



Máster en Bioinformática

Inferencia de CNVs en datos de scRNAseq en cáncer de ovario seroso de alto grado

Autor: Yacine Ibelaid Faid

Tutor: Matías Marín Falco

Curso 2023-24

Agradecimientos

A Matías, por su total predisposición y paciencia en este proceso. Gracias por todo tu apoyo. A mi familia y a mis seres queridos, quienes siempre me han acompañado y respaldado en mi camino en la ciencia, siendo mi pilar incondicional. Todo esto no habría sido posible sin vosotros.

RESUMEN

El cáncer de ovario tipo seroso de alto grado (HGSOC) es uno de los cánceres ginecológicos con peor pronóstico y con tasas de diagnóstico en estadios más avanzados. La heterogeneidad intratumoral que caracteriza a este cáncer implica un desafío en el estudio de mecanismos de resistencia a los tratamientos. Gran parte de las pacientes con cáncer de ovario desarrollan resistencia al tratamiento más utilizado, la quimioterapia basada en platino. En este estudio, aplicamos la herramienta inferCNV con el objetivo de identificar los subclones más relevantes y su comportamiento durante el tratamiento mediante la inferencia de variantes de número de copia (CNVs) a partir de datos de secuenciación de RNA en células individuales.

El conjunto de datos utilizado se obtuvo de muestras de pacientes con HGSOC de GEO (GSE165897), e incluyó muestras de 11 pacientes antes y después del tratamiento. Al enfrentar las células tumorales con las células de referencia identificamos poblaciones de subclones con patrones de ganancia o pérdida, que proporcionan una visión más profunda del crecimiento tumoral bajo el estrés del tratamiento. Asimismo, mediante el análisis de expresión diferencial y el enriquecimiento funcional se identificaron genes implicados en la evolución clonal, así como los procesos biológicos más relevantes en subclones con mecanismos de resistencia o de sensibilidad al tratamiento.

El estudio muestra el valor de la integración de la inferencia de CNVs en scRNAseq para identificar subclones tumorales y su comportamiento tras el tratamiento. Mediante trabajos que incorporen análisis multi-ómicos se podrían conseguir resultados más precisos.

Palabras clave: inferCNV, scRNAseq, cáncer de ovario, quimiorresistencia, HGSOC

ABSTRACT

High-grade serous ovarian cancer (HGSOC) is one of the gynecological cancers with the poorest prognosis and the highest rates of diagnosis at advanced stages. The intratumoral heterogeneity that characterizes this cancer poses a challenge in studying resistance mechanisms to treatments. A large proportion of ovarian cancer patients develop resistance to the most commonly used treatment, platinum-based chemotherapy. In this study, we applied the inferCNV tool with the aim of identifying the most relevant subclones and their behaviour during treatment through the inference of copy number variations (CNVs) from single-cell RNA sequencing data.

The dataset used was obtained from HGSOC patient samples from GEO (GSE165897), including samples from 11 patients before and after treatment. By comparing tumour cells with reference cells, we identified subclone populations with gain or loss patterns, providing deeper insights into tumour growth under treatment stress. Additionally, differential expression and functional enrichment analyses identified genes involved in clonal evolution, as well as the most relevant biological processes in subclones exhibiting resistance or sensitivity mechanisms to the treatment.

This study highlights the value of integrating CNV inference in scRNAseq to identify tumour subclones and their behaviour after treatment. More accurate results could be achieved by incorporating multi-omics analyses in future research.

Key words: inferCNV, scRNAseq, ovarian cancer, HGSOC, chemoresistance.

ÍNDICE GENERAL

Índice de figuras y tablas	6
Índice de figuras en anexo	8
0. ABREVIATURAS.....	9
1. INTRODUCCIÓN.....	11
1.1 Cáncer de ovario: una visión general.....	11
1.2 Secuenciación de RNA en células individuales: revelando la heterogeneidad tumoral ...	17
1.3 Inferencia de CNVs en scRNAseq: bioinformática y su aplicación	18
2. HIPÓTESIS.....	21
2.1 Hipótesis principal.....	21
2.1 Hipótesis secundarias	21
3. OBJETIVOS.....	21
3.1 Objetivo general.....	21
2.2 Objetivos específicos.....	22
4. MATERIALES Y MÉTODOS	22
4.1 Búsqueda de información bibliográfica	22
4.2 Plan de trabajo y tareas	23
4.3 Cronograma	24
4.4 Origen y manejo de los datos	25
4.4.1 InferCNV: Herramienta para la inferencia de CNVs.....	27
4.4.2 Estudio bioinformático en RStudio	29
5. RESULTADOS	33
5.1 Análisis de CNVs inferidas a partir de scRNAseq	33
5.2 Heterogeneidad tumoral y clusterización de subclones en HGSOc	35
5.3 Análisis de expresión diferencial con Seurat	40
5.4 Enriquecimiento funcional.....	43
5. DISCUSIÓN Y CONCLUSIONES	46
8. DECLARACIÓN DEL USO DE HERRAMIENTAS DE IA.....	52
9. BIBLIOGRAFÍA.....	53
9.1 Artículos y capítulos de libros	53
9.2 Bases de datos, herramientas bioinformáticas y páginas web.....	58
10. ANEXOS	59

Índice de figuras y tablas

Figura 1. Tasa de incidencia del OC en mujeres en 2022. La gráfica muestra la incidencia del OC por continentes. Origen: International Agency for Research on Cancer, 2022.	12
Figura 2. Distribución de los tipos de cáncer según el estadio en el que se detectan. Se puede observar como en el tipo seroso, el 90% de los casos se detectan en los estadios más tardíos. Origen: Torre et al., 2018.	13
Figura 3. Script de instalación paquete inferCNV.....	27
Figura 4. Script iniciación inferCNV.	27
Figura 5. Mapa de calor pre-filtración de ruido generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior), con el añadido de la especificación de los tipos de subc 29	29
Figura 6. Script análisis de expresión diferencial (Seurat) entre subclones en expansión y subclones en extinción, mostrando algunos genes del resultado. La tabla incluye información como el p-valor, el p-valor ajustado, log2FC promedio y el log10 del p-valor.....	31
Figura 7. Script análisis de enriquecimiento funcional mediante GO (clusterProfiler) en genes significativos para obtener los procesos biológicos más enriquecidos y genera las gráficas de barras con los procesos más representativos.	32
Figura 8. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 03.	35
Figura 9. Gráfico UMAP que muestra las diferentes poblaciones celulares presentes en los tejidos tumorales. Las células cancerosas están representadas en rojo, mientras que las células inmunes y estromales en verde y azul, respectivamente.	36
Figura 10. Gráfico UMAP que muestra las diferentes subpoblaciones celulares presentes en los tejidos tumorales.....	37
Figura 11. Gráfico UMAP que representa las células de los 11 pacientes.	37
Figura 12. Distribución de subclones en expansión y en extinción en cada paciente. El gráfico de barras muestra en azul el porcentaje de subclones del paciente que se expanden, y en rojo, los que se extinguen.	39
Figura 13. Stacked bar plot que muestra el porcentaje relativo de células correspondientes a los subclones de cada paciente. Cada color representa la proporción celular en cada subclón.....	40
Figura 14. Volcano plot genes obtenidos en análisis de expresión diferencial. Los puntos rojos representan genes significativos ($pval < 0,05$). A la izquierda se identifican los 10 genes con patrón de extinción más significativos, y a la derecha los 10 genes en expansión.....	42
Figura 15. Top 10 procesos GO enriquecidos en células cancerosas de todos los pacientes conjuntos, según el Gene Ontology. El eje Y muestra la descripción de los procesos biológicos, y el eje X indica el número de genes de cada proceso.....	44
Figura 16. Top 10 procesos biológicos enriquecidos en subclones que se expanden en el conjunto de pacientes. La intensidad del color refleja el valor de p ajustado, que indica la significancia estadística de los mecanismos. El eje Y muestra la descripción de los procesos biológicos y el eje X el número de genes en cada proceso.....	45
Figura 17. Top 10 procesos biológicos enriquecidos en subclones que se expanden en el conjunto de pacientes. La intensidad del color refleja el valor de p ajustado, que indica la significancia estadística de los mecanismos. El eje Y muestra la descripción de los.....	46

Tabla 1. Cronograma detallado de las tareas planificadas para la realización del proyecto, con fechas estimadas por meses.	24
Tabla 2. Tabla que muestra el conteo de células por paciente, fase del tratamiento y localización anatómica.	26
Tabla 3. Paquetes utilizados para el análisis en RStudio. Se puede observar el nombre del paquete, la fuente de origen, la versión y el porqué de su uso.....	30
Tabla 4. Top 10 genes más significativos en la expresión diferencial en el conjunto de subclones que se están expandiendo. Los genes aparecen ordenados según el p valor ajustado. La columna avg_log2FC muestra el cambio de expresión génica.	41
Tabla 5. Top 10 genes más significativos en la expresión diferencial en el conjunto de subclones que se están extinguiendo. Los genes aparecen ordenados según el p valor ajustado. La columna avg_log2FC muestra el cambio de expresión génica.	43

Índice de figuras en anexo

Figura S 1. Distribución de células pre y post tratamiento por subclón en los 11 pacientes. Los gráficos muestran en azul el número de células pre-tratamiento y en rojo post-tratamiento. En el eje X se pueden observar los subclones de cada paciente.....	60
Figura S 2. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 87.	60
Figura S 3. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 136.	61
Figura S 4. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 153.	61
Figura S 5. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 227.	62
Figura S 6. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 349.	62
Figura S 7. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 372.	63
Figura S 8. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 443.	63
Figura S 9. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 540.	64
Figura S 10. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 733.	64
Figura S 11. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 1005.	65
 Tabla S 1. Genes significativos con p-valor, avg_log2FC y p-valor ajustado.....	65

0. ABREVIATURAS

ACS	del inglés <i>American Cancer Society</i> , Sociedad Americana de Cáncer
aCGH	del inglés <i>array comparative genomic hybridization</i> , array de hibridación genómica comparativa.
CNVs	del inglés <i>copy number variants</i> , variantes de número de copia
DNA	del inglés <i>desoxiribonucleotic acid</i> , ácido desoxiribunucleico.
cDNA	del inglés, <i>complementary DNA</i> , DNA complementario.
DNTRseq	del inglés, <i>direct nuclear tagmentation and RNA sequencing</i> , secuenciación dirigida de DNA y RNA
FIGO	del francés, <i>Fédération Internationale de Gynécologie et d'Obstétrique</i> , Federación Internacional de Ginecología y Obstetricia.
HGSC	del inglés, <i>High-grade serous carcinoma</i> , carcinoma seroso de alto grado.
IARC	del inglés, <i>International Agency for Research on Cancer</i> , Agencia Internacional de Investigación en Cáncer
MLPA	del inglés, <i>multiple ligation-dependent probe amplification</i> , amplificación múltiple dependiente de ligación.
SIG	
SBM	del inglés <i>Special Interest Group</i> , grupo de interés especial. del inglés <i>spent blastocyst culture medium</i> , medio de cultivo de blastocisto.
NGS	del inglés <i>Next Generation Sequencing</i> , Secuenciación de Próxima Generación.
NCBI	del inglés, National Center for Biotechnology Information.
NLM	del inglés, National Library of Medicine.

