Marcu KB. Cartilage homeostasis in health and rheumatic diseases. *Arthritis Res Ther*. 2009;11(3):224. doi:10.1186/ar2592
42. Falconer J, Murphy AN, Young SP, Clark AR, Tiziani S, Guma M, et al. Review: Synovial Cell Metabolism and Chronic Inflammation in Rheumatoid Arthritis. *Arthritis & Rheumatology*. 2018 Jul 4;70(7):984–99. doi:10.1002/art.40504
43. Elshabrawy HA, Chen Z, Volin M V., Ravella S, Virupannavar S, Shahrara S. The pathogenic role of angiogenesis in rheumatoid arthritis. *Angiogenesis*. 2015 Oct 22;18(4):433–48. doi:10.1007/s10456-015-9477-2
44. Laurent L, Clavel C, Lemaire O, Anquetil F, Cornillet M, Zabraniecki L, et al. Fcγ receptor profile of monocytes and macrophages from rheumatoid arthritis patients and their response to immune complexes formed with autoantibodies to citrullinated proteins. *Ann Rheum Dis*. 2011 Jun;70(6):1052–9. doi:10.1136/ard.2010.142091
45. Negishi-Koga T, Gober HJ, Sumiya E, Komatsu N, Okamoto K, Sawa S, et al. Immune complexes regulate bone metabolism through FcRγ signalling. *Nat Commun*. 2015 Mar 31;6(1):6637. doi:10.1038/ncomms7637
46. Bartok B, Firestein GS. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. *Immunol Rev*. 2010 Jan 23;233(1):233–55. doi:10.1111/j.0105-2896.2009.00859.x
47. Curran AM, Girgis AA, Jang Y, Crawford JD, Thomas MA, Kawalerski R, et al. Citrullination modulates antigen processing and presentation by revealing cryptic epitopes in rheumatoid arthritis. *Nat Commun*. 2023 Feb 24;14(1):1061. doi:10.1038/s41467-023-36620-y PubMed PMID: 36828807.
48. Alghamdi M, Alasmari D, Assiri A, Mattar E, Aljaddawi AA, Alattas SG, et al. An Overview of the Intrinsic Role of Citrullination in Autoimmune Disorders. *J Immunol Res*. 2019 Nov 25;2019:1–39. doi:10.1155/2019/7592851
49. Loh TJ, Lim JJ, Jones CM, Dao HT, Tran MT, Baker DG, et al. The molecular basis underlying T cell specificity towards citrullinated epitopes presented by HLA-DR4. *Nat Commun*. 2024 Jul 23;15(1):6201. doi:10.1038/s41467-024-50511-w
50. Roudier J, Balandraud N, Auger I. How RA Associated HLA-DR Molecules Contribute to the Development of Antibodies to Citrullinated Proteins: The Hapten Carrier Model. *Front Immunol*. 2022 Jun 10;13. doi:10.3389/fimmu.2022.930112
51. Wang H, Kim SJ, Lei Y, Wang S, Wang H, Huang H, et al. Neutrophil extracellular traps in homeostasis and disease. *Signal Transduct Target Ther*. 2024 Sep 20;9(1):235. doi:10.1038/s41392-024-01933-x
52. Tang J, Xia J, Gao H, Jiang R, Xiao L, Sheng H, et al. IL33-induced neutrophil extracellular traps (NETs) mediate a positive feedback loop for synovial

