



**Universidad
Europea**

Máster en Bioinformática

**Caracterización funcional de una firma
transcriptómica diferencial en pacientes
jóvenes con mieloma múltiple**

Autor: Marta Hidalgo Soto

Tutor: Pablo Alberto Mínguez Paniagua

Curso 2024-2025

AGRADECIMIENTOS

A mi familia, con especial cariño a mi hermano, mi madre, mi prima y tías.

A mi padre, que se nos adelantó, pero nos dejó un legado de abrazos, paciencia infinita y chistes malos.

A mi tutor, por enseñarme nuevas posibilidades en bioinformática y por su constante disponibilidad.

Índice

Resumen	4
2. Introducción	5
2.1. Mieloma múltiple: aspectos clínicos y moleculares	5
2.2. Impacto de la edad en el mieloma múltiple	5
2.3. Análisis transcriptómico y mecanístico mediante RNA-Seq y redes moleculares	6
2.4. Repositorios de anotación: GO y KEGG	6
2.5. Justificación del presente trabajo	7
2. Hipótesis y Objetivos	7
2.1. Hipótesis	7
3.2. Objetivos	7
3.2.1. Objetivos específicos	8
4. Materiales y métodos	8
4.1. Fuente de datos (CoMMpass y GDC)	8
4.2. Diseño y población	10
4.3. Recogida y preparación de datos	10
4.4. Análisis estadístico	11
4.4.1. Expresión diferencial (DGE)	11
4.4.2. Enriquecimiento funcional (GSEA)	11
4.5. Redes de interacción	12
4.6. Software y paquetes empleados	12
5. Resultados	13
5.1. Diferencias ligadas a la edad en la aparición de MM: análisis por subcohortes.....	13
5.2. Análisis de expresión génica diferencial	14
5.3. Puntuación zUp–zDown (síntesis del DGE).....	17
5.4. Detección de funciones sobre e infrarepresentadas en el transcrito subcohortes)	18
5.5. Análisis de redes	21
6. Discusión	24
Conclusiones	26
8. Bibliografía	26
9. Anexos	29

RESUMEN

La edad al diagnóstico del mieloma múltiple (MM) podría modular programas biológicos tumorales, pero la huella transcriptómica diferencial entre pacientes jóvenes y mayores está poco definida. Este estudio contrasta perfiles de RNA-seq en pacientes con MM debutando a diferentes edades para buscar diferencias en la regulación molecular. Para ello hemos integrado expresión diferencial, enriquecimiento funcional y análisis de redes proteína-proteína, para obtener diferencias robustas entre grupos.

Objetivos:

Comparar los perfiles de expresión génica entre pacientes <50 y ≥50 años al diagnóstico del MM; identificar módulos funcionales e interacciones predominantes por grupo; y explorar una firma de expresión estratificada por edad.

Material y métodos:

Estudio observacional sobre RNA-seq basal de la cohorte MMRF-CoMMpass. Se analizaron 762 transcriptomas basales, tras la descarga de la matriz de conteos, el filtrado de genes de baja expresión (CPM) y normalización con *TMM*. Se compararon las dos subcohortes con modelización *limma-voom* contrastando los grupos <50 vs ≥50 años. La expresión diferencial se definió con $FDR < 0,05$ y $|\log_2FC| \geq 1$. La interpretación funcional combinó GSEA *preranked* y análisis de sobrerrepresentación (GO/KEGG). Las señales coordinadas se contextualizaron en redes (Cytoscape) evaluando centralidad y módulos densos.

Resultados:

El análisis global mostró solapamiento de genes diferencialmente expresados entre grupos, indicando que hay una parte de la señal que es independiente a la edad. Se identificaron 295 genes diferenciales ($FDR < 0,05$; $|\log_2FC| \geq 1$). En <50 años predominó una firma proliferativa/replicativa con nodos centrales (*BRCA2*, *ORC6*, *CDC25B*, *RBBP5*). En ≥50 años, destacó una firma inmune-adhesiva con mayor expresión coordinada de genes de presentación antigénica (*ICAM1*, *SELL*, *HLA-E*, *FCGR3A*). La concordancia entre expresión diferencial, enriquecimiento y estructura de redes respalda un desplazamiento funcional: “proliferativo” en jóvenes frente a “inmune/adhesivo” en mayores, sin reprogramaciones masivas.

Conclusiones:

Se ha observado dos perfiles transcriptómicos según la estratificación por edad. En pacientes <50 años, predomina un perfil más proliferativo; y en mayores, un perfil inmune-adhesivo, sin cambios globales masivos.

Palabras clave:

Mieloma múltiple; edad; transcriptómica; RNA-seq; DGE; GSEA; Redes.

2. Introducción

2.1. Mieloma múltiple: aspectos clínicos y moleculares

El mieloma múltiple (MM) es una neoplasia de células plasmáticas de la médula ósea (MO), que de acuerdo con la *International Myeloma Working Group* (IMWG), implica la presencia de $\geq 10\%$ de infiltración de células plasmáticas atípicas en MO, o bien presencia de plasmocitoma(s), junto con la asociación de daño de órgano diana según los criterios CRAB (disfunción renal, anemia, lesiones líticas o fracturas patológicas o hipercalcemia, no justificadas por otra causa (Rajkumar et al., 2014).

En ausencia de CRAB, el diagnóstico se establece si concurren biomarcadores SLiM por su altísimo riesgo de progresión: $\geq 60\%$ de plasmocitos en MO, cociente de cadenas ligeras libres (FLC) involucrada/no involucrada ≥ 100 (con FLC involucrada ≥ 100 mg/L) o más de una lesión focal en la resonancia magnética de cuerpo entero.

Desde el punto de vista molecular, el MM se estructura en dos grandes categorías citogenéticas iniciales. El MM hiperdiploide muestra ganancias trisómicas de cromosomas impares, mientras que el no hiperdiploide se caracteriza por translocaciones del locus IgH (14q32) que sitúan oncogenes bajo el control del promotor de IgH (por ejemplo, t(11;14)/CCND1, t(4;14)/FGFR3–MMSET, t(14;16)/MAF). Estas lesiones primarias coexisten con alteraciones secundarias de impacto pronóstico, como ganancia de 1q, del(1p), del(13q) o del(17p)/TP53. A nivel mutacional destacan las mutaciones recurrentes que afectan a las vías RAS/MAPK (KRAS, NRAS, BRAF), junto con genes de procesamiento de ARN (DIS3), biología secretora (FAM46C/TENT5C) y reparación del ADN (ATM, ATR, TP53). Este paisaje configura trayectorias de evolución clonal, con subclonas que emergen o declinan a lo largo del curso clínico y bajo presión terapéutica. En paralelo, los subtipos transcriptómicos y los estados funcionales derivados de RNA-Seq aportan una capa adicional de estratificación biológica que complementa a la citogenética clásica (Walker BA, et al., 2018; van de Donk NW et al., 2021; Awada H et al., 2021).

2.2. Impacto de la edad en el mieloma múltiple

Aunque la mayoría de los diagnósticos ocurren en personas mayores de 65 años, aproximadamente el 10 % de los casos se presentan en pacientes menores de 50 años (Tanguay M et al., 2023). La edad es un factor pronóstico crucial en el MM, con diferencias sustanciales en la biología, la presentación clínica, respuesta al tratamiento y pronóstico entre pacientes jóvenes y mayores.

Un ejemplo, es la amplia cohorte retrospectiva presentada por el IMWG (n=10 549), que corroboró un aumento de supervivencia en menores de 50 años significativamente superior, alcanzando tasas de supervivencia a 10 años de hasta el 43 % tras terapias

intensivas, frente al 29 % en pacientes mayores de 50 años (Ludwig et al., 2008). Sin embargo, a pesar de estas observaciones clínicas bien establecidas, los mecanismos moleculares específicos que explican las diferencias en función de la edad no están completamente definidos, destacando la importancia de estudios transcriptómicos específicos para dilucidar las vías biológicas subyacentes (Rojas et al., 2019).

2.3. Análisis transcriptómico mediante RNA-Seq y redes moleculares

La secuenciación del transcriptoma (RNA-Seq) es una técnica bien establecida para analizar la expresión génica, permitiendo la identificación precisa de genes diferencialmente expresados, nuevas isoformas y variantes de *splicing* involucradas en la fisiopatología del cáncer (Stark et al., 2019). En el contexto del MM, el análisis transcriptómico ha aportado información crítica sobre mecanismos moleculares asociados a la progresión tumoral y la respuesta terapéutica, permitiendo la caracterización de firmas génicas específicas relacionadas con diferentes perfiles clínicos y biológicos de la enfermedad (Jang et al., 2019).

Además, la integración del análisis transcriptómico con enfoques de enriquecimiento o basados en redes biológicas, permite interpretar los resultados a nivel funcional, facilitando la identificación de vías de señalización intracelulares y procesos biológicos específicos que podrían estar alterados diferencialmente entre grupos clínicos definidos, como pacientes jóvenes (<50 años) versus mayores (≥50 años). Esta combinación proporciona una perspectiva sistémica y funcional, permitiendo no solo describir diferencias transcriptómicas, sino también inferir posibles mecanismos regulatorios o terapéuticos específicos implicados en la patogénesis y evolución clínica del MM (Rojas et al., 2019).