-
- OC** del inglés *ovarian cancer*, cáncer ovárico.
- scSeq** del inglés, *single cell sequencing*, secuenciación de célula única.
- scRNAseq** del inglés, *single cell RNA sequencing*, secuenciación de RNA de célula única.
- SNVs** Del inglés, *single nucleotide variant*, variante de nucleótido único.
- UMI** del inglés, *unique molecular identifier*, identificador molecular único.
- UMAP** del inglés, *Uniform Manifold Approximation and Projection*, identificador molecular único., Aproximación y Proyección Uniforme de Variedades.
- WHO** del inglés, *World Health Organization*, Organización Mundial de la Salud.
- WGS** del inglés *whole generation sequencing*, secuenciación de completa generación.

1. INTRODUCCIÓN

1.1 Cáncer de ovario: una visión general

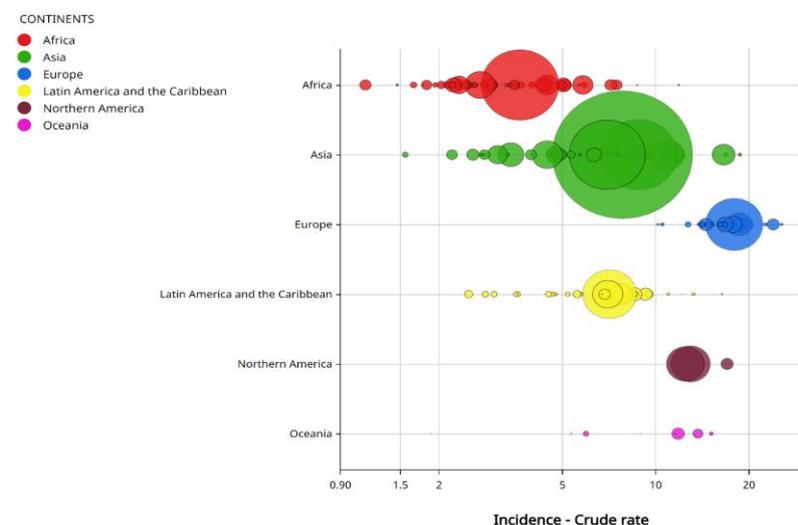
El cáncer es una enfermedad genética que puede afectar a distintos órganos y/o tejidos causada por una anomalía de la división celular que provoca una proliferación celular descontrolada o una pérdida de la muerte celular programada habitual que, además, puede expandirse y afectar a otros órganos y tejidos cercanos (metástasis). Según la WHO (del inglés, *World Health Organization*), el cáncer es la segunda causa de muerte a nivel mundial, y se estimó, en 2018, que 9,6 millones de muertes fueron causadas por esta enfermedad (World Health Organization, 2018). Asimismo, se estima que una de cada tres personas puede sufrir algún tipo de cáncer. En cuanto a las mujeres, una de cada seis mujeres desarrollará cáncer a lo largo de su vida, y una de cada once, fallecerán a causa de esta enfermedad, según la IARC (del inglés, *International Agency for Research on Cancer*).

La heterogeneidad celular en los organismos vivos se debe, principalmente, a las diferencias en la expresión génica, que junto a otros mecanismos y elementos proporcionan un equilibrio, denominado homeostasis, que dota a la célula de capacidad para adaptarse al medio. Cuando aparecen errores en la homeostasis celular existen mecanismos de apoptosis o senescencia para eliminarlas. El cáncer se origina por un error en esta regulación, dando lugar a una proliferación descontrolada, multiplicándose exponencialmente y expandiéndose a otros tejidos. En resumen, las células cancerosas pierden la capacidad de morir y comienzan a dividirse sin límite hasta formar el tumor. Estos grupos de células cancerosas presentan alteraciones genéticas tanto a nivel de DNA como a nivel de RNA.

El cáncer surge, en gran parte de las ocasiones, debido a mutaciones en las células somáticas del organismo de manera espontánea o inducida por factores externos. En un número menor de los casos, cuando las mutaciones afectan a las células de la línea germinal, se pueden transmitir a la descendencia, tratándose de casos de cáncer hereditario.

El cáncer de ovario (OC, del inglés *Ovarian Cancer*) ocurre cuando las células que forman el ovario comienzan a crecer de forma descontrolada. Este tipo de cáncer engloba además a los tumores que se producen en las trompas de Falopio y en el peritoneo. Según el IARC, en 2022 se diagnosticaron más de 300.000 casos de dicha enfermedad a nivel mundial, de los cuales, más de 200.000 mujeres fallecieron a causa de la patología (*International Agency for Research on Cancer, 2022*). Por lo tanto, este tipo de cáncer es, en la clasificación de tipos de cánceres en mujeres, el octavo más común y con mayor mortalidad, y el primero como causa de muerte entre los cánceres ginecológicos. Una de las principales razones de su mal pronóstico es que el diagnóstico suele realizarse en estadios avanzados de la enfermedad. Según la ACS (del inglés, *American Cancer Society*), en el año 2024, en los Estados Unidos, se diagnosticarán alrededor de 20.000 mujeres con cáncer de ovario, de las cuales casi 13.000 fallecerán (*American Cancer Society, 2024*). En números de incidencia, Europa es el continente con mayor ratio de mujeres con cáncer de ovario en el mundo (ver Figura 1)

Region grouping vs Incidence - Crude rate, Females, in 2022
 Ovary



Cancer TODAY | IARC - <https://gco.iarc.who.int/today>
 Data version : Globocan 2022 (version 1.1)
 © All Rights Reserved 2024



Figura 1. Tasa de incidencia del OC en mujeres en 2022. La gráfica muestra la incidencia del OC por continentes.
 Origen: International Agency for Research on Cancer, 2022.

La mayoría de los cánceres de ovario son detectados en estadios avanzados (estadio FIGO III y IV) debido a la escasez de sintomatología específica al inicio de la enfermedad.

Además, la falta de una metodología de diagnóstico precoz eficaz para dicha enfermedad es, además, un factor fundamental en la elevada mortalidad asociada a este tipo de cáncer (Torre et al., 2018).

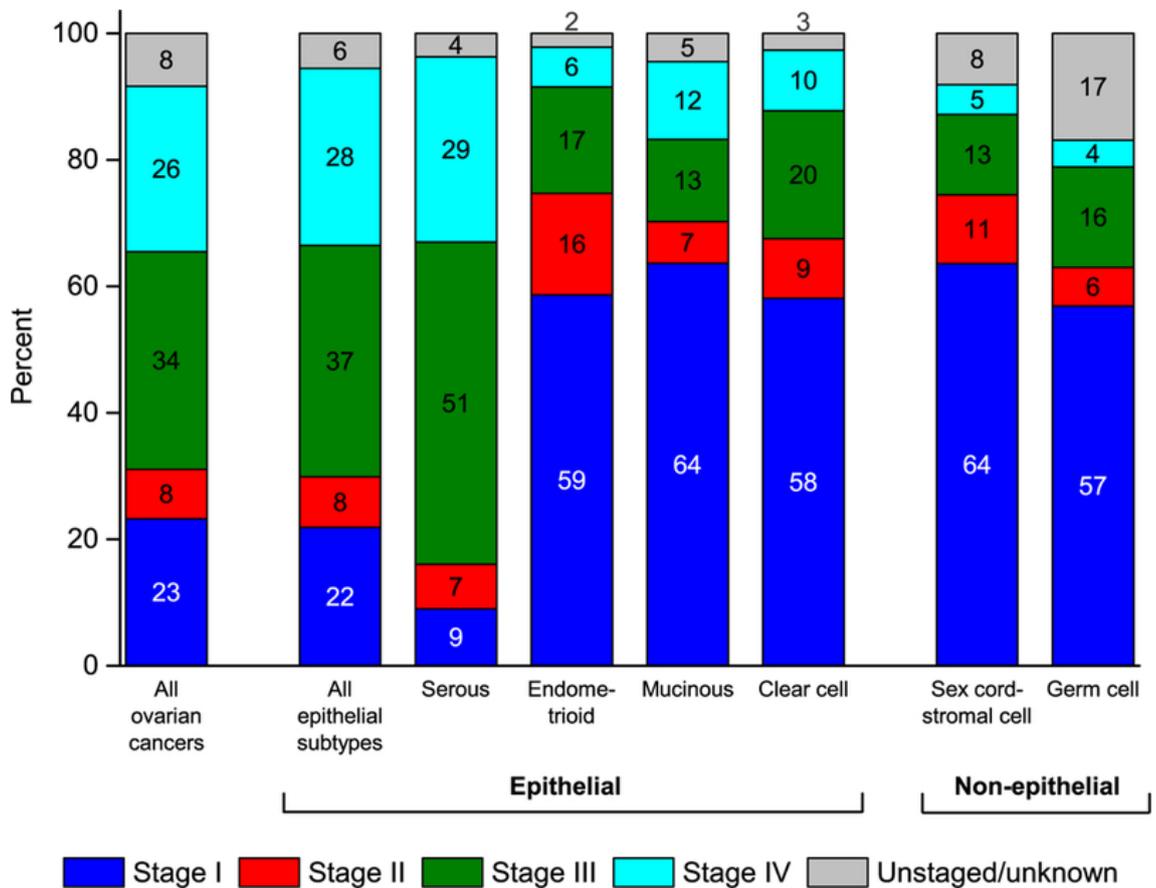


Figura 2. Distribución de los tipos de cáncer según el estadio en el que se detectan. Se puede observar como en el tipo seroso, el 90% de los casos se detectan en los estadios más tardíos. Origen: Torre et al., 2018.