- inflammation and NET amplification in rheumatoid arthritis. *Exp Mol Med*. 2024 Dec 2;56(12):2602–16. doi:10.1038/s12276-024-01351-7
53. Zhang Y, Wu Y, Yang H, Liu X, Pan B, Ding C, et al. NETs promote invasive behavior of fibroblast-like synoviocytes through GPIIb/IIIa in rheumatoid arthritis. *Front Immunol*. 2025 Nov 21;16. doi:10.3389/fimmu.2025.1667319
 54. Frade-Sosa B, Sanmartí R. Neutrophils, neutrophil extracellular traps, and rheumatoid arthritis: An updated review for clinicians. *Reumatología Clínica (English Edition)*. 2023 Nov;19(9):515–26. doi:10.1016/j.reumae.2023.10.002
 55. Liu F, Wang XQ, Zou JW, Li M, Pan CC, Si YQ. Association between serum antinuclear antibody and rheumatoid arthritis. *Front Immunol*. 2024;15:1358114. doi:10.3389/fimmu.2024.1358114 PubMed PMID: 38711518.
 56. Smith SL, Plant D, Eyre S, Barton A. The potential use of expression profiling: implications for predicting treatment response in rheumatoid arthritis. *Ann Rheum Dis*. 2013 Jul;72(7):1118–24. doi:10.1136/annrheumdis-2012-202743
 57. Giannopoulou EG, Elemento O, Ivashkiv LB. Use of RNA sequencing to evaluate rheumatic disease patients. *Arthritis Res Ther*. 2015 Jul 1;17(1):167. doi:10.1186/s13075-015-0677-3
 58. Sanayama Y, Ikeda K, Saito Y, Kagami S, Yamagata M, Furuta S, et al. Prediction of Therapeutic Responses to Tocilizumab in Patients With Rheumatoid Arthritis: Biomarkers Identified by Analysis of Gene Expression in Peripheral Blood Mononuclear Cells Using Genome-Wide DNA Microarray. *Arthritis & Rheumatology*. 2014 Jun 27;66(6):1421–31. doi:10.1002/art.38400
 59. Sellam J, Marion-Thore S, Dumont F, Jacques S, Garchon H, Rouanet S, et al. Use of Whole-Blood Transcriptomic Profiling to Highlight Several Pathophysiologic Pathways Associated With Response to Rituximab in Patients With Rheumatoid Arthritis: Data From a Randomized, Controlled, Open-Label Trial. *Arthritis & Rheumatology*. 2014 Aug 28;66(8):2015–25. doi:10.1002/art.38671
 60. Mao X, Shi M, Zhang B, Fu R, Cai M, Yu S, et al. Integration of single-cell and bulk RNA sequencing revealed immune heterogeneity and its association with disease activity in rheumatoid arthritis patients. *Immunol Res*. 2024 Oct 16;72(5):1120–35. doi:10.1007/s12026-024-09513-5
 61. Luo Y, de Lange KM, Jostins L, Moutsianas L, Randall J, Kennedy NA, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat Genet*. 2017 Feb 9;49(2):186–92. doi:10.1038/ng.3761
 62. Gianfrancesco M, Yazdany J, Robinson PC. Epidemiology and outcomes of novel coronavirus 2019 in patients with immune-mediated inflammatory diseases. *Curr Opin Rheumatol*. 2020 Sep;32(5):434–40. doi:10.1097/BOR.0000000000000725
 63. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014 Feb 25;506(7488):376–81. doi:10.1038/nature12873
 64. Humby F, Lewis M, Ramamoorthi N, Hackney JA, Barnes MR, Bombardieri M, et al. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann Rheum Dis*. 2019 Jun;78(6):761–72. doi:10.1136/annrheumdis-2018-214539
 65. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating

- single-cell transcriptomics and mass cytometry. *Nat Immunol.* 2019 Jul 6;20(7):928–42. doi:10.1038/s41590-019-0378-1
66. Liu Y, Fan S, Meng S. Identification of the candidate genes of diagnosing rheumatoid arthritis using the single-cell sequencing technology and T cell subclusters analysis of patients with rheumatoid arthritis. *Arch Rheumatol.* 2023 Mar 1;38(1):109–18. doi:10.46497/ArchRheumatol.2022.9573
 67. Sharma S, Bluett J. Towards Personalized Medicine in Rheumatoid Arthritis. *Open Access Rheumatol.* 2024 May;Volume 16:89–114. doi:10.2147/OARRR.S372610
 68. Xue M, Wang H, Campos F, Jackson CJ, March L. Rheumatoid Arthritis: Biomarkers and the Latest Breakthroughs. *Int J Mol Sci.* 2025 Oct 30;26(21):10594. doi:10.3390/ijms262110594
 69. Chang MJ, Feng QF, Hao JW, Zhang YJ, Zhao R, Li N, et al. Deciphering the molecular landscape of rheumatoid arthritis offers new insights into the stratified treatment for the condition. *Front Immunol.* 2024 Jun 25;15. doi:10.3389/fimmu.2024.1391848
 70. Tasaki S, Suzuki K, Kassai Y, Takeshita M, Murota A, Kondo Y, et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. *Nat Commun.* 2018 Jul 16;9(1):2755. doi:10.1038/s41467-018-05044-4
 71. Satija R, Shalek AK. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol.* 2014 May;35(5):219–29. doi:10.1016/j.it.2014.03.004 PubMed PMID: 24746883.
 72. Leone P, Susca N, Racanelli V. Editorial: Autoimmune diseases: from molecular mechanisms to therapy development. *Front Immunol.* 2026 Jan 6;16. doi:10.3389/fimmu.2025.1768955
 73. Zhang Y, Xu H, Xu L, Jiang S, Yu Y, Zhao H. Multi-omics-driven biomarker discovery in autoimmune diseases: a comprehensive review. *Front Immunol.* 2025 Dec 8;16. doi:10.3389/fimmu.2025.1652211
 74. Gunkl-Tóth L, Sütő G, Kumánovics G, Kun J, Urbán P, Gyenesei A, et al. Transcriptomic Analysis of Peripheral Blood Mononuclear Cells Reveals Pain and Inflammation Specific Alterations in Difficult-to-treat Rheumatoid Arthritis. 2023 Nov.
 75. Tang M, Haese-Hill W, Morton F, Goodyear C, Porter D, Siebert S, et al. RNAcare: integrating clinical data with transcriptomic evidence using rheumatoid arthritis as a case study. *BMC Med Genomics.* 2025 May 21;18(1):93. doi:10.1186/s12920-025-02162-z