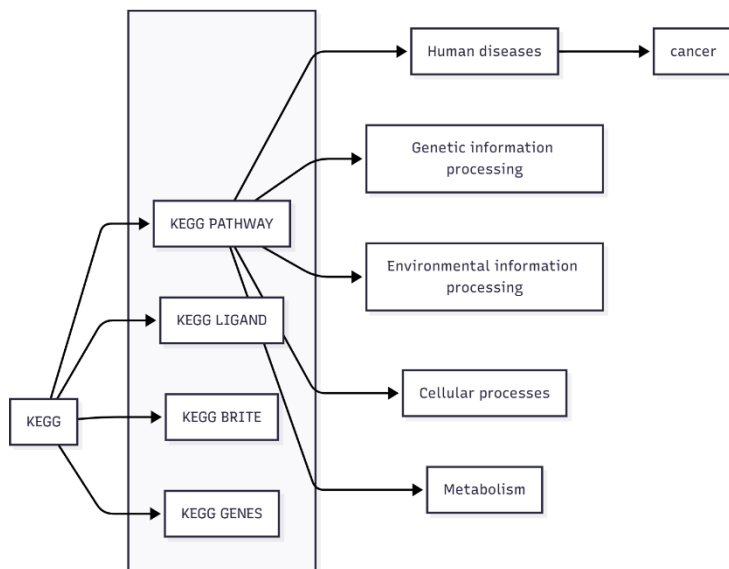
2.4. Repositorios de anotación: GO y KEGG

Gene Ontology (GO) es un repositorio donde se recogen anotaciones funcionales estructurado como un grafo acíclico dirigido (DAG) en tres dominios: Proceso Biológico (BP), Función Molecular (MF) y Componente Celular (CC). Cada término integra relaciones jerárquicas (p. ej., *is_a*, *part_of*) y códigos de evidencia. En el presente trabajo, GO se utiliza como colección de conjuntos génicos para el análisis de enriquecimiento. (Gene Ontology Consortium, 2023).

Otro punto de vista, es la base de datos *Kyoto Encyclopedia of Genes and Genomes* (KEGG). Recopila rutas dibujadas que integran interacciones, reacciones y módulos funcionales a distintos niveles de organización. Sus mapas de vía se emplean para evaluar los conjuntos génicos que están coordinadamente sobreexpresados (o infraexpresados) y en qué vías celulares están implicados. KEGG ofrece una representación coherente y compacta de procesos celulares (Kanehisa M et al., 2020); no obstante, es un modelo, por lo que la interpretación se contrasta con resultados de

DGE y con la organización en redes (módulos y hubs) para reforzar la lectura biológica (Kanehisa et al., 2025).

Figura 1. Componentes de KEGG



2.5. Justificación del presente trabajo

Aunque existen estudios previos que han comparado la expresión génica en pacientes con MM según su edad (Keskin M et al., 2022), aún son escasos aquellos que integran un análisis funcional y de redes basado en redes biológicas para caracterizar las diferencias moleculares específicas asociadas a pacientes jóvenes. Entender estas diferencias puede revelar nuevos biomarcadores pronósticos o predictivos, además de identificar potenciales dianas terapéuticas adaptadas específicamente a esta población, lo que mejoraría el manejo clínico y la precisión terapéutica en MM.

3. Hipótesis y Objetivos

3.1. Hipótesis

La integración de un *pipeline* reproducible de análisis transcriptómico, estudio de expresión génica diferencial (DGE) y de enriquecimiento de *set* de genes (GSEA), con visualización y análisis de redes en Cytoscape permite modelizar mecanismos celulares y subredes funcionales asociados a la edad al diagnóstico en mieloma múltiple. En concreto, los pacientes <50 años presentarán una firma transcriptómica diferencial respecto a ≥50 años, sugiriendo alteraciones biológicas específicas por edad.

3.2 Objetivos

Desarrollar y optimizar un *pipeline* de R para el análisis de RNA-Seq en la cohorte *Multiple Myeloma Research Foundation* (MMRF)-CoMMpass, que integre DGE (*limma-voom*), GSEA/ORR (*clusterProfiler/fgsea*) y redes de interacciones proteína-proteína (PPI) mediante *STRING* y *Cytoscape*), aplicándolo como caso de uso a la comparación <50 vs. ≥50 años en el debut del mieloma múltiple.

3.2.2 Objetivos específicos

- a) Obtener y procesar datos de RNA-Seq de dos cohortes de MM mayores y menores de 50 años.
- b) Identificar genes diferencialmente expresados en cada una de las subcohortes.
- c) Detectar e interpretar funciones biológicas asociadas a cada una de las cohortes.
- d) Construir y analizar redes de interacción proteica para detectar módulos funcionales específicos de cada cohorte.
- e) Evaluar firmas transcriptómicas: Calcular puntuación *zUp* – *zDown* tras estandarizar la expresión génica (z-score) y comparar entre grupos.

4. Materiales y métodos

4.1. Fuente de datos (CoMMpass y GDC)

CoMMpass (MMRF; NCT01454297) es un estudio prospectivo y longitudinal de pacientes con mieloma múltiple de nuevo diagnóstico (NDMM), diseñado para vincular resultados clínicos con perfiles genómicos. El protocolo contempla la recogida seriada de información clínica (por ej.: tratamientos, respuestas o eventos) y muestras biológicas durante 8 años de seguimiento (Figura 1).

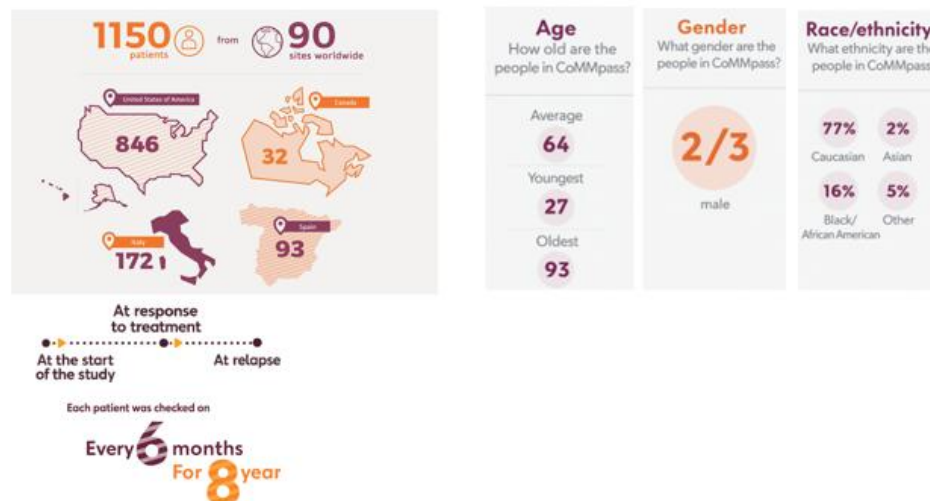


Figura 2. Infografía del estudio MMRF-CoMMpass

Imagen extraída de: <https://themmrf.org/> [Último acceso 15/Sep/2025]

Las muestras tumorales proceden de aspirado de médula ósea al diagnóstico, recaídas y confirmación de respuestas completas. Para los análisis de *bulk* que usa la comunidad (WGS, WES y RNA-seq), se enriquecen por CD138 para purificar células plasmáticas (selección inmunomagnética), lo que proporciona una fracción tumoral de alta pureza para la secuenciación masiva.

El tamaño y la disponibilidad por recurso se sintetizan en la Tabla 1.

Tabla 1. Disponibilidad de casos y RNA-seq basal

Recurso	Alcance	Pacientes.muestras	Ventana.temporal	Notas
CoMMpass (cohorte prospectiva)	Estudio longitudinal de NDMM, multi-omics	1.143 pacientes	2011–2016	Multicéntrico. Seguimiento trimestral.
GDC: Proyecto MMRF-COMMPASS	Datos armonizados	~995 casos con ficheros ómicos y clínicos	Actualización continua	RNA-seq “STAR-Counts” disponible vía API y descarga controlada.
RNA-seq basal	Subconjuntos filtrados por calidad (FASTQ)	714 (consenso de expresión) / 721 (IA9 alta calidad) / 768 (SPECTRA)	N/A	Variación según tamaño de biblioteca y filtros del estudio.

Fuente: Elaboración propia a partir de los datos de CoMMpass y GDC (2025).

El portal *Genomic Data Commons* (GDC), es el repositorio del *National Cancer Institute* para datos de cáncer armonizados y accesibles mediante acceso web y API, con control de acceso a ficheros protegidos y trazabilidad de *pipelines* de armonización (Colaprico et al., 2016).

En este trabajo se emplea *TCGAbiolinks* (R/Bioconductor) como interfaz de alto nivel para consultar, descargar y preparar datos. El paquete permite filtrar por proyecto y tipo de dato, emparejar metadatos clínicos con expresión y estructurar la salida en objetos *SummarizedExperiment*, dejando rastro de versiones y parámetros, clave para la trazabilidad y reproducibilidad (Figura 2).