Este cáncer se origina principalmente en mujeres de edad avanzada. Aproximadamente, la mitad de las mujeres diagnosticadas tienen más de 60 años. En cuanto a la predisposición a padecer dicho cáncer, la información es escasa, siendo los factores más comunes relacionados con el ciclo ovulatorio. Por este motivo, la toma de anticonceptivos orales, la multiparidad y la lactancia materna reducen el riesgo de desarrollar cáncer de ovario. Además, la significativa diferencia observada en la incidencia entre países desarrollados y no desarrollados, indica que pueden existir factores relacionados con los países desarrollados que pueden aumentar la posibilidad

de padecer este tipo de cáncer, tales como la obesidad, dietas ricas en cRNAes, terapias hormonales sustitutivas, herencia, edad, endometriosis y/o el consumo de agentes tóxicos. No obstante, el factor de riesgo más relevante es el genético, siendo gran parte de las pacientes portadoras de alteraciones genéticas hereditarias como pueden ser mutaciones (15%) en los genes *BRCA1* o *BRCA2* que proporcionan un riesgo elevado a padecer este tipo de cáncer o las mutaciones en los genes *MSH2* o *MLH1* relacionadas con el síndrome de Lynch, también asociado con cáncer de ovario.

Según la histopatología del tumor, se pueden diferenciar tres tipos de OC: **carcinoma epitelial**, que representa el 90% de los cánceres, derivan del epitelio que recubre la superficie ovárica y suele darse en mujeres a partir de los 50 años; **tumores de células germinales**, representan un 5-10% de los casos y son desarrollados en las células generadas en los óvulos siendo más frecuentes en mujeres jóvenes (<25 años); **tumores del estroma**, representan el 5% de los tumores y están formados en células que producen las hormonas sexuales femeninas, son muy poco frecuentes. Además, según la malignidad, todos estos tumores se pueden clasificar en **benignos** (60%), **borderline o de bajo potencial maligno** (10%) y **malignos** (30%). Asimismo, se pueden clasificar los tumores según el grado de diferenciación histológico, **grado 1 o bajo, grado 2 o moderado y grado 3 o alto**.

Debido a la relevancia y la gran incidencia en la población con OC, los tumores epiteliales del ovario son los más estudiados, y dentro de este tipo se pueden encontrar diversos tipos: **tumores serosos (HGSC, del inglés, High-grade serous carcinoma)**, son los cánceres de ovario más frecuentes; **tumores mucinosos**, simbolizan el 15% de los tumores malignos de ovario y producen mucina que se va extendiendo por la cavidad peritoneal, tienen buen pronóstico dada la precocidad de los diagnósticos; **tumores endometrioides**, representan el 15-25% de los casos; **tumores de células claras**, el 5% de los cánceres de ovario son de este tipo; **tumores indiferenciados**, representan el 1,5% de los casos.

El carcinoma seroso de alto grado (HGSOC) es el tipo de cáncer de ovario más común provocando el 70-80% de las muertes (Bowtell et al., 2015). Además, la supervivencia no ha variado en los últimos años por la falta de mejora en los tratamientos dada la dificultad que implica este tipo de OC. El HGSOC se caracteriza por una alta

heterogeneidad molecular y muestra unas pocas anomalías genéticas relacionadas. Gran parte de los pacientes con HGSOC presentan una pérdida de función del gen *TP53* que da lugar a numerosas variantes de número de copia, o, en menor medida de otros genes como *BRCA1*, *BRCA2*, *NF1*, *CDK12*, *GABRA6*, *RB1*, *PTEN* y *RAD51B*, entre otros.

El HGSOC se caracteriza por ser un tumor genéticamente inestable y con una gran heterogeneidad clonal con distintos genotipos. La heterogeneidad intratumoral se ha asociado con la recurrencia, la resistencia a la quimioterapia e incluso a un pronóstico desfavorable (Denisenko et al., 2024). Asimismo, se ha observado que la diversidad clonal aumenta con la progresión del cáncer y la metástasis, lo que se asocia a peores pronósticos y a una peor respuesta al tratamiento (Morden et al., 2021; Patch et al., 2015).

La heterogeneidad intratumoral ha sido investigada durante muchos años, dada la dificultad de este estudio, hoy en día, se siguen buscando nuevos métodos para caracterizar las poblaciones de células tumorales con el fin de caracterizar individualmente a cada paciente y así poder realizar un tratamiento personalizado. Gracias a las investigaciones que han ido surgiendo durante los años, se ha demostrado la heterogeneidad mutacional que existe dentro de un mismo tumor y la capacidad que tienen estas poblaciones de evolucionar selectivamente para formar células resistentes a los tratamientos.

La mayoría de las pacientes con HGSOC responden al tratamiento inicialmente, sin embargo, gran parte de ellas desarrollan resistencia al tratamiento. El 80% de las pacientes recaen y responden peor al tratamiento (Jayson et al., 2014). Este aspecto es uno de los desafíos principales en la investigación del mecanismo de la enfermedad. El tratamiento estándar actual es la quimioterapia basada en fármacos de platino como el carboplatino o el cisplatino que ayudan a eliminar las células cancerosas que quedan tras la cirugía citorreductora (Kelland et al., 2007). La resistencia al tratamiento con platino se da cuando la recaída se produce en los 6 meses posteriores a la terapia más reciente y, actualmente, las opciones terapéuticas son muy limitadas, derivando en una supervivencia escasa (Havasi et al., 2023; McNeish et al., 2014; Zhou et al., 2020). Recientemente, se han aprobado nuevos fármacos como los inhibidores de poli (ADP-

ribosa) polimerasas y un inhibidor del factor de crecimiento endotelial vascular (Ortiz et al., 2022; Senkowski et al., 2023).

La resistencia al tratamiento ocurre por diversos factores como la selección de clones resistentes, la sobreexpresión de vías de protección tumoral, sobreexpresión de células envueltas en la activación inmune, hipoxia o remodelación de la matriz extracelular, los microRNA, los RNA no codificantes y/o los cambios epigenéticos, que también contribuyen a la quimiorresistencia (Olbrecht et al., 2021). Uno de los objetivos de la medicina de precisión en la actualidad es el de caracterizar a los pacientes con el fin de realizar tratamientos según los perfiles genómicos y transcriptómicos de la causa de la enfermedad, en este caso, de los tumores. Estos enfoques aumentan el éxito del tratamiento considerablemente. Actualmente se necesita abrir nuevas vías de investigación que permitan descubrir nuevos patrones de expresión o mecanismos de resistencia a la quimioterapia con el fin de conseguir nuevos tratamientos personalizados para los pacientes de HGSOC.

Recientemente, se ha abierto una línea de investigación novedosa basada en la inferencia de variantes de número de copia en células tumorales mediante scRNA-seq. Gracias a estos resultados se pueden observar los distintos subclones que presenta la muestra tumoral en el momento de la extracción. Esto da la posibilidad de estudiar el comportamiento de las células cancerosas antes y después de un tratamiento, proporcionando información de qué clones están respondiendo o no a la quimioterapia y así, poder enfocar el tratamiento a las vías implicadas en la expresión de dichos genes (Y. Wu et al., 2020; Yin et al., 2019). Este enfoque afronta uno de los mayores desafíos clínicos en el tratamiento del HGSOC, la resistencia a la quimioterapia basada en platino (L. Wang et al., 2024).

En diversas publicaciones se ha demostrado que el estudio cuantitativo de la heterogeneidad intratumoral en HGSOC ayuda a predecir la supervivencia al tratamiento (Schwarz et al., 2015).

1.2 Secuenciación de RNA en células individuales: revelando la heterogeneidad tumoral

Las técnicas tradicionales de secuenciación solo son capaces de obtener un valor medio de las señales de poblaciones celulares, lo que dificulta la caracterización precisa de la heterogeneidad tumoral. En cambio, la secuenciación de células individuales permite identificar con precisión las diferencias genómicas y/o transcriptómicas.

El estudio de los diferentes tipos de cáncer ha ido variando con el paso de los años siendo la transcriptómica una de las áreas más beneficiadas. Este ámbito de estudio trata del análisis del RNA, comúnmente, para medir la expresión de los genes de interés. Gracias a la diferencia de expresión en los genes, se pueden observar patrones en distintas enfermedades, dando lugar a un amplio abanico de posibilidades en la investigación actual. Esto sumado a la secuenciación de nueva generación dio lugar a la aparición de la secuenciación de RNA o RNAseq (del inglés, *RNA sequencing*) que permite analizar todos los tipos de RNA posibles, y a partir de ahí, relacionar los hallazgos con las enfermedades.

Una de las técnicas más recientes en el análisis de datos de secuenciación de RNA es el realizado en células individuales. El objetivo de esta técnica es el de analizar el comportamiento a nivel celular y surge debido al constante avance que existe en las técnicas de secuenciación. Este tipo de estudio se puede utilizar, según el objetivo, para analizar, por ejemplo, distintas poblaciones celulares y su contribución durante un proceso concreto como una enfermedad. Una de las aplicaciones más interesantes es el análisis del transcriptoma de cada célula, o lo que es similar, el análisis celular a nivel de expresión también denominado scRNAseq (del inglés, *single cell RNA sequencing*) (Kolodziejczyk et al., 2015). Esto supuso un gran avance en las tecnologías transcriptómicas, siendo el primer estudio publicado en 2009 (Tang et al., 2009).