Anexos

Anexo 1. Control de calidad de ARN extraído de PBMC

<i>Numero De muestra</i>	Muestra	Concentración	RIN
1	G1P128T1	38,0	4,7
2	G1P129T1	25,6	4,7
3	G1P131T1	33,3	4,5
4	G1P29T1	30,7	6,7
5	G1P33T1	9,0	7,9
6	G1P36T1	27,4	6,4
7	G1P37T1	5,2	7,8
8	G1P42T1	8,1	7,1
9	G1P90T1	---	---
10	G1P44T1	10,0	6,3
11	G1P53T1	13,9	8,4
12	G1P54T1	6,5	8,4
13	G1P62T1	26,6	8,4
14	G1P64T1	21,6	8,4
15	G1P106T1	29,3	8,4
16	G1P97T1	9,9	8,4
17	G1P98T1	24,0	8,4
18	G1P105T1	35,0	8,4
19	G1P83T1	29,9	8,4
20	G1P85T1	18,7	8,4
21	G1P86T1	13,2	8,4
22	G1P92T1	5,1	8,4
23	G1P93T1	11,3	8,4
24	G1P66T1	9,1	8,4
25	G1P67T1	8,9	8,4
26	G1P69T1	20,7	8,4
27	G1P72T1	31,9	8,4
28	G1P74T1	43,5	8,4
29	G1P77T1	48,6	8,4
30	G1P78T1	26,2	8,4
31	G1P95T1	18,9	8,4
32	G1P80T1	39,5	8,4
33	G1P58T1	21,6	8,4
34	G1P99T1	18,5	8,4
35	G1P135T1	19,3	8,4
36	G1P75T1	9,0	8,4
37	G1P127T1	51,4	8,4
38	G1P27T1	13,5	8,4

Tabla A1. Resultados de concentración y calidad de ARN de las muestras incluidas en el estudio. Se presenta la concentración de ARN (ng/μL) y el índice de integridad (RIN) obtenidos tras la extracción de ARN total a partir de células mononucleares de sangre periférica (PBMC). Los valores de RIN fueron determinados mediante análisis de integridad de ARN, indicando la calidad de las muestras para su uso en secuenciación. Las muestras con valores no disponibles se indican como “—”.

Anexo 2. Enriquecimiento funcional complementario en las categorías de Componentes Celulares (CC) y Funciones Moleculares (MF). Sección 7.8.