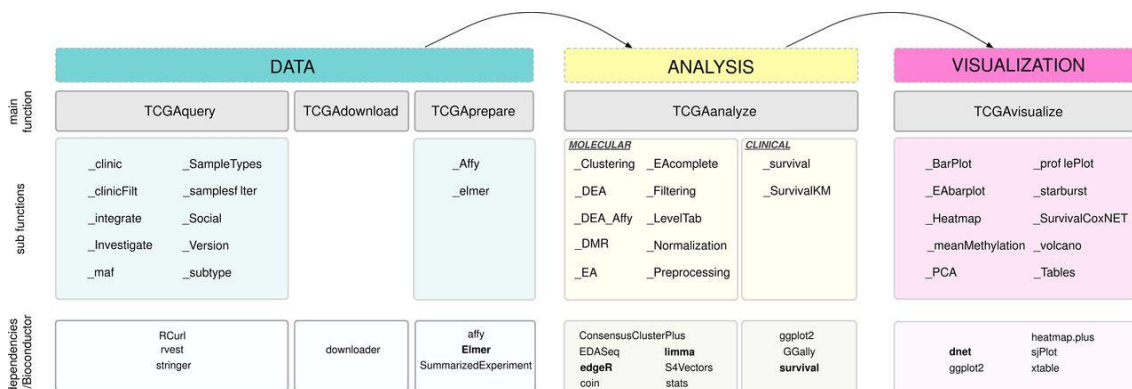


Figura 3. Visión general de las funciones de TCGAbiolinks

Figura extraída de Colaprico et al., 2016

4.2. Diseño y población

Se llevó a cabo un estudio observacional retrospectivo que integra datos transcriptómicos y clínicos procedentes de CoMMpass/GDC, con el objetivo de comparar el perfil de expresión tumoral entre pacientes <50 y ≥50 años en el debut y realizar su interpretación funcional mediante análisis de vías y redes. La población corresponde a los pacientes de la cohorte multicéntrica internacional (centros del consorcio). El periodo de la cohorte original abarca 2011–2016. Para el tamaño analítico, se adoptó un enfoque censal, incluyendo todas las muestras basales con edad y RNA-seq válidos en las MO de diagnóstico.

4.3. Recogida y preparación de datos

Los datos se obtuvieron del GDC mediante TCGAbiolinks (*GDCquery*, *GDCdownload*, *GDCprepare*), integrando la información clínica y de expresión por identificadores de muestra del GDC y depurando duplicados. A partir de los HTSeq-Counts del pipeline armonizado (alineamiento con STAR y cuantificación por gen con HTSeq-count), se construyó la matriz de conteos. El análisis se realizó con identificadores de ENSEMBL (IDs ENSEMBL) por su estabilidad y unicidad; los *símbolos HUGO Gene Nomenclature Committee* (HGNC) se emplearon solo para la presentación de resultados. Se aplicó un filtro de baja expresión (CPM > 1 en ≥2 muestras) y la normalización *Trimmed Mean of M-values* (TMM) (*edgeR*) (Chen Y et al., 2025). Sobre las log-CPM se empleó *voom* (*limma*) para estimar pesos de precisión y estabilizar la relación media–varianza del

RNA-seq (Robinson et al., 2010; Gene Ontology Consortium, 2015). El flujo de trabajo se muestra en la Figura 3.

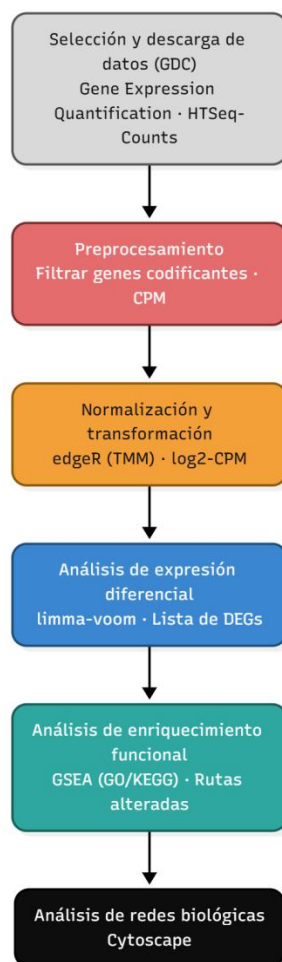


Figura 4. Pipeline de trabajo

4.4. Análisis estadístico

4.4.1. Expresión diferencial (DGE)

Se realizó un contraste de jóvenes (<50 años) frente a mayores (≥ 50 años). Se modeló cada gen con *limma* usando log-CPM y pesos *voom*; el modelo incluyó intercepto y un indicador de grupo de edad, tomando ≥ 50 años como referencia. Se aplicó eBayes (Ritchie ME et al, 2015), con tendencia para moderar varianzas y la corrección por múltiples contrastes se hizo con Benjamini–Hochberg (FDR). Se consideraron DGE, los genes con $FDR < 0,05$ y, un $|\log_2FC| \geq 1$. La salida incluyó la tabla completa (ENSEMBL con mapeo a símbolo HGNC, $\log_2FC$, t , p , FDR), un volcano-plot y un heatmap basado en *z-scores* por gen. Asimismo, para resumir el patrón diferencial por muestra, se calculó

una puntuación $z_{Up}-z_{Down}$. Primero se definió la vía para el contraste. Se empleó la vía KEGG “Cell cycle” (hsa04110), por la implicación en la oncogénesis que presenta (Wu D et al., 2020).

Posteriormente, se catalogaron conjuntos a partir del DGE (Up: genes más expresados en <50 ; Down: más expresados en ≥ 50). La expresión se estandarizó por gen (z-score) y, en cada muestra, se calculó la media de z de los genes Up menos la media de z de los genes Down. Una puntuación alta indica que la muestra se parece más al perfil “joven”; una baja/negativa, al perfil “ ≥ 50 ”. Esta puntuación se comparó entre grupos con Wilcoxon como verificación adicional del contraste.

4.4.2. Enriquecimiento funcional (GSEA)

A partir de los resultados de expresión diferencial, se generó un ranking de genes por el estadístico t con signo (jóvenes vs ≥ 50) y se aplicó GSEA preranked con *fgsea* para obtener el score de enriquecimiento y q-values en KEGG y Reactome (Kanehisa M. et al., 2025). La presentación de resultados se realizó con *dotplots* y *representación de pathways*.

4.5. Redes de interacción

Las interacciones proteína–proteína se obtuvieron de STRING usando, para cada subcohorte, únicamente los genes diferencialmente expresados. Con *stringApp* se importaron en Cytoscape (Shannon P. et al., 2003; Doncheva N.T. et al., 2019) y se construyó la red correspondiente. Sobre cada red se calcularon métricas topológicas con *NetworkAnalyzer* (por ejemplo, *degree* y *betweenness*), se identificaron módulos con *clusterMaker2* y se priorizaron nodos centrales con *cytoHubba*. Para la interpretación, se superpusieron atributos ómicos ($\log_2FC$, FDR y anotaciones) sobre nodos y aristas. Estas redes recogen conocimiento recopilado de bases de datos y evidencias reportadas; los módulos y nodos ayudan a contextualizar la firma, sin implicar direccionalidad ni causalidad.

4.6. Software y paquetes empleados

El estudio se realiza empleando el lenguaje de programación R (v. 3.4.1) mediante el entorno RStudio y la aplicación Cytoscape 3.10.4. Para ello, se hace uso de una serie de librerías (Tabla 2):

Tabla 2. Paquetes empleados en R

Librería	Fuente	Versión	Uso
DESeq2	Bioconductor	1.42.0	Expresión génica diferencial con RNA-Seq
edgeR	Bioconductor	3.42.4	Modelos de dispersión para RNA-Seq
limma	Bioconductor	3.62.2	Análisis con voom para RNA-Seq y microarrays
clusterProfiler	Bioconductor	4.14.6	Enriquecimiento funcional (GO, KEGG, Reactome, GSEA, ORA)
enrichplot	Bioconductor	1.26.6	Visualización de resultados de enriquecimiento
ReactomePA	Bioconductor	1.42.0	Enriquecimiento de rutas Reactome
DOSE	Bioconductor	4.0.1	Enriquecimiento de términos relacionados con enfermedades
fgsea	Bioconductor	1.32.4	Análisis de enriquecimiento GSEA rápido
org.Hs.eg.db	Bioconductor	3.20.0	Anotaciones de genes humanos
AnnotationDbi	Bioconductor	1.64.0	Acceso a bases de datos de anotación
GOSemSim	Bioconductor	2.32.0	Similitud semántica entre términos GO
RCy3	Bioconductor	2.26.0	Comunicación directa con Cytoscape desde R
igraph	CRAN	2.2.0	Análisis y visualización de redes
ggplot2	Tidyverse	4.0.0	Gráficos personalizados
dplyr	Tidyverse	1.1.4	Manipulación de datos
tidyr	Tidyverse	1.3.0	Transformación de estructuras de datos

5. Resultados

5.1. Diferencias ligadas a la edad en la aparición de MM: análisis por subcohortes

Se identificaron 995 pacientes con NDMM en la cohorte. De ellos, 762 disponían de RNA-seq basal utilizable tras la deduplicación por paciente (una muestra basal por MMRF_ID), que constituyeron la población de análisis para DGE, GSEA y redes. La edad estuvo disponible para todas las muestras con RNA-seq, y no se aplicaron exclusiones adicionales por calidad. En cuanto a los grupos por edad, se identificaron 94 pacientes <50 años y 668 pacientes ≥50 años (Figura 5).