La secuenciación de RNA en células individuales o scRNAseq permite la exploración de la diversidad transcriptómica tumoral en los distintos tipos de cáncer, mediante el análisis simultáneo de miles de células de una sola muestra. Gracias al análisis transcriptómico de las células, se puede observar la expresión celular en distintos contextos, como, por ejemplo, durante la diferenciación celular o en respuesta a ciertos tratamientos. Todo esto lleva a la posibilidad de mejorar el diagnóstico de distintas

enfermedades tales como distintos tipos de cáncer, enfermedades inflamatorias, respiratorias o neurodegenerativas (Li et al., 2022). Dado la complejidad de la técnica, siguen apareciendo desafíos que llevan a los investigadores a buscar nuevos enfoques. El estudio de células individuales presenta una principal dificultad respecto al RNAseq tradicional, que es la dispersión y la complejidad de los datos a analizar. Sin embargo, hoy en día, ya tiene gran relevancia en diversos campos de la biomedicina. Uno de los ejemplos que, además, se va a tratar en este trabajo, es la aplicación en el análisis oncológico. Gracias al análisis de las células en tejido tumoral, se ha podido entender mejor el complejo entorno que supone y la heterogeneidad intratumoral que se puede observar. Estos estudios aplicados tanto en cánceres tempranos como en tumores metastásicos ayudan a mejorar el diagnóstico y jugar un papel importante en la personalización del tratamiento, siendo clave en la oncología de precisión (Keyvani et al., 2019; Yuan et al., 2024).

Al igual que en gran parte de los análisis transcriptómicos, surgen retos debido a distintas limitaciones que van apareciendo durante el proceso, como la falta de datos, la aparición de artefactos o la sensibilidad de la técnica (Kim et al., 2015; Stegle et al., 2015). El ruido de fondo y la falta de material analizable plantean retos en el análisis computacional de los datos de scRNAseq. Debido a esta situación surgen numerosas herramientas estadísticas para optimizar este análisis y conseguir los mejores resultados posibles (Haque et al., 2017).

1.3 Inferencia de CNVs en scRNAseq: bioinformática y su aplicación

Las variantes de número de copia son reordenamientos desbalanceados de secuencias de DNA que pueden ser de un tamaño desde kilobases hasta cromosomas completos. Según diversos factores tales como región cromosómica, genes incluidos, secuencia de DNA afectada y/o estado de cigosidad, las CNVs pueden tener gran relevancia sobre el fenotipo de un paciente. La patogenicidad de dichas CNVs depende del efecto directo sobre la expresión de los genes afectados según su relevancia. Estas variantes pueden ser detectadas por diversos métodos, entre los más comunes, se encuentra la hibridación genómica comparativa o aCGH (del inglés, *array comparative*

genomic hybridization), el MLPA (del inglés, *multiple ligation-dependent probe amplification*), la secuenciación de nueva generación o NGS (del inglés, *next generation sequencing*) o la secuenciación del genoma completo o WGS (del inglés, *whole genome sequencing*). Estas técnicas tienen sus limitaciones debido a la resolución limitada de algunas de ellas o el coste/duración de otras. En contexto de cáncer, se han observado un número de CNVs mayor, probablemente, debido a la gran cantidad de divisiones mediante mitosis y la falta de control del ciclo celular en las células cancerosas. Diversos estudios han destacado la importancia de las CNVs en algunos tipos de cáncer como el pancreático, el cáncer de ovario o de mama, dónde pueden alcanzar hasta un 10% de las variantes patogénicas causantes de la enfermedad (Al-Sukhni et al., 2012). En el caso de los estudios de CNVs en scRNAseq en pacientes con HGSOC se han realizado menos investigaciones dada la dificultad que implica.

No obstante, la realización de perfiles de CNVs inferidos de scRNAseq es capaz de dRNAos información genética sobre cada célula tumoral en una muestra y con estos datos caracterizar las poblaciones clonales, su nivel de expresión durante el tratamiento y comprender mejor los mecanismos de resistencia o sensibilidad a la quimioterapia (Chai et al., 2024; Glasgow et al., 2019).

Actualmente, existe un número limitado de métodos capaces de detectar CNVs a partir de datos de scRNAseq. Estos métodos se basan en la correlación entre la diferencia de expresión génica y las CNVs. Algunos de los métodos más usados son: inferCNV, HoneyBADGER, CaSpER o CopyKAT.

InferCNV es una herramienta perteneciente al proyecto TrinityCTAT (Tickle, Ti, Brown, & Haas, 2019), centrado en el uso de RNAseq en células individuales para entender mejor el análisis transcriptómico en muestras tumorales. El objetivo de esta herramienta es la identificación de alteraciones somáticas a gran escala, ya sean ganancias o pérdidas de partes concretas de los cromosomas o completas. Esto se consigue mediante la exploración de la intensidad de la expresión génica en el genoma del tumor, comparándolo con células de referencia sin alteraciones (Zhang et al., 2022). Tras esto, se pueden observar los resultados en diferentes gráficas como los mapas de calor en el que se puede ilustrar las diferencias de expresión según la intensidad de la señal, identificando así que regiones de los cromosomas presentan una ganancia o una

pérdida de material respecto a la referencia. Una de las características de dicho paquete es la variedad de soluciones que ofrece, como el uso de filtros con el objetivo de minimizar el ruido obtenido (típico en resultados de scRNAseq en muestras tumorales), el uso de métodos de predicción de regiones más susceptibles a presentar CNVs y la clasificación de grupos celulares para estudiar la heterogeneidad tumoral. Otro método, como puede ser HoneyBADGER, utiliza un modelo Markov basado en un enfoque bayesiano para identificar las CNVs (Fan et al., 2018), que se basa en la calculación de la desviación de la fracción de alelos de las variantes heterocigotas del valor esperado. Aunque su particularidad reside en que es necesario detectar primero las variantes de nucleótido único o SNVs (del inglés, *single number variants*) para poder confirmar la presencia de CNVs. CopyKAT y CaSpER son otros ejemplos de herramientas con un enfoque bayesiano y agrupamiento jerárquico que se ha demostrado que es un método fiable en la identificación de células tumorales entre las células epiteliales en pacientes con cáncer de mama (Gao et al., 2021; Serin Harmanci et al., 2020).

A pesar de la aparición de nuevas herramientas para la inferencia de CNVs a partir de datos de scRNAseq, inferCNV sigue siendo una de las más usadas en el ámbito científico. La capacidad de reducir el ruido típico de datos de scRNAseq, junto su flexibilidad en cuanto al uso de datos de gran capacidad y la posibilidad de aplicar diferentes opciones como el filtrado o el procesamiento, hacen de inferCNV una de las opciones preferidas para los investigadores en el contexto oncológico.

Las CNVs de gran tamaño afectan de modo más drástico a la expresión génica de la región, por este motivo, la detección de este tipo de variantes es más precisa y mucho más fiable que la identificación de CNVs de menor tamaño. El ruido, aunque se minimice, puede seguir acarreando una dificultad adicional a la hora de detectar CNVs de menor tamaño. Además, la diferenciación entre variantes reales y artefactos puede suponer una gran dificultad en una escala menor debido a la heterogeneidad, la falta de profundidad en el estudio o el fenómeno de *dropout* o pérdida alélica debido a un error de amplificación.

2. HIPÓTESIS

2.1 Hipótesis principal

La heterogeneidad tumoral en pacientes con cáncer de ovario se puede caracterizar mediante la inferencia de CNVs y la expresión diferencial, lo que permitirá detectar los subclones de cada muestra y observar si el tratamiento surge efecto o, por lo contrario, no implica mejora debido a la presencia de mecanismos de resistencia.

2.1 Hipótesis secundarias

- La inferencia de CNVs a partir de datos de scRNAseq puede revelar alteraciones genómicas específicas correlacionadas con subclones tumorales en expansión o extinción.
- El uso de herramientas como inferCNV junto a un análisis transcriptómico puede mejorar la comprensión de la heterogeneidad tumoral, proporcionando una mayor precisión en la identificación de eventos clonales.
- El estudio de los genes y rutas biológicas implicadas en la expansión o extinción de los subclones tumorales puede mejorar la identificación de biomarcadores clave asociados con la resistencia o sensibilidad al tratamiento.

3. OBJETIVOS

3.1 Objetivo general

Determinar si, mediante la inferencia de CNVs, es posible caracterizar la heterogeneidad intratumoral y el comportamiento de subclones tumorales en pacientes con HGSOC, con el propósito de entender la causa de la resistencia o sensibilidad a la estrategia terapéutica utilizada.

2.2 Objetivos específicos

- ◆ Comparar células tumorales pre y post tratamientos para la expansión o extinción de los subclones presentes en el tumor.
- ◆ Aplicación de herramientas aprendidas durante el curso como estudios de expresión diferencial y enriquecimiento funcional.
- ◆ Identificar posibles biomarcadores que expliquen la resistencia a la quimioterapia neoadyuvante.

4. MATERIALES Y MÉTODOS

4.1 Búsqueda de información bibliográfica

Para realizar la introducción y poner en contexto sobre el cáncer de ovario, la tecnología de scRNAseq y la relevancia de las CNVs en estos estudios, se ha utilizado el motor de búsqueda de acceso libre PubMed, implementado por el NCBI (del inglés, *National Center for Biotechnology Information*), perteneciente a la NLM (del inglés, *National Library of Medicine*) de Estados Unidos. Dicho motor de búsqueda permite el acceso a la base de datos MEDLINE de citas y resúmenes de artículos de investigación biomédica.

En la búsqueda se emplearon combinaciones de las siguientes palabras: inferCNV, scRNAseq, ovarian cancer, copy number variants, HGSOC, chemoresistance.

Como criterio de inclusión para la revisión bibliográfica, se estableció que *quimiorresistencia en cáncer ovárico* fuera el objeto principal del estudio o que entre sus objetivos principales se incluyeran uno de los siguientes: inferencia de CNVs en cáncer ovárico mediante scRNAseq y/o *high-grade serous ovarian cancer* en scRNAseq. Todos los artículos que no cumplieron estos criterios fueron excluidos.

4.2 Plan de trabajo y tareas

El cronograma que se muestra a continuación (ver Tabla 1) muestra la disposición temporal del trabajo realizado, desde la asignación de tema hasta la entrega y presentación.

Para la realización del trabajo, se llevarán a cabo ciertas tareas enumeradas a continuación:

1. Revisión de la literatura científica del cáncer ovárico, scRNAseq y la detección de CNVs.
2. Revisión de la literatura científica de aplicación bioinformática de estos aspectos.
3. Revisión información sobre herramienta inferCNV
4. Redacción de la introducción
5. Preparación y preprocesamiento de datos
6. Aplicar paquete inferCNV
7. Análisis de resultados
8. Redacción de resultados
9. Corrección
10. Entrega
11. Defensa

4.3 Cronograma

Tabla 1. Cronograma detallado de las tareas planificadas para la realización del proyecto, con fechas estimadas por meses.