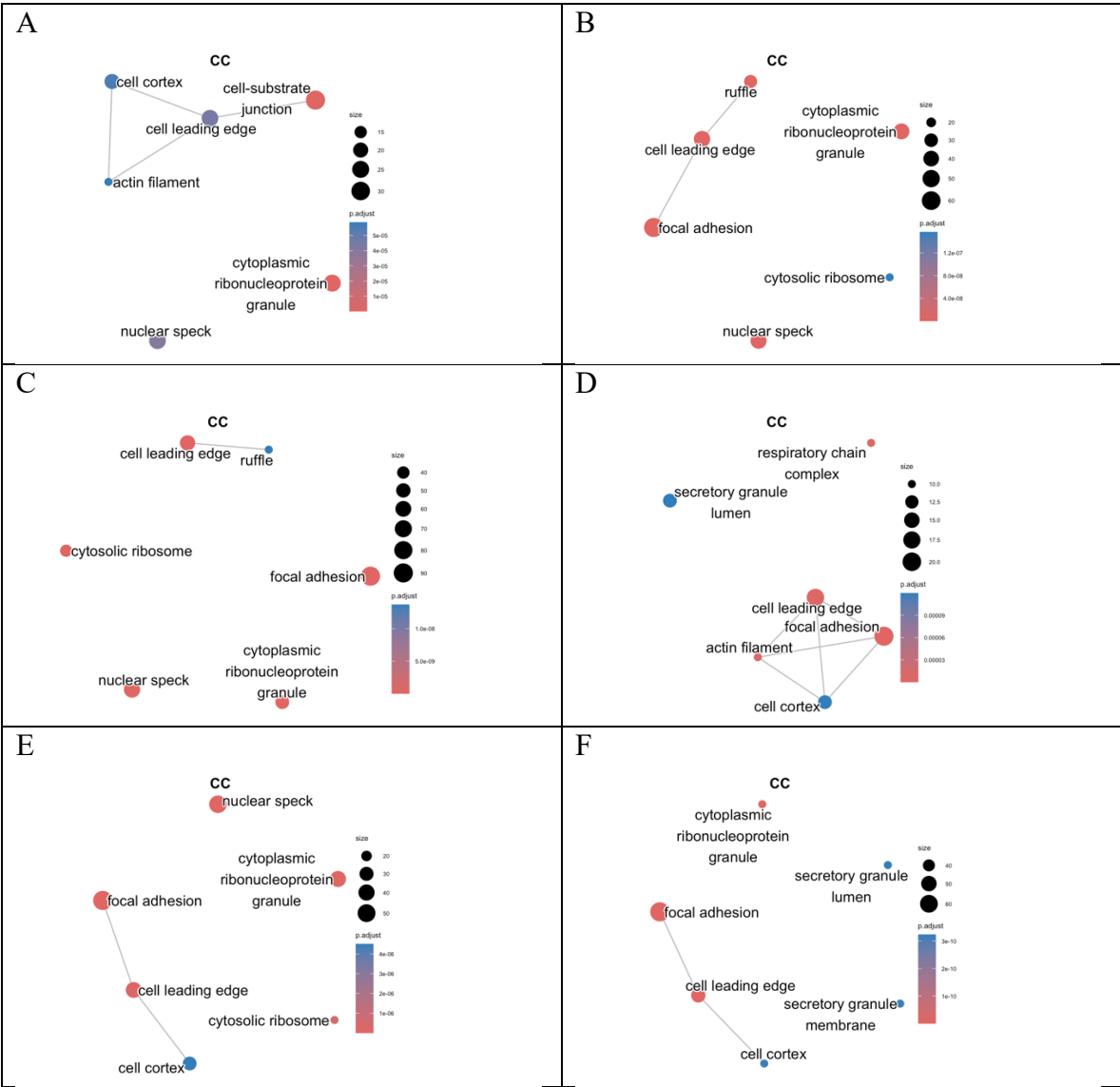


Figura A1. Enriquecimiento funcional en la categoría de Componentes Celulares (CC) para los genes diferencialmente expresados en las comparaciones entre endofenotipos. Cada panel (A–F) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 7, respectivamente. Los nodos representan términos enriquecidos de Gene Ontology (Componentes Celulares), cuyo tamaño es proporcional al número de genes asociados (Count), mientras que el color indica la significancia estadística ajustada (p.adjust). Las conexiones entre nodos reflejan relaciones funcionales entre los términos enriquecidos.