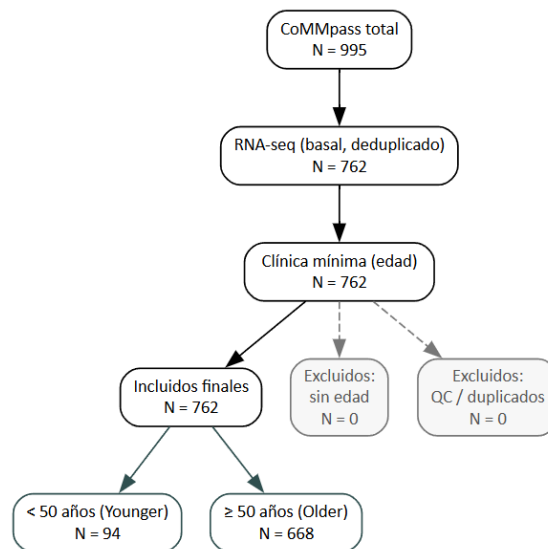


Figura 5. Diagrama de flujo de selección de pacientes

La distribución por sexo fue comparable entre grupos ($p=0.5548$). En ISS, las proporciones I/II/III no mostraron diferencias significativas ($p=0.1856$). En cuanto etnia, la distribución difirió entre grupos ($p=0.0031$), con menor proporción de caucásicos en el grupo más joven y mayor proporción relativa de afroamericanos (Tabla 3).

Tabla 3. Características demográficas basales

Características	<50 años (n=94)	≥50 años (n=668)	p-valor
Edad, mediana [IQR]	46.0 [42.0–48.0]	64.0 [59.0–71.0]	
Sexo			
Femenino (%)	41 (44%)	270 (40%)	0.55
Masculino (%)	53 (56%)	398 (60%)	
ISS			
I	41 (44%)	220 (34%)	0.18
II	30 (32%)	240 (37%)	
III	23 (24%)	188 (29%)	
Etnia			
Caucásico	46 (49%)	452 (67%)	0.003
Afroamericano	18 (19%)	79 (12%)	
Asiático	1 (1%)	11 (2%)	
No diposnible	29 (31%)	126 (19%)	

IQR: Rango intercuartil.

5.2. Análisis de expresión génica diferencial

Se evaluaron 26.198 genes en el contraste (*Younger (<50 años)* – *Older (≥50 años)*) tras el filtrado de baja expresión. Un total de 6.288 entradas carecían de anotación de tipo génico (GENETYPE), es decir, transcritos no codificantes o no mapeables en *org.Hs.eg.db*, por lo que no se consideraron en las proyecciones a vías (KEGG/Reactome) ni a interacciones proteína–proteína.

Tras la aplicación de los dos criterios elegidos para considerar a un gen como diferencialmente expresados de forma significativa, se identificaron 295 genes: 87 genes se mostraban sobreexpresados de forma significativa en jóvenes y 208, infraexpresados.

Para facilitar la lectura, se presentan una tabla resumida con los principales genes por dirección (log2FC y FDR) (Tabla 4), mientras que el listado completo de 295 DEG y la sublista de 141 codificantes se adjuntan en el Anexo 3 y 4 (.cvs).

Tabla 4. Top 20 genes más diferencialmente expresados en menores de 50 años

Gene	Direction	log2FC	t	FDR
SLC22A31	Upregulated	+1.47	4.09	0.0182
CFHR1	Upregulated	+1.04	3.87	0.0248
TUBB2B	Upregulated	+1.01	3.73	0.0318
LINC02466	Downregulated	-1.68	-5.25	0.00172
ENSG00000277156	Downregulated	-1.18	-5.43	0.00172
LOC101928087	Downregulated	-1.48	-5.13	0.00187
BDKRB1	Downregulated	-1.47	-5.12	0.00187
PCDHB3	Downregulated	-1.04	-5.10	0.00187
ENSG00000277128	Downregulated	-1.45	-5.02	0.00213
ENSG00000248685	Downregulated	-1.43	-4.99	0.00216
ENSG00000225210	Downregulated	-1.52	-4.89	0.00321
ENSG00000215270	Downregulated	-1.09	-4.78	0.00399
LOC101928622	Downregulated	-1.10	-4.64	0.00620
OVAAL	Downregulated	-1.67	-4.55	0.00766
DUXAP8	Downregulated	-1.64	-4.54	0.00766
LINC00922	Downregulated	-1.39	-4.55	0.00766
ENSG00000287136	Downregulated	-1.17	-4.50	0.00841
ENSG00000224899	Downregulated	-1.14	-4.41	0.0103
PTP4A3	Downregulated	-1.50	-4.38	0.0109
ENSG00000258927	Downregulated	-1.37	-4.38	0.0109

El análisis de componente principal, no separa los dos grupos; las muestras se solapan casi por completo. En consonancia, el número de DEG es moderado y los cambios de expresión, discretos. Así, las diferencias por grupo no explican la mayor parte de la variabilidad transcriptómica (Figura 6).

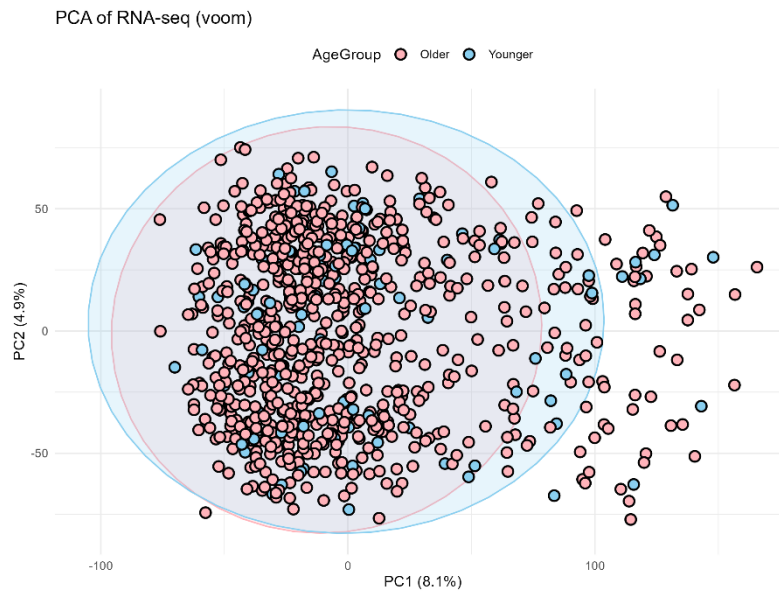


Figura 6. Análisis de componente principal (PCA)

En la figura 7, se resume el contraste de expresión génica entre los dos grupos. La nube de puntos se concentra alrededor de $\log_2\text{FC} \approx 0$ y por debajo del umbral de FDR, lo que indica que, globalmente, no hay cambios masivos asociados a la edad. Entre los genes que sí superan los filtros se aprecia una asimetría clara: predominan los situados en el hemisferio negativo, es decir, con mayor expresión en el grupo de mayor edad, mientras que solo un pequeño número aparece aumentado en los jóvenes. Además, los tamaños de efecto de los genes significativos son en su mayoría moderados y cercanos al umbral de $|\log_2\text{FC}|=1$, sin colas extremas, lo que sugiere diferencias finas más que reprogramaciones amplias (Figura 7).



Figura 7. Volcano plot. El eje X muestra el cambio de expresión ($\log_2FC$, jóvenes vs ≥ 50 ; positivo = mayor en jóvenes) y el eje Y la significación ($-\log_{10} FDR$). Se resaltan como DEG los genes con $FDR < 0,05$ y $|\log_2FC| \geq 1$ (resto en gris), con líneas de referencia en $|\log_2FC| = 1$ y $FDR = 0,05$; se anotan los genes más destacados por FDR o magnitud del efecto.

Otra forma de representación es el heatmap, construido con todos los genes retenidos tras el filtrado (no solo los significativos), escalados por z-score por gen y con *clustering* en genes y muestras. La variabilidad interindividual dentro de cada grupo es evidente y sugiere desplazamientos finos entre grupos, no un reprogramado masivo, en coherencia con la asimetría descrita previamente (Figura 8).

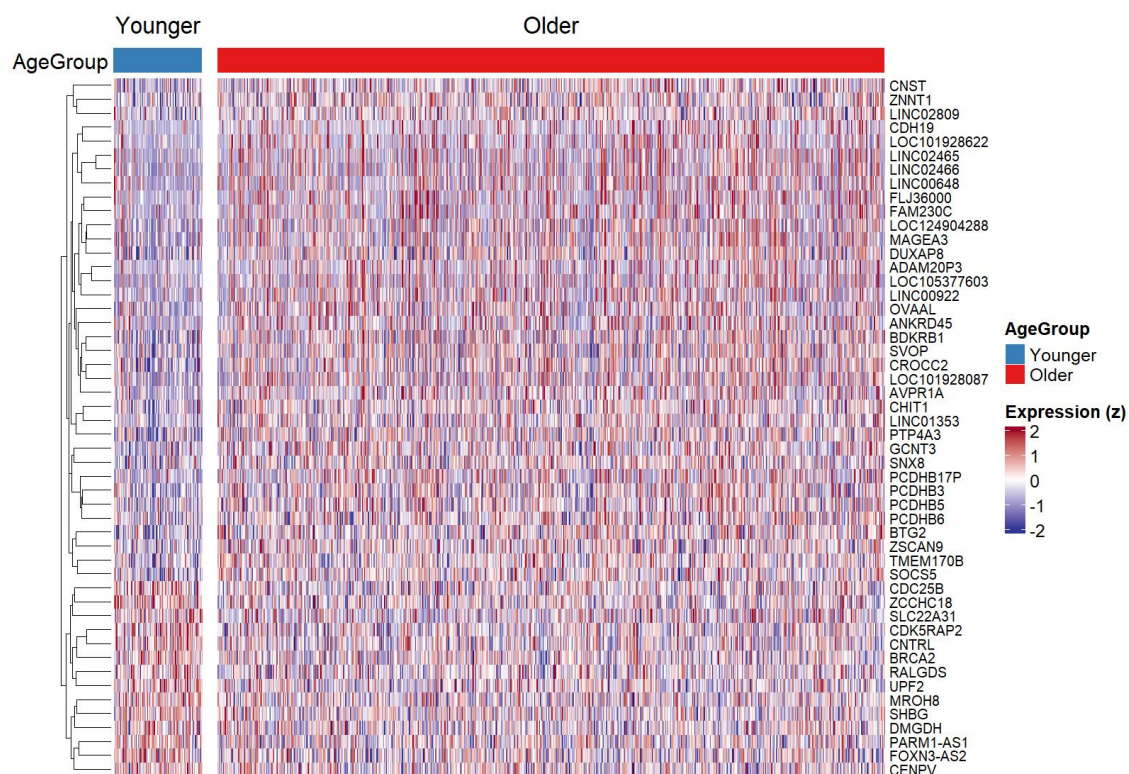


Figura 8. Heatmap de expresión global por grupos de edad

5.3 Puntuación zUp–zDown (síntesis del DGE)

Para poder resumir una “firma” de expresión, se ha realizado un *análisis zUp–zDown*. Dicho constructo permite sintetizar el patrón diferencial identificado por el DGE.