	TFM Inferencia de CNVs en scRNAseq en cáncer de ovario															
	MARZO		ABRIL		MAYO		JUNIO		JULIO		AGOSTO		SEPTIEMBRE		OCTUBRE	
	S1-S2	S3-S4	S1-S2	S3-S4	S1-S2	S3-S4	S1-S2	S3-S4	S1-S2	S3-S4	S1-S2	S3-S4	S1-S2	S3-S4	S1-S2	S3-S4
Asignación tema de trabajo																
Búsqueda bibliográfica			[Barra azul]													
Tutorías																
PIANIFICACIÓN	[Barra azul]															
Redacción	[Barra azul]															
Abstract																
Objetivos																
Introducción	[Barra azul]															
Métodos y Resultados																
Discusión y Conclusión																
Revisión																
Entrega																
Defensa																

4.4 Origen y manejo de los datos

Los datos utilizados para este trabajo provienen de la publicación con título: Longitudinal single-cell RNA-seq data of metastatic ovarian cancer (Zhang et al., 2022) y contiene datos de scRNAseq de pacientes con cáncer de ovario seroso de alto grado. Cada paciente tiene muestras pareadas de antes y después del tratamiento con carboplatino. Los resultados incluyen información de las células tumorales y de las células de referencia. Dichos datos se encuentran accesibles en la base de datos NCBI GEO (GSE165897), asegurando su accesibilidad. Dado que los datos se encuentran públicamente accesibles, no es necesaria una aprobación del comité de ética para el estudio realizado.

El estudio realizado es un estudio de diseño descriptivo obtenido de una fuente de datos en abierto. El objetivo principal es el de identificar CNVs en diferentes poblaciones celulares del tumor ovárico que nos permita establecer la dinámica de evolución clonal en respuesta al tratamiento. La población de estudio consta de pacientes diagnosticados con cáncer ovárico en distintas fases de la enfermedad y se caracteriza la resistencia a quimioterapia en el cáncer de ovario seroso de alto grado usando muestras de tejido antes y después del tratamiento. En todos ellos se analiza una muestra de tejido tumoral mediante scRNAseq y células estromales como referencia para la comparación.

La población de estudio consiste en pacientes con cáncer de ovario seroso de alto grado de la cohorte del proyecto DECIDER (NCT04846933) y tratados en el Hospital Universitario de Turku. El estudio incluye resultados sobre 11 pacientes, de los cuales se obtuvieron muestras antes y después de la quimioterapia neoadyuvante. Tras el control de calidad, se analizaron un total de 51,786 células, de las cuales 8,806 eran células tumorales, 8,045 células estromales y 34,935 células inmunes (Ver Tabla 2). Además, se puede observar información adicional como la fase de tratamiento, la localización anatómica, el tipo y subtipo celular y los datos de expresión. El proyecto DECIDER (Innovative personalized treatments for ovarian cancer, n.d.) tiene como objetivo

desarrollar herramientas diagnósticas y tratamientos para pacientes con cáncer de ovario seroso de alto grado con la ayuda de métodos de inteligencia artificial.

Tabla 2. Tabla que muestra el conteo de células por paciente, fase del tratamiento y localización anatómica.

patient_id	treatment_phase	anatomical_location	cell_count
EOC1005	post-NACT	Tumor	1288
EOC1005	treatment-naive	Peritoneum	3071
EOC136	post-NACT	Omentum	3602
EOC136	treatment-naive	Mesentery	2501
EOC153	post-NACT	Omentum	2401
EOC153	treatment-naive	Omentum	1380
EOC227	post-NACT	Omentum	2015
EOC227	treatment-naive	Omentum	2178
EOC3	post-NACT	Omentum	2503
EOC3	treatment-naive	Peritoneum	3142
EOC349	post-NACT	Omentum	1471
EOC349	treatment-naive	Peritoneum	1483
EOC372	post-NACT	Peritoneum	4671
EOC372	treatment-naive	Peritoneum	711
EOC443	post-NACT	Omentum	4463
EOC443	treatment-naive	Omentum	2122
EOC540	post-NACT	Omentum	2043
EOC540	treatment-naive	Omentum	1921
EOC733	post-NACT	Omentum	4077
EOC733	treatment-naive	Peritoneum	1742
EOC87	post-NACT	Omentum	1491
EOC87	treatment-naive	Peritoneum	1510
			51786

Dicho estudio se llevó a cabo durante el año 2021 en el Hospital Universitario de Turku, en Finlandia. Entre los criterios de inclusión más relevantes se encuentran que el paciente esté diagnosticado con cáncer de ovario seroso de alto grado, que hayan firmado el consentimiento informado para el análisis del tejido tumoral y aquellos que tuviesen disponibilidad para extraer tejido antes y después del tratamiento. Los criterios de exclusión fueron los pacientes que no dieron el consentimiento o con historia clínica incompleta, e incluso pacientes en los que la muestra de tejido tumoral no fue apta para su procesamiento.

La secuenciación realizada por el equipo de investigación se llevó a cabo con las plataformas Illumina HiSeq 2500, HiSeq 4000 y NovaSeq 6000. Para la recogida de datos, se utilizaron plataformas de análisis de datos de secuenciación obtenidos mediante scRNAseq. Se usó el software R y la aplicación de paquetes específicos para dicho análisis junto a la herramienta inferCNV (Tickle, Ti, Brown, & Haas, 2019).

4.4.1 InferCNV: Herramienta para la inferencia de CNVs.

A continuación, se describirán los procedimientos realizados en el trabajo para el análisis de la inferencia de CNVs en datos obtenidos en estudios de scRNAseq en pacientes con cáncer de ovario. La metodología sigue el plan de trabajo mencionado anteriormente y se estructura en varias fases.

La recopilación de datos, donde se obtuvieron los datos del estudio experimental original y se clasificó la naturaleza de dichos datos. Estos datos constan de una matriz de conteo o *UMI counts* en la que aparece información sobre las células y los genes. Además, consta de otro archivo de anotación celular que contiene información de la célula y el subtipo celular y, por último, un archivo de posición genómica con información de la localización de los genes en cada cromosoma con sus respectivas coordenadas.

```
```{r cars}
if (!requireNamespace("BiocManager", quietly = TRUE))
 install.packages("BiocManager")
BiocManager::install("infercnv")
```
```

Figura 3. Script de instalación paquete inferCNV.

Seguidamente, se procede a la aplicación de la herramienta inferCNV mediante R con la extensión inferCNVapp (ver Figura 4), una aplicación creada para ejecutar el paquete de inferCNV de manera más sencilla.

```
```{r}
library("devtools")
devtools::install_github("broadinstitute/infercnvApp")
infercnvApp::infercnvApp()
```
```

Figura 4. Script iniciación inferCNV.

Las características utilizadas en la aplicación fueron las siguientes: **selección del modelo HMM** (del inglés, Hidden Markov Model) i6, un modelo de 6 estados que se basa en las intensidades de expresión a nivel génico; **modo de análisis por subclones**, lo que permite que se analicen las CNVs por subclones; **método Leiden**, un algoritmo de agrupamiento usado por su fiabilidad; **filtro de eliminación de ruido**, con el objetivo de reducir las señales de ruido que pueden surgir en el análisis; y el **parámetro de corte** se fijó en 0.1.

El proceso realizado por inferCNV consiste en:

- Disposición de datos de entrada, que constan de una matriz de expresión génica con los genes en las filas y las células en las columnas, un archivo de anotación celular con la información de cada célula, incluidas las de referencia. Además, se necesita un archivo de posición genómica.
- Tras cargar los datos en inferCNV hay que especificar qué tipo celular servirá de referencia, en nuestro caso las células inmunes y estromales. A su vez, en este paso se especifican las características seleccionadas.
- InferCVN analiza la expresión génica en cada célula comparándola con las de las células de referencia, gracias al modelado HMM se asigna el estado de la CNV y mediante el método de análisis procede a la agrupación de las células cancerosas en subclones según su perfil.
- Se aplica un filtro de eliminación de ruido (ver Figura 5) clave en resultados de scRNAseq. Por último, se obtienen los resultados en forma de mapas de calor y objetos de inferCNV que fueron usados más adelante para la expresión diferencial.

Tabla 3. Paquetes utilizados para el análisis en RStudio. Se puede observar el nombre del paquete, la fuente de origen, la versión y el porqué de su uso.

| Paquete | Fuente | Versión | Uso |
|-----------------|--------------|---------|--|
| BiocManager | Bioconductor | 1.30.25 | Instalar y administrar paquetes de Bioconductor. |
| DESeq2 | Bioconductor | 1.44.0 | Análisis de expresión diferencial de secuenciación de ARN. |
| Seurat | Bioconductor | 5.1.0 | Análisis de datos de expresión de célula única y visualización de UMAP. |
| SeuratObject | Bioconductor | 5.0.2 | Gestión de objetos Seurat |
| biomaRt | Bioconductor | 2.60.1 | Acceso y conversión de anotaciones genómicas para la identificación de genes. |
| clusterProfiler | Bioconductor | 4.12.6 | Análisis de enriquecimiento funcional y análisis de rutas. |
| devtools | CRAN | 2.4.5 | Instalación de paquetes y herramientas de desarrollo. |
| dplyr | CRAN | 1.1.4 | Manipulación y transformación de datos en la preparación y filtrado de datos. |
| ggplot2 | CRAN | 3.5.1 | Visualización de datos mediante gráficos para el análisis de resultados. |
| ggrepel | CRAN | 0.9.5 | Evitar superposiciones en las etiquetas de gráficos para mejorar la claridad visual. |
| infercnv | Bioconductor | 1.20.0 | Inferencia de variaciones en el número de copias (CNVs) a partir de datos scRNA-seq. |
| pheatmap | CRAN | 1.0.12 | Creación de mapas de calor para visualizar datos de expresión génica. |
| writexl | CRAN | 1.5.0 | Exportar resultados de análisis a |
| org.Hs.eg.db | Bioconductor | 3.19.1 | Mapa de anotaciones genéticas de Homo Sapiens |

Tras la obtención del objeto inferCNV, se procedió a la selección de subclones celulares y su posterior clasificación según el estado:

→ **Expansión:** cuando el número de células del subclón después del tratamiento aumenta de manera significativa respecto al número de células en el mismo subclón antes del tratamiento.