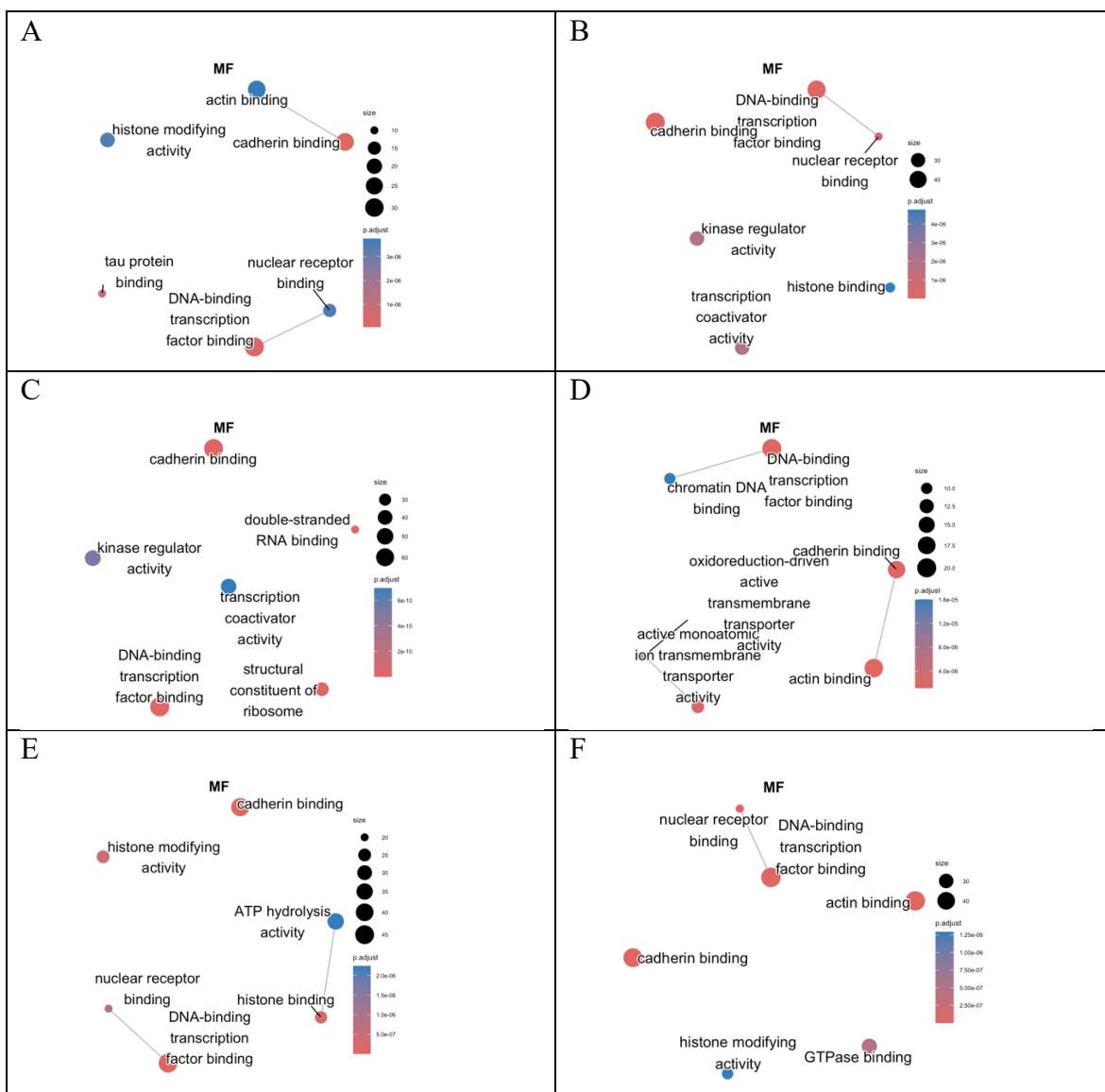


Figura A2. Enriquecimiento funcional en la categoría de Funciones Moleculares (MF) para los genes diferencialmente expresados en las comparaciones entre endofenotipos. Cada panel (A–F) corresponde a una comparación entre el endofenotipo 1 y los endofenotipos 2 a 7, respectivamente. Los nodos representan términos enriquecidos de Gene Ontology (Funciones Moleculares), cuyo tamaño es proporcional al número de genes asociados (Count), mientras que el color indica la significancia estadística ajustada (p.adjust). Las conexiones entre nodos reflejan relaciones funcionales entre los términos enriquecidos.

Se presentan los resultados del análisis de enriquecimiento funcional de los genes diferencialmente expresados (DEGs) en las categorías de Componentes Celulares (CC) y Funciones Moleculares (MF) de Gene Ontology.

En la categoría de Componentes Celulares (CC) se observan términos asociados principalmente a estructuras relacionadas con organización celular y dinámica del citoesqueleto, incluyendo cell cortex, focal adhesion, cell leading edge y actin filament, así como componentes asociados a actividad ribosomal y complejos ribonucleoproteicos. Estos resultados sugieren la participación de procesos relacionados con adhesión celular, migración y organización estructural en los perfiles transcriptómicos evaluados.

En la categoría de Funciones Moleculares (MF) se identifican términos asociados a regulación transcripcional y señalización celular, destacándose DNA-binding transcription factor binding, nuclear receptor binding, kinase regulator activity y transcription coactivator activity. Adicionalmente, se observan funciones relacionadas con interacción proteína-proteína y actividad estructural, como cadherin binding y actin binding, así como procesos asociados a regulación epigenética (histone binding y histone modifying activity).