En la figura 9, se resume por muestra el peso relativo de dos conjuntos definidos por el DGE: genes Up (con mayor expresión en <50) y genes Down (con mayor expresión en ≥ 50). Un valor positivo indica predominio del patrón Up y un valor negativo, del Down.

En nuestros datos, que evalúa el score de “Cell Cycle” y proliferación celular, la distribución de la puntuación se desplaza al alza en <50 y a la baja en ≥50, con diferencia significativa entre grupos (Wilcoxon), lo que confirma el desplazamiento global observado en el análisis gen-a-gen.

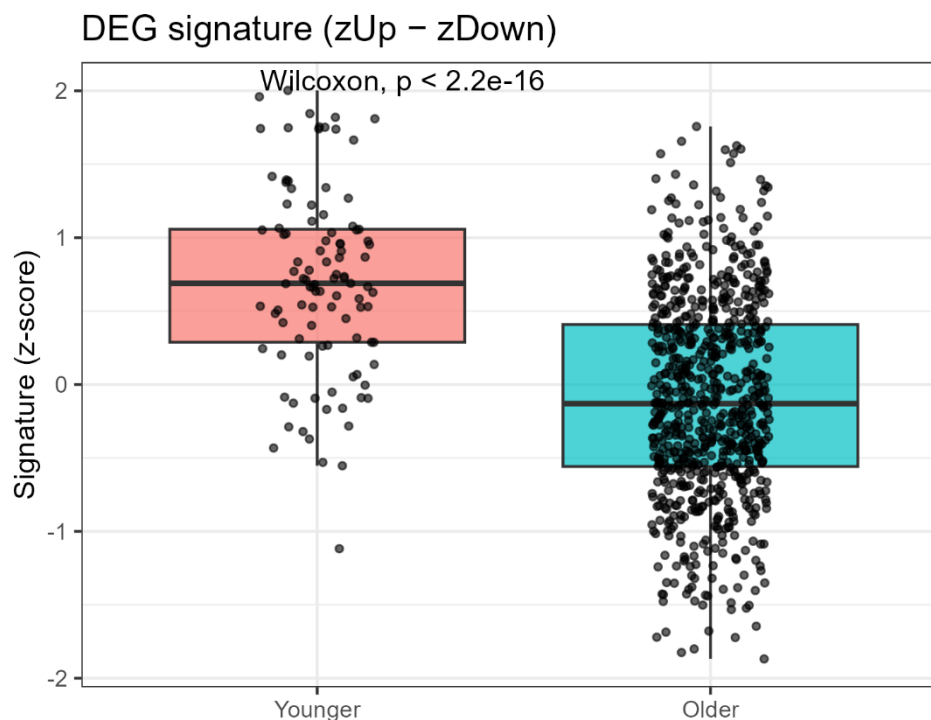


Figura 9. Firma de expresión de proliferación (KEGG “Cell cycle”) por grupo. Eje Y: score ssGSEA (single-sample Gene Set Enrichment Analysis), de la vía KEGG Cell cycle (centrado y escalado). Eje X: grupo de edad. Cada punto es una muestra; cajas: mediana e IQR. Se muestra el p de Wilcoxon.

5.4. Detección de funciones sobre e infrarepresentadas en el transcrito de los dos subcohortes

Con el contraste ya definido a nivel génico, el siguiente paso fue interpretar la señal de expresión en términos de vías moleculares. Para ello se realizó GSEA preranked sobre el listado completo ordenado por el estadístico t de *lima*.

El *Nature Enrichment Score* (NES) mide si los genes de una vía se acumulan de forma no aleatoria en alguno de esos extremos, normalizando por el tamaño del conjunto. Así, $NES > 0$ indica que la vía está enriquecida hacia la parte alta del ranking (patrón coordinado a favor de <50 años) y $NES < 0$ que lo está hacia la parte baja (patrón coordinado a favor de ≥50 años). Importa subrayar que el signo del NES refleja la

dirección de una diferencia de expresión coordinada; no implica, por sí mismo, una “activación” o “inhibición” bioquímica causal de la vía.

El análisis detectó 74 vías significativas (FDR<0.05), en el repositorio de KEGG, distribuidas en dos polos biológicos:

- Procesos mitóticos y de dinámica cromosómica con más señal hacia los casos más jóvenes.
- Bloque inmune/adaptativo y de presentación antigénica más marcado en el grupo de mayor edad.

A continuación, en la tabla 5 se nombran las vías más enriquecidas en pacientes jóvenes, las cuales cuentan con NES negativo (más sobreexpresadas en el grupo de referencia, los cuales corresponde a los pacientes mayores).

Tabla 5. Top 10 rutas enriquecidas de KEGG

ID	Descripción	NES	FDR	Size
hsa05150	Staphylococcus aureus infection	-2.255	0.000000263	67
hsa05322	Systemic lupus erythematosus	-2.099	0.000000369	115
hsa05332	Graft-versus-host disease	-2.172	0.000016900	36
hsa05320	Autoimmune thyroid disease	-2.099	0.000043000	34
hsa04612	Antigen processing and presentation	-2.064	0.000043000	66
hsa04080	Neuroactive ligand-receptor interaction	-1.773	0.000043000	259
hsa05033	Nicotine addiction	-2.074	0.000057200	34
hsa05330	Allograft rejection	-2.087	0.000096500	33
hsa05323	Rheumatoid arthritis	-1.978	0.000096500	86
hsa04060	Cytokine-cytokine receptor interaction	-1.703	0.000149000	238

Top 10 KEGG (GSEA). NES: normalized enrichment score; FDR: p.adjust.

Estas mismas vías de KEGG enriquecidas se pueden representar con el dotplot de *clusterProfiler*. El dotplot muestra, para cada vía, la significación ($-\log_{10}$ FDR), el tamaño del conjunto (tamaño del punto) y el NES (color), facilitando la comparación visual entre términos (Fig. 10).

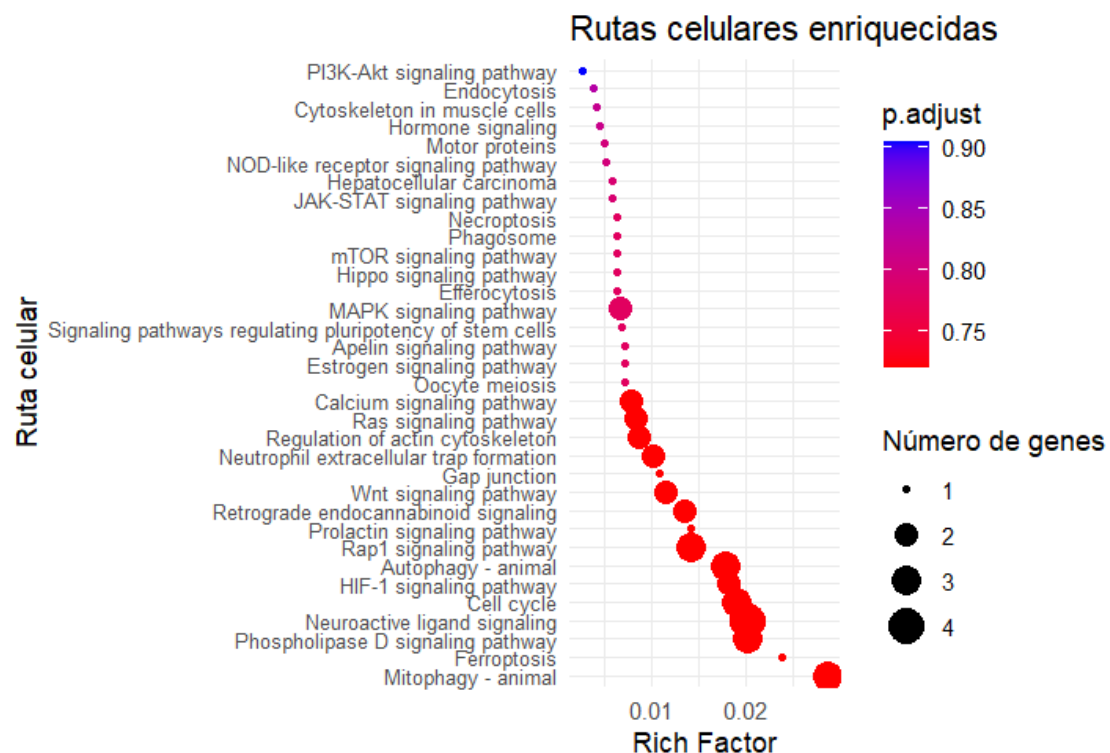


Figura 10. Dotplot de vías celulares (Cellular processes). En el GeneRatio, se refleja la fracción de genes de esa vía que contribuyen a la señal y el tamaño de la burbuja indica cuántos genes (Count) participan. El color, codifica la significación.