→ **Extinción:** cuando el número de células en el subclón se ve disminuido tras el tratamiento respecto al número de antes.

→ **Neutral:** cuando el número de células no varía antes y después del tratamiento.

En nuestro caso, se realizó una selección de aquellos grupos celulares pertenecientes a un subclón con un mínimo de 20 células totales, con el fin de alcanzar una correcta representatividad. Para cada paciente se generó un *dataframe* con las células antes del

tratamiento y después del tratamiento, comparando el porcentaje de evolución de los subclones. Esto nos dio una primera aproximación sobre cómo se comportaron en cada paciente.

Seguidamente, se procedió a realizar el análisis de expresión diferencial, con el objetivo de identificar los genes con una expresión diferencial significativa en los subclones estudiados (Ju et al., 2009).

En primer lugar, el preprocesado de los datos incluyó un paso de normalización logarítmica basada en la expresión génica de cada célula, una selección de características altamente variables para obtener únicamente los genes con expresión diferencialmente significativa y, por último, una reducción de la dimensionalidad mediante el análisis de componentes principales. Para la visualización de los resultados se utilizó la técnica UMAP (por sus siglas en inglés, *Uniform Manifold Approximation and Projection*) que nos permitió observar las poblaciones celulares en un gráfico bidimensional. Finalmente, el análisis de expresión diferencial se llevó a cabo con la función *FindMarkers()*, que nos identificó los genes significativamente expresados (Ver Figura 6).

```

# Unimos las capas del objeto Seurat combinado
combined_seurat_obj <- JoinLayers(combined_seurat_obj)

# Realizo el análisis de expresión diferencial con las identidades "Expandiéndose" y "Extinguiéndose"
resultados_diff_expr_conjunto <- FindMarkers(combined_seurat_obj, ident.1 = "Expandiéndose", ident.2 = "Extinguiéndose", min.pct = 0.25)

# Verificamos los resultados
head(resultados_diff_expr_conjunto)

```

| Description | p_val | avg_log2FC | pct.1 | pct.2 | p_val_adj | log10_p_val |
|-------------|--------------|------------|-------|-------|--------------|-------------|
| | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| RPS16 | 1.781336e-40 | 0.5503330 | 1.000 | 1.000 | 2.903043e-36 | 39.74925 |
| TM4SF1 | 2.187482e-34 | -0.6810132 | 0.980 | 0.993 | 3.564939e-30 | 33.66006 |
| SAT1 | 3.548840e-30 | -0.6812444 | 0.993 | 0.996 | 5.783544e-26 | 29.44991 |
| EEF1B2 | 3.037481e-29 | 0.7797493 | 0.985 | 0.984 | 4.950184e-25 | 28.51749 |
| RPS27A | 6.897502e-29 | 0.3057267 | 1.000 | 1.000 | 1.124086e-24 | 28.16131 |
| RPS25 | 2.734996e-27 | 0.5063591 | 1.000 | 0.998 | 4.457222e-23 | 26.56304 |

Figura 6. Script análisis de expresión diferencial (Seurat) entre subclones en expansión y subclones en extinción, mostrando algunos genes del resultado. La tabla incluye información como el p-valor, el p-valor ajustado, log2FC promedio y el log10 del p-valor.

A continuación, con el fin de identificar los procesos biológicos asociados a estos genes diferencialmente expresados se realizó un análisis de enriquecimiento funcional mediante el paquete *clusterProfiler* (M. Wu et al., 2020).

El enriquecimiento funcional se realizó mediante el uso de los genes diferencialmente expresados (previamente obtenidos en la expresión diferencial) y la base de datos de

Gene Ontology (GO). Esta base de datos nos permitió identificar procesos biológicos, funciones moleculares y componentes celulares en los que estaban involucrados los genes significativamente expresados. La función empleada fue *enrichGO()* y se visualizaron los datos mediante la función *ggplot()* y *barplot()* (Ver Figura 7).

```

# Aplicar el análisis de enriquecimiento GO
enriquecimiento_GO_conjunto <- enrichGO(
  gene = genes_interes_conjunto,
  OrgDb = org.Hs.eg.db,
  keyType = "SYMBOL",
  ont = "BP", # "BP" para Biological Process
  pvalueCutoff = 0.05
)

# Lo convertimos en dataframe
enriquecimiento_GO_conjunto_df <- as.data.frame(enriquecimiento_GO_conjunto$result)
print(head(enriquecimiento_GO_conjunto_df))

# Ordeno el df por GeneRatio de forma descendente
enriquecimiento_GO_conjunto_df$GeneRatio_num <- sapply(enriquecimiento_GO_conjunto_df$GeneRatio, function(x) as.numeric(strsplit(x, "/")[1][1]))
enriquecimiento_GO_conjunto_df <- enriquecimiento_GO_conjunto_df[order(-enriquecimiento_GO_conjunto_df$GeneRatio_num), ]

# Seleccionamos las 10 más representativas
top_enriquecimiento_conjunto <- head(enriquecimiento_GO_conjunto_df, 10)
print(top_enriquecimiento_conjunto)

# Gráfico de barras
library(ggplot2)
ggplot(top_enriquecimiento_conjunto, aes(x = reorder(Description, -GeneRatio_num), y = GeneRatio_num)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  coord_flip() +
  labs(title = "Top 10 procesos GO enriquecidos",
        x = "Descripción GO",
        y = "Número de genes") +
  theme_minimal()

```

Figura 7. Script análisis de enriquecimiento funcional mediante GO (clusterProfiler) en genes significativos para obtener los procesos biológicos más enriquecidos y genera las gráficas de barras con los procesos más representativos.

5. RESULTADOS

El diseño experimental realizado se basa en los datos de perfiles de expresión de células individuales en 22 muestras de tejido tumoral de 11 pacientes con HGSOE. Se obtuvieron dos muestras pareadas por paciente, una muestra tumoral antes del tratamiento y otra tras recibir quimioterapia neoadyuvante. Los datos utilizados constan de 51,786 células de las cuales el 17% (8,806/51,786) son células cancerosas, el 16% células estromales (8,045/51,786) y el 67% (34,935/51,786) células inmunes.

El estudio realizado tiene como objetivo caracterizar las poblaciones celulares y su comportamiento durante el tratamiento con el objetivo de identificar el estado celular en cada caso y así predecir, mediante biomarcadores, la respuesta a este tipo de quimioterapia.

5.1 Análisis de CNVs inferidas a partir de scRNAseq

Dada la heterogeneidad intratumoral que caracteriza al HGSOE, el estudio de las CNVs puede reflejar la presencia de ganancias o pérdidas de regiones cromosómicas específicas que pueden revelar la expansión o extinción de las células tumorales durante el tratamiento. Estos perfiles nos dieron la posibilidad de caracterizar las subpoblaciones celulares en el tejido analizado.

Los gráficos obtenidos mediante inferCNV constan de dos mapas de calor claramente diferenciados, en los que, la parte superior representa las células de referencia no cancerosas, que incluyen a las células inmunes y a las células estromales, y la parte inferior, donde se observa el perfil de las células tumorales (Ver Figura 8). En el eje X están representadas las regiones genómicas, abarcando todos los autosomas. Los colores que resaltan en cada región la ganancia o sobreexpresión, de color rojo, y la pérdida o subexpresión, de color azul. Además, se puede observar un dendograma en el lado izquierdo, cuyo objetivo es mostrar la agrupación de células en clústeres según su expresión. A simple vista se debería observar un patrón más estable en las células de referencia respecto a las células tumorales.

Como era de esperar, observamos una gran variabilidad en las CNVs detectadas en los 11 pacientes. Aunque se pueden observar patrones mayoritarios que coinciden en varios pacientes, como la duplicación completa o parcial de cromosoma 19 (11/11). El cromosoma 19 se ve afectado frecuentemente en los carcinomas de ovario, observándose duplicaciones en 19q o 19p en el 25% de los casos (Smebye et al., 2014). Por otro lado, también se observan ganancias del cromosoma 8 frecuentemente entre nuestra cohorte (8/11). Esto coincide con diversos estudios que reportan reordenamientos en el cromosoma 8 frecuentes entre diversos tipos de cáncer, incluyendo el OC (Dimova et al., 2009; Mark et al., 1999). En cuanto a las pérdidas completas o parciales, observamos algunos patrones comunes, como las deleciones en los cromosomas 4, 13 o 15. No obstante, se observa una gran variabilidad en las CNVs entre los pacientes (Ver Figuras S2-S11).

En el paciente 03 se pueden observar distintas regiones cromosómicas afectadas por ganancias y pérdidas características de las células tumorales. Este resultado revela un gran número de alteraciones que evidencia la inestabilidad genómica y la heterogeneidad tumoral en este paciente.

Además, podemos observar los distintos subclones detectados por inferCNV que muestran la diferencia de comportamiento en el medio tumoral.

En este caso, el mapa de calor muestra deleciones completas de los cromosomas 4, 9, 13, 18 y 22, y duplicaciones de gran tamaño en los cromosomas 8 y 19. Las deleciones de cromosomas como el 4, 13, 18 o 22 se han reportado con frecuencia en la literatura científica y se ha relacionado con un aumento de la progresión del tumor y un peor pronóstico (Mayr & Diebold, 2003).