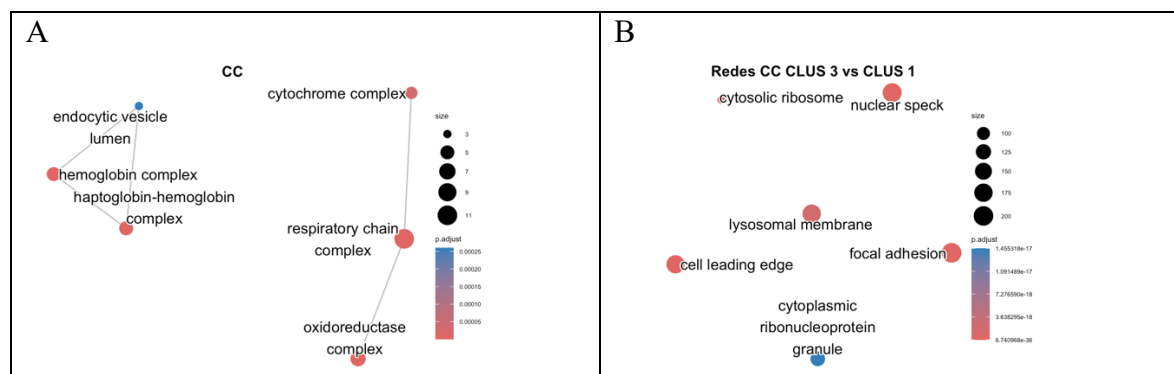
Estos resultados complementan los hallazgos descritos en Procesos Biológicos (BP), aportando resolución adicional a nivel estructural y funcional, aunque con menor capacidad integradora para la interpretación global de los programas biológicos implicados en la cohorte.

Anexo 3. Distribución de las muestras según clúster transcriptómico

	MUESTRAS	TOTAL MUESTRAS
CLÚSTER_1	G1P42T1, G1P44T1, G1P48T1, G1P53T1, G1P54T1, G1P58T1, G1P62T1, G1P64T1	8
CLÚSTER_2	G1P29T1, G1P72T1, G1P74T1, G1P92T1, G1P128T1, G1P129T1, G1P131T1	7
CLÚSTER_3	G1P37T1, G1P66T1, G1P67T1, G1P75T1, G1P78T1, G1P80T1, G1P83T1, G1P85T1, G1P86T1, G1P90T1, G1P93T1, G1P95T1, G1P99T1, G1P105T1, G1P106T1, G1P127T1, G1P135T1	17
CLÚSTER_4	G1P27T1, G1P33T1, G1P36T1, G1P69T1, G1P77T1, G1P97T1, G1P98T1	7

Tabla A2. Asignación de muestras a los clústeres transcriptómicos (CLUS_1–CLUS_4). Se presenta la distribución de las muestras incluidas en el estudio según su agrupación en clústeres transcriptómicos definidos a partir del análisis de expresión génica. La tabla incluye el listado de identificadores de muestra por clúster y el número total de muestras en cada grupo.

Anexo 4. Enriquecimiento funcional complementario en las categorías de Componentes Celulares (CC) y Funciones Moleculares (MF). Sección 7.9.