Por otro lado, KEGG ilustra gráficamente la localización de los genes implicados dentro de cada vía en un mapa que representa los *pathways* celulares (Figura 11). La presencia de múltiples nodos coloreados dentro de la misma ruta indica una alteración concertada de su regulación transcripcional, un patrón compatible con procesos de oncogénesis y relevante en el mieloma múltiple, donde la desregulación de redes de señalización y control del ciclo celular puede favorecer proliferación y resistencia.

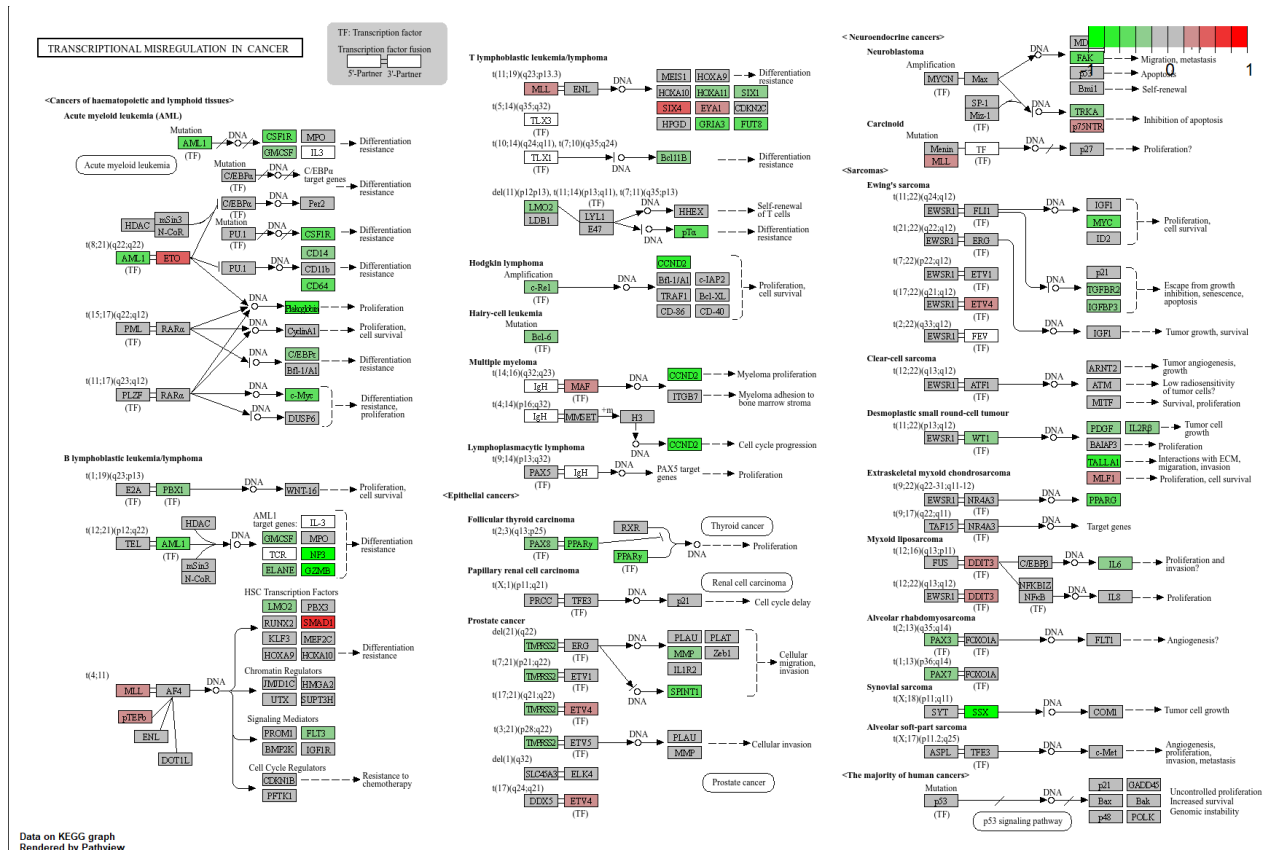


Figura 11. Activación diferencial de genes implicados en la vía “Transcriptional misregulation in cancer”. Los genes sobreexpresados en el grupo <50 años se muestran en rojo y los subexpresados en verde

5.5. Análisis de redes y firma de expresión

Se proyectó el conjunto completo de genes diferencialmente expresados, de forma significativa, en Cytoscape mediante *stringApp* (cutoff de confianza $\geq 0,40$; sin interactores añadidos). Para centrar la interpretación en relaciones de interés, se eliminaron nodos aislados; los colores de nodo codifican el $\log_2FC$ (rojo: mayor expresión en <50 años; azul: mayor en ≥ 50) y el grosor de arista refleja la confianza de STRING (Figura 12).

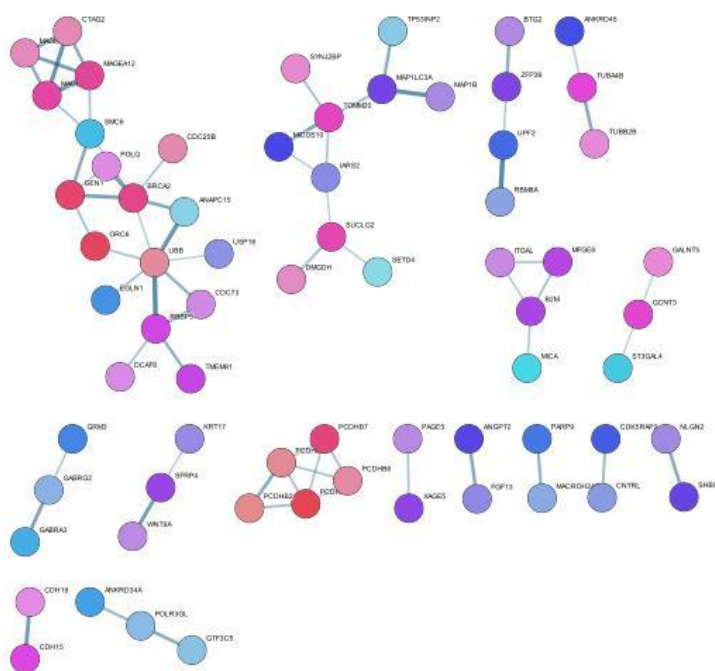


Figura 12. Red de interacciones de genes DEG significativos. Los nodos se colorean por cambio de expresión: rojo indica mayor expresión en <50 y azul en ≥50 años; la intensidad del color refleja la magnitud ($|\log_2FC|$).

La red resultante muestra varios componentes conectados, destacando los genes de replicación (BRCA2, ORC6, CDC25B, RBBP5, ANAPC15, UBB, con conexiones a GEN1 y POLQ). La mayor parte de estos nodos aparece en rojo (más altos en <50), consistente con un programa proliferativo/replicativo más activo en el grupo joven.

También, observamos genes involucrados en la estabilidad del ARNm (conjunto ZFP36–UPF2–RBM8A y adyacentes), metabolismo energético (TOMM20–IARS2–DMGDH–SUCLG2), y de la presentación antigénica (ITGAL–B2M–MFG8, RGS1 y vecinos) (Tabla 6).

Tabla 6. Genes más significativos de la red (StringApp)

Gen	LogFC	P.ajustado	Tipo.de.gen	Ruta.molecular.enriquecida
CD160	1.75	0.0041	Inmunológico	Activación de linfocitos T
SLAMF6	2.12	0.0034	Inmunológico	Interacciones linfocito-linfocito
NCR3	1.63	0.0056	Inmunológico	Activación células NK
LILRA1	-1.78	0.0047	Inmunológico	Señalización de receptores inmunes
PTGS1	2.05	0.0068	Metabólico	Metabolismo del ácido araquidónico
ACER3	-1.91	0.0055	Metabólico	Metabolismo de esfingolípidos
SUCLG2	1.89	0.0032	Metabólico	Metabolismo del propanoato
CHIT1	2.20	0.0028	Metabólico	Metabolismo de aminoazúcares

Genes diferencialmente expresados y sus rutas moleculares enriquecidas (GSEA/KEGG).

Para complementar el análisis de redes, se generó una segunda vista incluyendo de todo el set de genes analizados, coloreados por el estadístico t de *limma* (ranking completo).

La estructura de la red presenta un núcleo compacto con alta conectividad donde convergen genes de adhesión/inmunidad (p. ej., ICAM1, SELL, HLA-E, FCGR3A, KLRD1 y receptores relacionados). El predominio de tonos naranja-rojo están asociados al grupo ≥ 50 años, los cuales corresponden al módulo inmune/adhesivo, y tonos verde-amarillo, al grupo < 50 años, visibles en subredes ligadas a división/replicación (Figura 13).