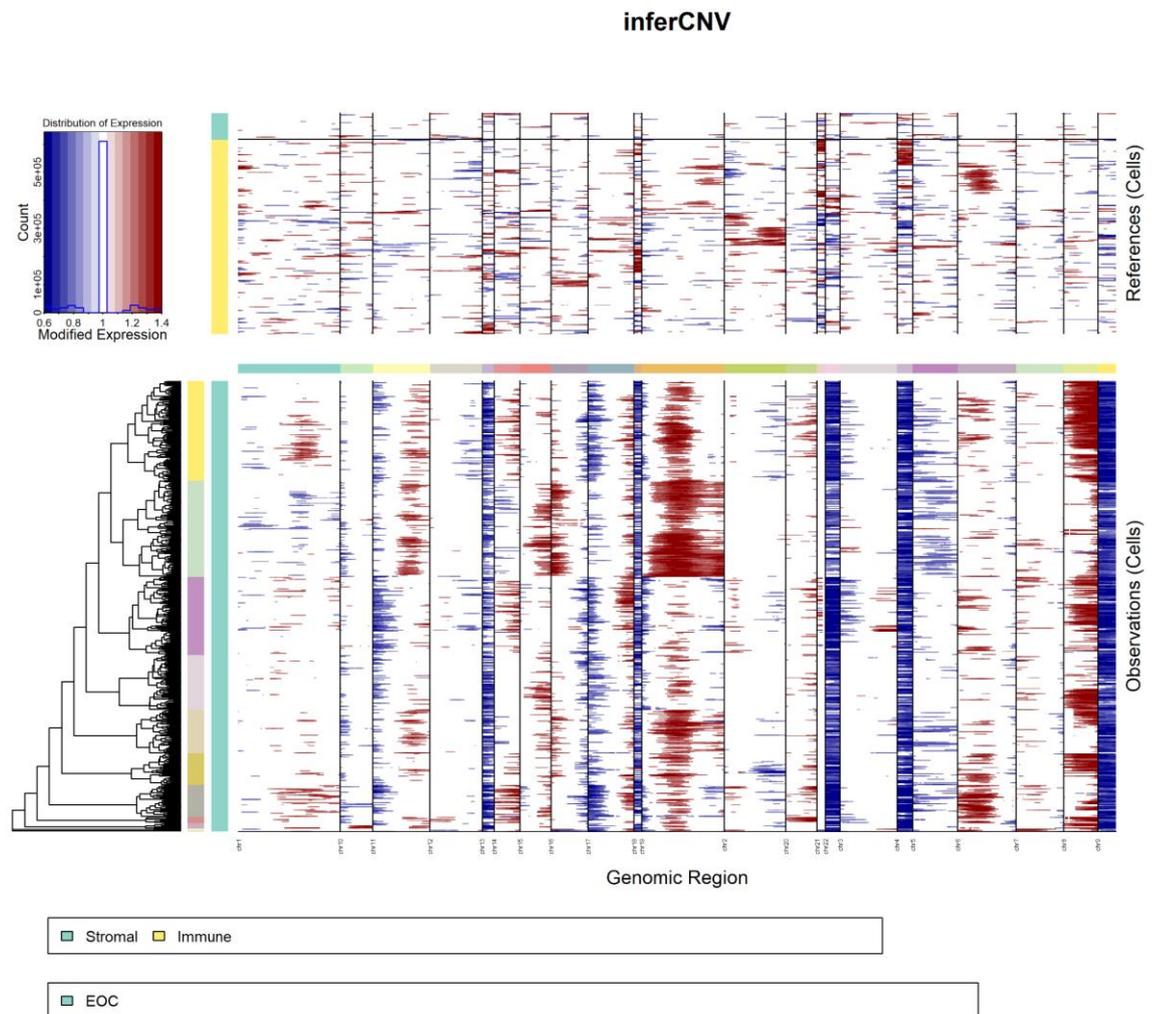


Figura 8. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 03.

5.2 Heterogeneidad tumoral y clusterización de subclones en HGSOC

Como ya se ha mencionado anteriormente, la heterogeneidad intratumoral que caracteriza al cáncer de ovario, y concretamente al HGSOC, se ve manifestada a la hora de analizar los perfiles genómicas o transcriptómicos de las muestras. Esta diferencia afecta de manera representativa al comportamiento del tumor a la hora de adaptarse al medio durante el tratamiento.

En cuanto a las células que presentan los datos analizados, se dividieron en dos grandes grupos: células cancerosas y células no cancerosas. Las células cancerosas, denominadas EOC forman las células tumorales que a su vez se dividen en 12 subtipos

celulares. Mientras que, las células no cancerosas, se dividen en células inmunes (células T, plasmáticas, NK, dendríticas, B, macrófagos, mastocitos y linfoides) y células estromales (fibroblastos, mesoteliales y endoteliales) (Ver Figuras 9, 10 y 11).

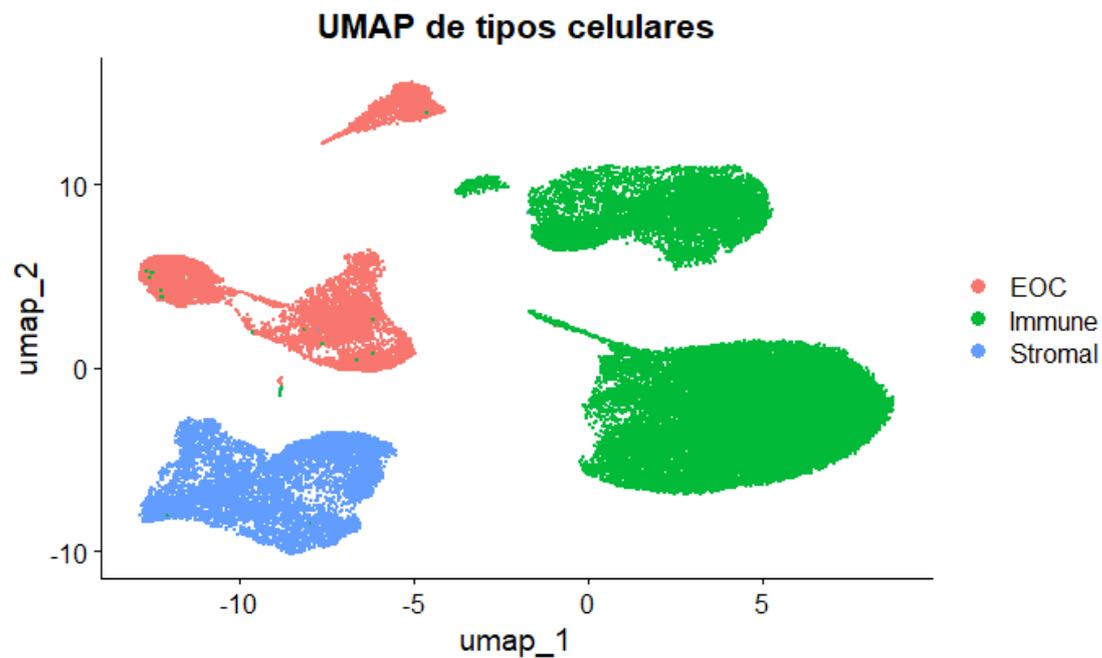


Figura 9. Gráfico UMAP que muestra las diferentes poblaciones celulares presentes en los tejidos tumorales. Las células cancerosas están representadas en rojo, mientras que las células inmunes y estromales en verde y azul, respectivamente.

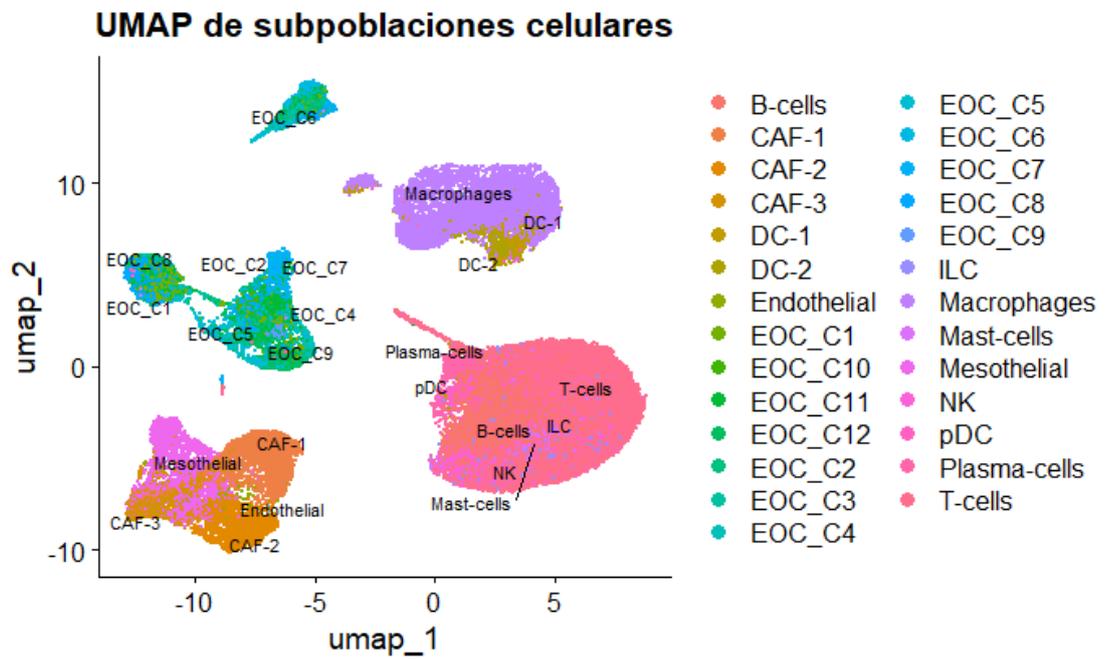


Figura 10. Gráfico UMAP que muestra las diferentes subpoblaciones celulares presentes en los tejidos tumorales.

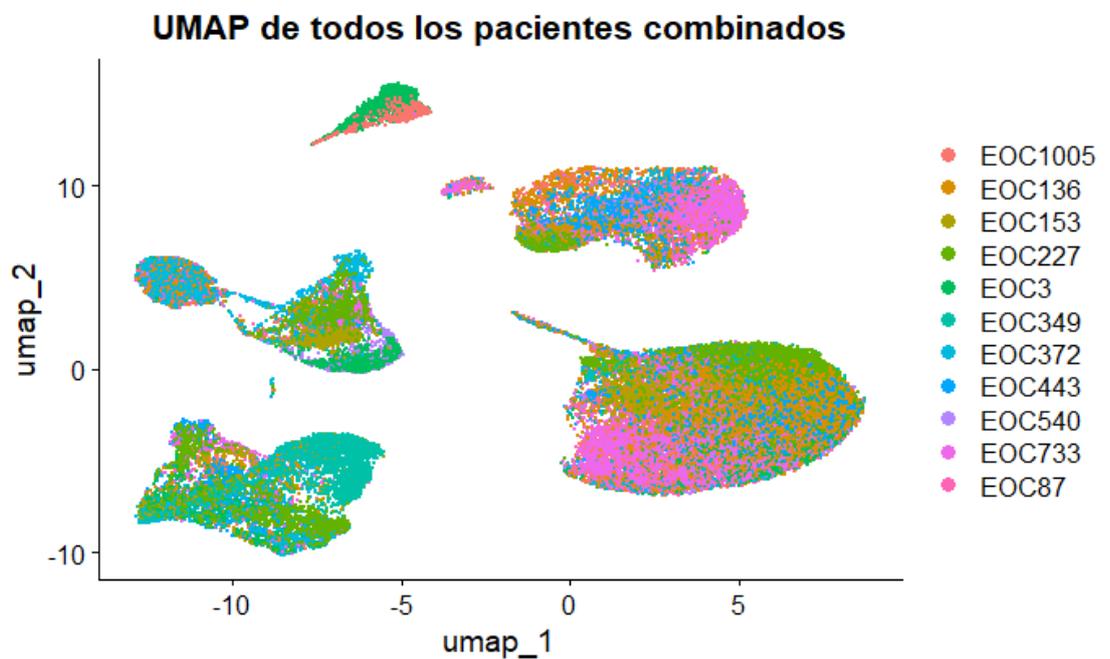


Figura 11. Gráfico UMAP que representa las células de los 11 pacientes.