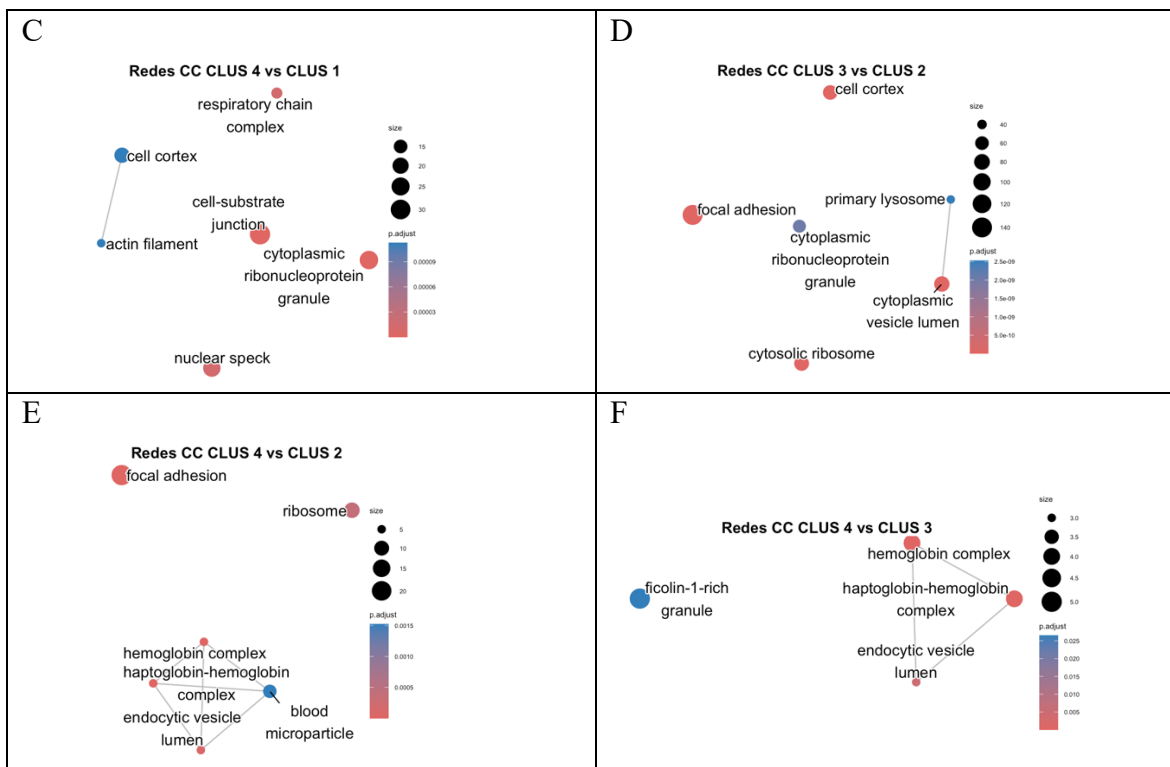
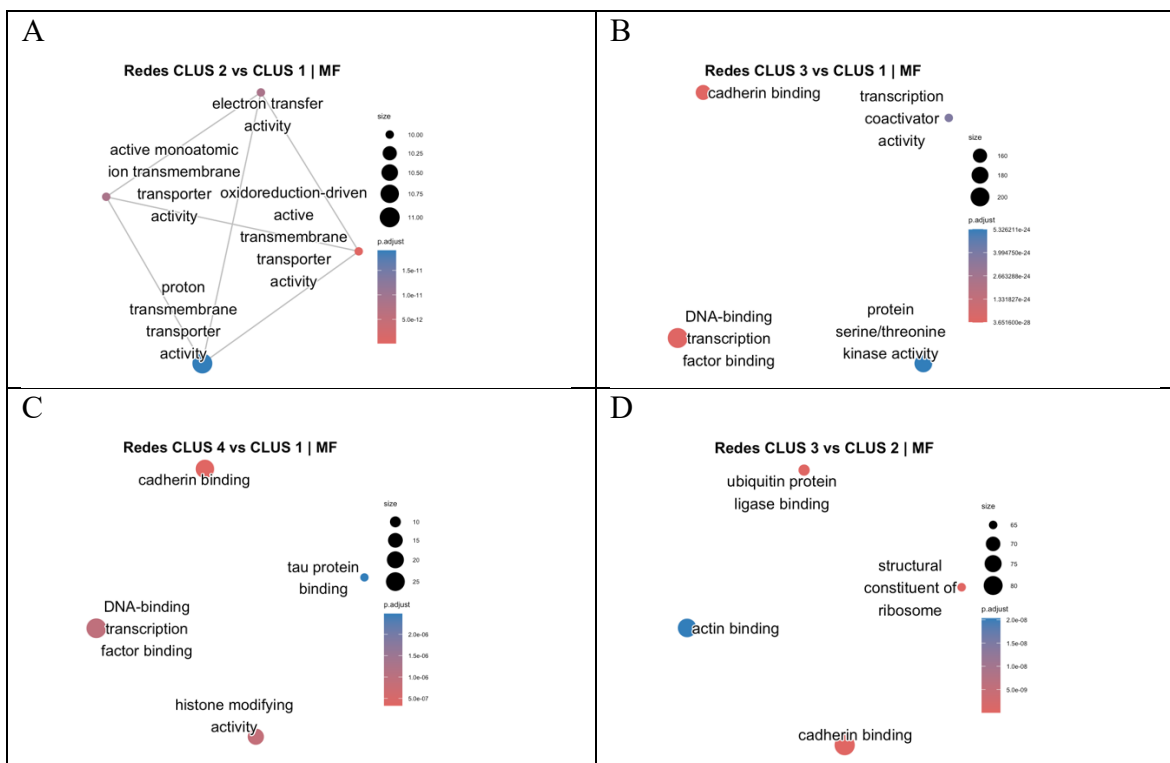


Figura A3. Enriquecimiento funcional en la categoría de Componentes Celulares (CC) para los genes diferencialmente expresados en las comparaciones entre endofenotipos. Cada panel (A–F) corresponde a una comparación entre clusters. Los nodos representan términos enriquecidos de Gene Ontology (Componentes Celulares), cuyo tamaño es proporcional al número de genes asociados (Count), mientras que el color indica la significancia estadística ajustada (p.adjust). Las conexiones entre nodos reflejan relaciones funcionales entre los términos enriquecidos.