Red de interacciones (STRING) - Genes enriquecidos

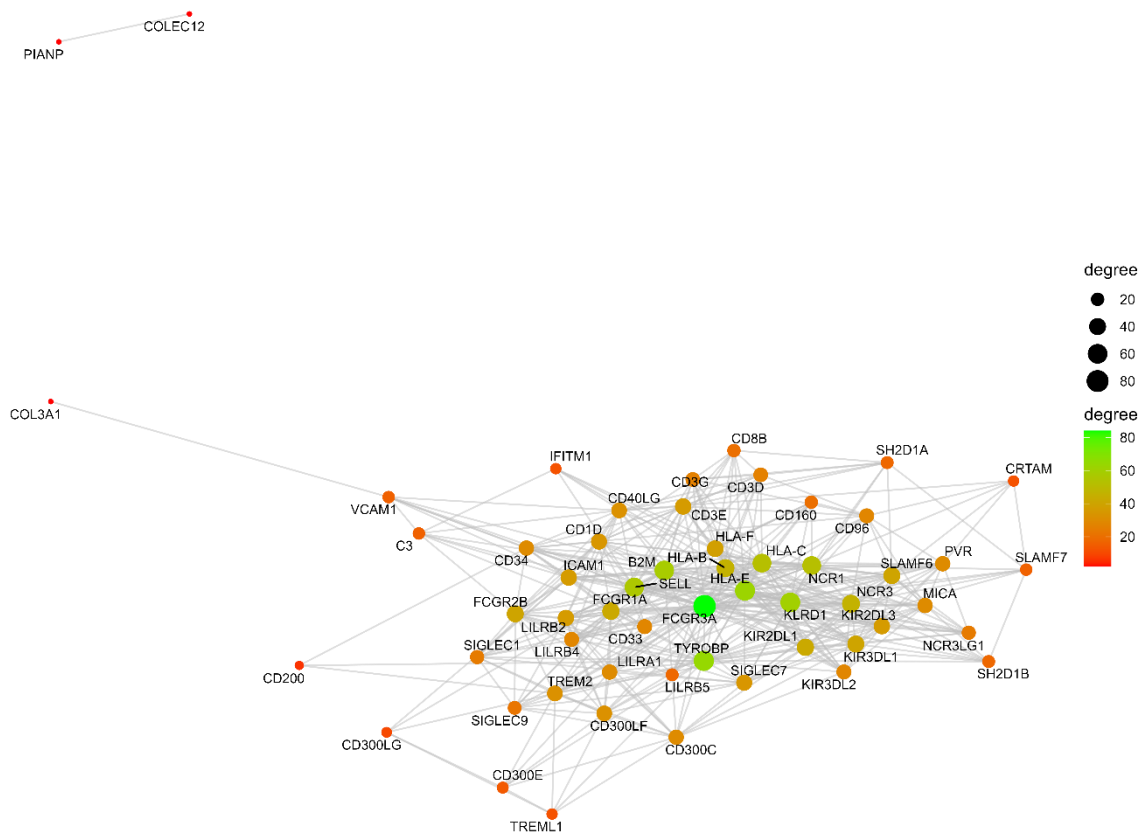


Figura 13. Red de interacciones de genes enriquecidos. Los nodos representan genes pertenecientes a vías KEGG enriquecidas; el color identifica la vía/módulo y la saturación refleja la magnitud del $\log_2 FC$. El tamaño del nodo es proporcional a su centralidad (degree). Las aristas indican interacciones PPI (STRING); si aparece grosor, es proporcional al score de la interacción.

6. Discusión

En nuestro estudio, empleamos una cohorte amplia y armonizada del proyecto CoMMpass/MMRF descargada desde el repositorio público GDC. Identificamos 762 muestras de MO con datos clínicos disponibles. Realizamos una comparación del transcriptoma tumoral entre pacientes < 50 y ≥ 50 años en el momento del diagnóstico. El análisis DGE identificó 295 genes diferencialmente expresados, de los que 141 fueron

codificantes. A partir de este contraste génico, realizamos GSEA y ORA sobre los repositorios públicos KEGG y GO. Posteriormente, proyectamos la firma a una red de asociaciones funcionales empleando la herramienta Cytoscape para contextualizar módulos funcionales.

En conjunto, los resultados apuntan a distintas huellas transcriptómicas entre los dos grupos: En pacientes más jóvenes, predomina una firma proliferativa (procesos de ciclo celular y algunos genes asociados a estructura del citoesqueleto), y en el grupo de más edad, se observa una sobrerrepresentación de ejes inmunes (antígeno–presentación, redes de receptores inmunitarios y vías autoinmunes).

Entre el grupo de más edad, observamos mayor expresión relativa de varios antígenos “*cancer-testis*” (CTA), como: MAGEA3, MAGEA1 y CTAG2/NY-ESO-2. Los CTA están bien descritos en mieloma múltiple: MAGE-C1/CT7 y MAGE-C2/CT10 muestran alta prevalencia y asociación con agresividad/recurrencia, y NY-ESO-1/LAGE-1 se ha utilizado como diana en estrategias inmunoterapéuticas (vacunas y linfocitos T con receptor de antígeno redirigido) específicamente en MM (Atanackovic D et al., 2010).

En los menores, los incrementos más consistentes incluyen SLC22A31, CFHR1 y TUBB2B, junto con otros genes implicados en ciclo y organización del huso mitótico. Aunque varios genes con log2FC positivo (p. ej., CLIP4, CITED4, MAP1B) no alcanzaron significación ajustada, el patrón global fue apoyado por GSEA sobre el ranking completo.

En el enriquecimiento, se identificó conjuntos de genes que actúan de forma coordinada en diferentes funciones celulares según el grupo de edad. Se ha identificado sobre-representación de ejes mitóticos/ciclo entre los DGE “positivos” del grupo joven, y de procesamiento/presentación de antígenos en el grupo mayor.

La firma de los más jóvenes, es congruente con el patrón clásico que, históricamente, han definido subtipos proliferativos en MM (p. ej., perfiles con alto índice proliferativo en *microarrays*) (Corchete Sánchez LA. 2016). Aunque nuestro diseño no subdivide por subtipos de CoMMpass, podría ser plausible que en el estrato joven haya mayor peso de casos con mayor índice proliferativo. Estudios previos, mostraron que un índice proliferativo elevado se asocia a peor pronóstico en MM, independientemente de la edad (Alexandrakis, MG et al., 2004).

Respecto al análisis de redes, pone de relieve dos bloques: uno de ciclo celular y replicación (p. ej., BRCA2, ORC6, CDC25B, SMC8, POLQ), con conectividad alta y gradientes de color acordes al log2FC del contraste, en los menores. Y otro, un módulo inmune (HLA-E, FCGRs, KLRD1/KIRs, SELL, CD48, etc.) con aristas múltiples y enriquecimiento relativo en ≥ 50 años, alineado con el GSEA.

Un anterior trabajo análisis transcriptómico de CoMMpass (Skerget S et al., 2024) consolidó 12 subtipos transcriptómicos estables que se alinean con alteraciones genómicas clásicas (hiperdiploidía, translocaciones del locus IGH, etc., mencionados anteriormente). Nuestro objetivo no era reclasificar por subtipo, sino intentar

estratificar por edad e intentar indagar los mecanismos que subyacen a una presentación tan precoz del MM.

Una fortaleza técnica de este trabajo es el uso de datos armonizados del GDC con pipeline documentado (*STAR-Counts*) y descarga reproducible (*TCGAbiolinks*), lo que minimiza heterogeneidad preanalítica. Además, al centrarnos en muestras basales y realizar un enfoque censal (todas las que cumplen criterios), reducimos sesgo de selección. La combinación *limma-voom*, DGE, GSEA y posteriormente, análisis de redes con *Cytoscape (stringApp)* se ajusta a estándares actuales y está ampliamente respaldada por la comunidad (Law CW et al., 2016; Doncheva NT et al., 2019).

Respecto a las limitaciones, el estudio muestra una diferencia del número de sujetos entre grupos, lo que requiere de validación externa. Por otra parte no se ha realizado un análisis estratificado por etnia, hallando en las características basales, una diferencia con más población afroamericana en pacientes jóvenes. Además, no se puede descartar por completo que en la MO enriquecida en CD138+, no se halle una contaminación residual, lo que puede contribuir a la señal inmune. Por otra parte, el acceso a datos clínicos es muy limitado: no se ha tenido acceso a datos de tratamientos, resultados ni supervivencia. R-ISS tampoco estaba categorizado ni se disponía de citogenética. Por ello, no se puede realizar ninguna asociación de firmas con *outcomes*. La generalización por edad es operativa, pero la edad biológica y la mezcla de subtipos pueden modular el efecto.

Por último, en la representación de interacciones de genes por debajo de los criterios de significación para aumentar la cobertura, el ruido crece. Por ello, vinculamos su interpretación a la coherencia con GSEA.

En cuanto a pasos futuros, se podría valorar la validación externa del estudio, y completar la GSEA con *leading-edge analysis* para identificar los genes núcleo de cada vía y priorizar candidatos. Así como integrar el análisis de redes para ubicar los puntos de convergencia transcripcional.

Evaluar la puntuación *zUp-zDown* (y perfiles por muestra con *ssGSEA*) en otras “*releases*” de CoMMpass o en cohortes externas. Cuando exista seguimiento, explorar valor pronóstico con modelos de Cox ajustados por ISS/citogenética.

Por último, sería interesante evaluar el impacto en el valor pronóstico con modelos de Cox ajustados y respuesta al tratamiento.