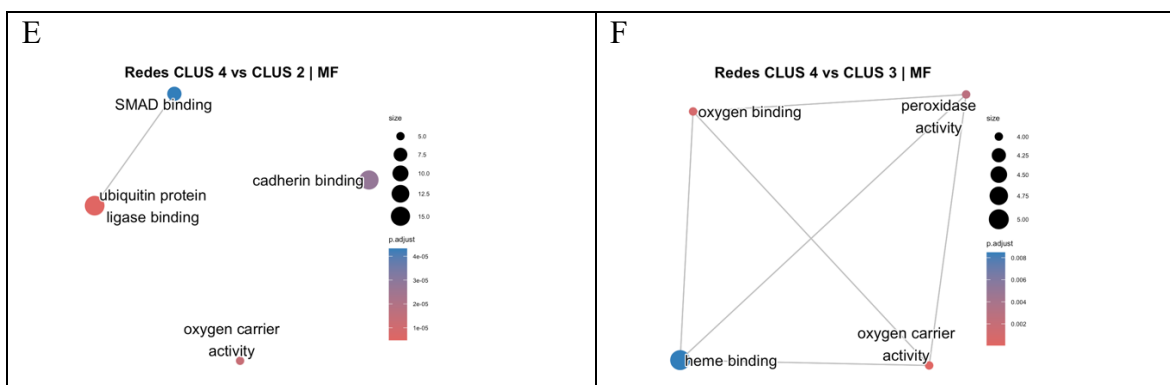


Figura A4. Enriquecimiento funcional en la categoría de Funciones Moleculares (MF) para los genes diferencialmente expresados en las comparaciones entre endofenotipos. Cada panel (A–F) corresponde a una comparación entre los clusters. Los nodos representan términos enriquecidos de Gene Ontology (Funciones Moleculares), cuyo tamaño es proporcional al número de genes asociados (Count), mientras que el color indica la significancia estadística ajustada (p.adjust). Las conexiones entre nodos reflejan relaciones funcionales entre los términos enriquecidos.

Se presentan los resultados del análisis de enriquecimiento funcional de los genes diferencialmente expresados (DEGs) en las categorías de Componentes Celulares (CC) y Funciones Moleculares (MF) de Gene Ontology, derivados de las comparaciones entre clústeres transcriptómicos.

En la categoría de Componentes Celulares (CC) se identifican términos asociados a estructuras relacionadas con metabolismo energético, complejos ribonucleoproteicos y organización celular. Destacan componentes como respiratory chain complex, cytochrome complex y oxidoreductase complex, junto con estructuras vinculadas a adhesión y dinámica celular, incluyendo cell cortex, focal adhesion, cell leading edge y actin filament. Asimismo, se observan términos relacionados con compartimentos intracelulares como cytoplasmic ribonucleoprotein granule, nuclear speck y lysosomal membrane. Este conjunto sugiere la participación de procesos asociados tanto a la organización estructural como a la actividad metabólica y la dinámica subcelular en los estados transcriptómicos definidos.

En la categoría de Funciones Moleculares (MF) se evidencian funciones relacionadas con transporte transmembrana, regulación transcripcional y actividad enzimática. Se destacan términos como transmembrane transporter activity, active ion transmembrane transporter activity y oxidation-reduction-driven active transmembrane transporter activity, asociados a procesos de intercambio iónico y metabolismo celular. De igual forma, se identifican funciones vinculadas a control génico, incluyendo DNA-binding transcription factor binding, transcription coactivator activity y histone modifying activity. Adicionalmente, se observan funciones relacionadas con interacción proteína-proteína y organización estructural, como cadherin binding, actin binding y ubiquitin protein ligase binding, así como actividades asociadas a metabolismo del oxígeno, como oxygen binding y heme binding.

Estos resultados amplían la interpretación funcional de los hallazgos descritos en Procesos Biológicos (BP), evidenciando que las diferencias entre clústeres transcriptómicos no solo se reflejan en programas biológicos generales, sino también en la organización subcelular y en funciones moleculares específicas. Aunque estas categorías aportan mayor resolución a nivel estructural y mecanístico, su capacidad integradora es menor en comparación con BP para la interpretación global de los estados biológicos presentes en la cohorte.