7. Conclusiones

- a) La edad al diagnóstico no explica por sí sola la variación global del transcriptoma en MM, como refleja el solapamiento de las muestras en PCA y la magnitud moderada de los efectos.

- b) Aun así, se observan diferencias funcionales reproducibles entre grupos de edad, detectables mediante análisis de conjuntos génicos y redes.
- c) En <50 años predomina un patrón proliferativo, con mayor actividad de ciclo celular/replicación y procesos ligados a dinámica cromosómica.
- d) En ≥50 años emerge un módulo inmune-adhesivo, con mayor señal de presentación antigénica y adhesión/tráfico celular.
- e) La convergencia de DGE, GSEA y redes PPI sostiene una dicotomía “proliferativa vs. inmune” ligada a la edad, sin evidencias de reprogramación masiva.

8. Bibliografía

1. Rajkumar, S. V., Dimopoulos, M. A., Palumbo, A., Blade, J., Merlini, G., Mateos, M. V., Kumar, S., Hillengass, J., Kastiris, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., ... Miguel, J. F. (2014). International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet. Oncology*, 15(12), e538–e548. [https://doi.org/10.1016/S1470-2045\(14\)70442-5](https://doi.org/10.1016/S1470-2045(14)70442-5)
2. Walker, B. A., Mavrommatis, K., Wardell, C. P., Ashby, T. C., Bauer, M., Davies, F. E., Rosenthal, A., Wang, H., Qu, P., Hoering, A., Samur, M., Towfic, F., Ortiz, M., Flynt, E., Yu, Z., Yang, Z., Rozelle, D., Obenauer, J., Trotter, M., Auclair, D., ... Morgan, G. J. (2018). Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood*, 132(6), 587–597. <https://doi.org/10.1182/blood-2018-03-840132>
3. Van de Donk, N. W., Pawlyn, C., & Yong, K. L. (2021). *Multiple myeloma*. *The Lancet*, 397(10272), 410–427. [https://doi.org/10.1016/S0140-6736\(21\)00135-5](https://doi.org/10.1016/S0140-6736(21)00135-5)
4. Awada, H., Thapa, B., Awada, H., Dong, J., Gurnari, C., Hari, P., & Dhakal, B. (2021). *A Comprehensive Review of the Genomics of Multiple Myeloma: Evolutionary Trajectories, Gene Expression Profiling, and Emerging Therapeutics*. *Cells*, 10(8), 1961. <https://doi.org/10.3390/cells10081961>
5. Tanguay, M., LeBlanc, R., Ahmad, I., & Claveau, J.-S. (2023). *Young adults with multiple myeloma: A systematic review of manifestations, biology, and outcomes*. *Cancers*, 15(4), 1239. <https://doi.org/10.3390/cancers15041239>
6. Ludwig, H., Durie, B. G. M., Bolejack, V., Turesson, I., Kyle, R. A., Blade, J., ... San Miguel, J. (2008). *Myeloma in patients younger than age 50 years presents with more favorable features and shows better survival: an analysis of 10 549 patients from the International Myeloma Working Group*. *Blood*, 111(8), 4039–4047. <https://doi.org/10.1182/blood-2007-03-081018>

7. Rojas, E. A., Corchete, L. A., Mateos, M. V., García-Sanz, R., Misiewicz-Krzeminska, I., & Gutiérrez, N. C. (2019). *Transcriptome analysis reveals significant differences between primary plasma cell leukemia and multiple myeloma even when sharing a similar genetic background*. Blood Cancer Journal, 9(12), 90. <https://doi.org/10.1038/s41408-019-0253-1>
8. Stark, R., Grzelak, M., & Hadfield, J. (2019). *RNA sequencing: the teenage years*. Nature Reviews Genetics, 20(11), 631–656. <https://doi.org/10.1038/s41576-019-0150-2>
9. Jang, J. S., Li, Y., Mitra, A. K., Bi, L., Abyzov, A., van Wijnen, A. J., et al. (2019). *Molecular signatures of multiple myeloma progression through single cell RNA-Seq*. Blood Cancer Journal, 9(1), 2. <https://doi.org/10.1038/s41408-018-0160-x>
10. Gene Ontology Consortium, Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., Van Auken, K., ... Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1), iyad031. <https://doi.org/10.1093/genetics/iyad031>
11. Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein science : a publication of the Protein Society*, 29(1), 28–35. <https://doi.org/10.1002/pro.3711>
12. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. *Nucleic acids research*, 53(D1), D672–D677. <https://doi.org/10.1093/nar/gkae909>
13. Keskin, M. (2022). *Comparison of Transcriptomic Changes in Younger and Older Multiple Myeloma Patients from the MMRF-CoMMpass Study RNA-Seq Data*. bioRxiv. <https://doi.org/10.1101/2022.09.14.507921>
14. Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data*. Nucleic Acids Research, 44(8), e71. <https://doi.org/10.1093/nar/gkv1507>
15. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L., & Smyth, G. K. (2025). *edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets*. Nucleic Acids Research, 53(2), gkaf018. <https://doi.org/10.1093/nar/gkaf018>
16. The Gene Ontology Consortium (2015). *Gene Ontology Consortium: going forward*. Nucleic Acids Research, 43(D1), D1049–D1056. <https://doi.org/10.1093/nar/gku1179>
17. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). *limma powers differential expression analyses for RNA-sequencing and*

microarray studies. *Nucleic acids research*, 43(7), e47.
<https://doi.org/10.1093/nar/gkv007>

18. Wu, D., Zhang, P., Li, F., Shen, Y., Chen, H., Feng, Y., He, A., & Wang, F. (2020). CD138⁺ multiple myeloma cells express high level of CHK1 which correlated to overall survival in MM patient. *Aging*, 12(22), 23067–23081.
<https://doi.org/10.18632/aging.104066>

19. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>

20. Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of proteome research*, 18(2), 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>

21. Atanackovic, D., Hildebrandt, Y., Jadcak, A., Cao, Y., Luetkens, T., Meyer, S., et al. (2010). Cancer-testis antigens *MAGE-C1/CT7* and *MAGE-A3* promote the survival of multiple myeloma cells. *Haematologica*, 95(5), 785–793.
<https://doi.org/10.3324/haematol.2009.014464>

22. Corchete Sánchez, L. A. (2016). *Expresión génica en mieloma múltiple: análisis de datos de RNA-seq y microarrays en combinación con estudios de metaanálisis y predicción de respuesta al tratamiento*. [Tesis doctoral, Universidad de Salamanca]. Repositorio Documental Gredos. <https://gredos.usal.es/handle/10366/130978>

23. Alexandrakis, M. G., Passam, F. H., Kyriakou, D. S., Dambaki, K., Niniraki, M., & Stathopoulos, E. (2004). *Ki-67 proliferation index: correlation with prognostic parameters and outcome in multiple myeloma*. *American Journal of Clinical Oncology*, 27(1), 8–13. <https://doi.org/10.1097/01.coc.0000045810.91816.4>

24. Skerget, S., Penaherrera, D., Chari, A., Jagannath, S., Siegel, D. S., Vij, R., et al. (2024). *Comprehensive molecular profiling of multiple myeloma identifies refined copy number and expression subtypes*. *Nature Genetics*, 56(9), 1878–1889.
<https://doi.org/10.1038/s41588-024-01853-0>

25. Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., & Ritchie, M. E. (2016). *RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR*. *F1000Research*, 5, ISCB Comm J-1408. <https://doi.org/10.12688/f1000research.9005.3>

9. Anexos

Anexo 1. Índice de tablas:

- p. 9 — Tabla 1. Disponibilidad de casos y RNA-seq basal.*
- p. 12 — Tabla 2. Paquetes empleados en R.*
- p. 14 — Tabla 3. Características demográficas basales.*
- p. 15 — Tabla 4. Top 20 genes más diferencialmente expresados en menores de 50 años.*
- p. 19 — Tabla 5. Top 10 rutas enriquecidas de KEGG*
- p. 22 — Tabla 6. Genes más significativos de la red (StringApp)*

Anexo 2. Índice de figuras

- p. 7 — Figura 1. Componentes de KEGG.*
- p. 9 — Figura 2. Infografía del estudio MMRF-CoMMpass.*
- p. 10 — Figura 3. Visión general de las funciones de TCGAbiolinks.*
- p. 11 — Figura 4. Pipeline de trabajo.*
- p. 14 — Figura 5. Diagrama de flujo de selección de pacientes.*
- p. 16 — Figura 6. Análisis de componente principal (PCA).*
- p. 17 — Figura 7. Volcano plot... (leyenda completa en la página).*
- p. 17 — Figura 8. Heatmap de expresión global por grupos de edad.*
- p. 18 — Figura 9. Firma de expresión de proliferación (KEGG “Cell cycle”) por grupo.*
- p. 20 — Figura 10. Dotplot de vías celulares (Cellular processes).*
- p. 21 — Figura 11. Activación diferencial de genes implicados en la vía “Transcriptional misregulation in cancer”.*
- p. 22 — Figura 12. Red de interacciones de genes DEG significativos.*
- p. 23 — Figura 13. Red de interacciones de genes enriquecidos.*

Anexo 3. Material complementario:

Archivo Resultados_DEG.csv

Anexo 4. Material complementario:

Archivo Genes_codificantes_significativos_unicos.csv

Anexo 5. Código

El código desarrollado para este trabajo se encuentra disponible en GitHub, accesible mediante el siguiente enlace: https://github.com/mhidalgosoto92-debug/TFM_MartaHidalgo