Gracias a inferCNV hemos podido observar la evolución clonal durante el tratamiento en los pacientes con HGSOC. Tras la inferencia de CNVs, obtenemos datos de expresión de las distintas células y las clasificamos en subclones de antes y después del tratamiento, obteniendo tablas en las que podemos observar el cambio en el número de células respecto al total de células cancerosas en esa fase, clasificándose según su estado.

De los 11 pacientes, se obtuvieron un total de 117 subclones con más de 20 células totales. De todos los subclones seleccionados, se observó una tendencia a la expansión tras el tratamiento en un total de 63 clones. Por otro lado, se observó una extinción del número de células tras el tratamiento en un total de 54 clones (Ver Figura S1).

Los pacientes 136, 227 y 349, mostraron un gran porcentaje de subclones expandiéndose, lo que revela una falta de respuesta al tratamiento. Podríamos decir que en los pacientes 136, 227 y 349, prácticamente la mayoría de las células tumorales analizadas se han adaptado y son resistentes al tratamiento mediante quimioterapia neoadyuvante. Solo dos pacientes (153 y 540) muestran un porcentaje mayoritario de subclones extinguiéndose. El resto de los pacientes (03, 1005, 372, 443, 733 y 87) muestran porcentajes menos significativos (Ver Figura 12).

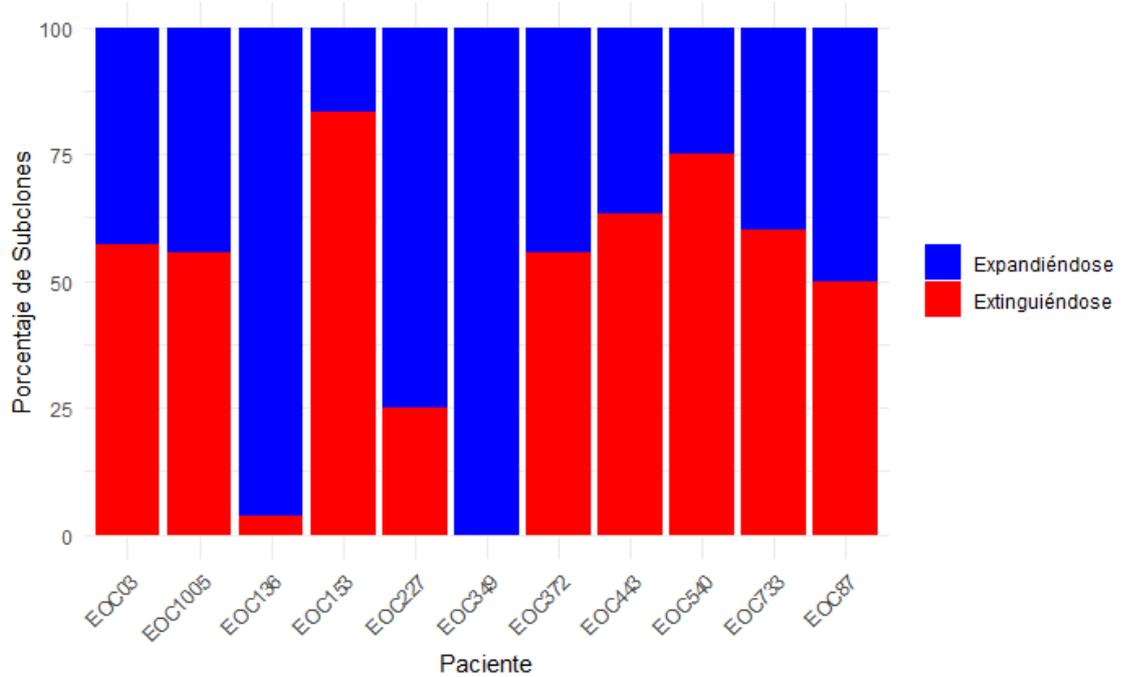


Figura 12. Distribución de subclones en expansión y en extinción en cada paciente. El gráfico de barras muestra en azul el porcentaje de subclones del paciente que se expanden, y en rojo, los que se extinguen.

Por otro lado, la clasificación de subclones para cada paciente, muestra, por lo general, una amplia diversidad de subclones entre paciente. Algunos pacientes como EOC03, EOC153 o EOC349, presentan un número reducido de subclones analizables, mientras que algunos como EOC136 o EOC443 mostrar una diversidad clonal analizable mucho mayor que otros pacientes (Ver Figura 13).

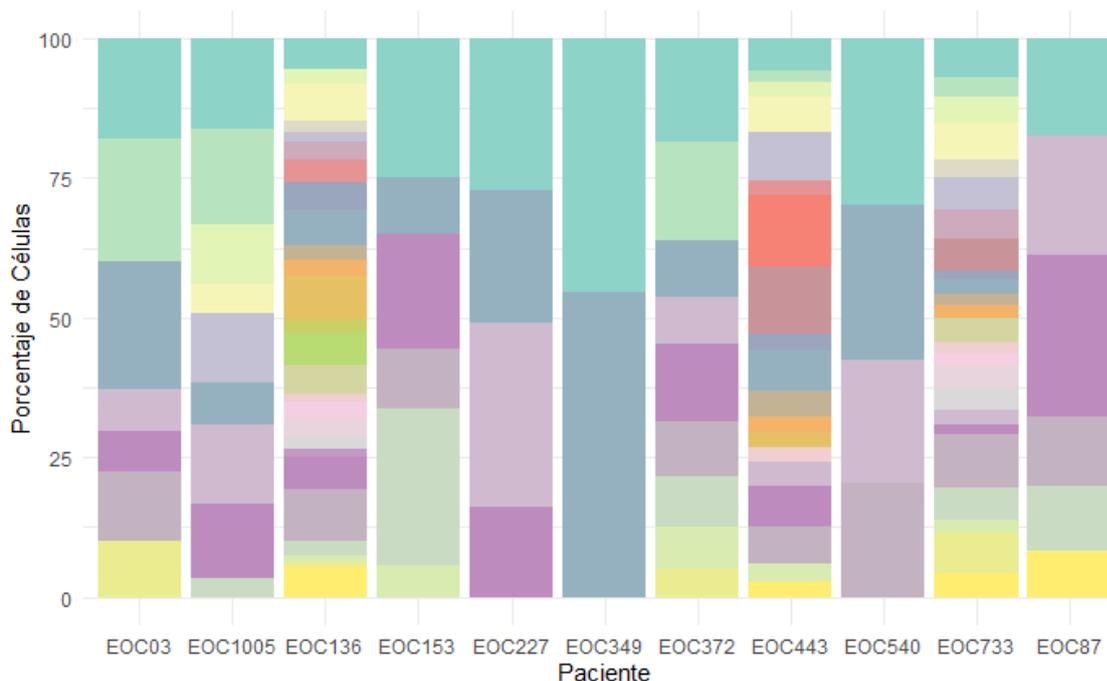


Figura 13. Stacked bar plot que muestra el porcentaje relativo de células correspondientes a los subclones de cada paciente. Cada color representa la proporción celular en cada subclón.

5.3 Análisis de expresión diferencial con Seurat

El análisis de expresión diferencial nos permitió identificar los mecanismos moleculares asociados a la expansión o extinción de los subclones. El objetivo fue el de reseñar los genes más significativos, sabiendo que un gen sobreexpresado tras el tratamiento puede estar asociado a resistencia al tratamiento, mientras que los genes que disminuyeron su expresión tras el tratamiento podrían estar asociados a una sensibilidad terapéutica significativa.

El valor que representa el cambio de expresión se observó en forma de *log fold changes* (\log_2FC). Aquellos genes que sufrieron un aumento en su valor se clasificaron como genes sobreexpresados (Ver Tabla 4), mientras que aquellos que disminuyeron se clasificaron como genes subexpresados (Ver Tabla 5).

La identificación de los genes diferencialmente expresados se obtuvo gracias a las funciones *FindMarkers* de Seurat. Se detectaron un total de 43,883 genes, de los cuales 30,896 mostraron un valor \log_2FC positivo y 12,987 un valor negativo (volcano plot).

Esto quiere decir que el 70% de los genes diferencialmente expresados muestran un patrón de expansión, mientras que, el 30% restante disminuyen su expresión. Los genes con valores negativos pueden indicar que su expresión ha disminuido tras el tratamiento y/o estar relacionados con subclones en extinción.

De los genes detectados, 134 fueron significativamente relevantes ($p\text{-val adj} < 0.05$) (Ver Tabla S1). Entre los genes más destacados podemos observar los genes *GTSF1* y *ATP* de los que no se encontró información relevante. No obstante, sus funciones biológicas pueden tener relación con el crecimiento tumoral (Gao et al., 2021).

Tabla 4. Top 10 genes más significativos en la expresión diferencial en el conjunto de subclones que se están expandiendo. Los genes aparecen ordenados según el p valor ajustado. La columna *avg_log2FC* muestra el cambio de expresión génica.

| Gene | avg_log2FC | p_val_adj |
|-----------|-------------|------------|
| GTSF11 | 6,518555215 | 3,2971E-24 |
| ATP6V0B3 | 1,590629618 | 8,286E-09 |
| UQCRH3 | 5,977208274 | 2,4862E-06 |
| JUN5 | 1,684186307 | 6,6706E-06 |
| MT1F3 | 2,536165111 | 2,105E-05 |
| TMEM595 | 2,050867574 | 3,8532E-05 |
| MT2A5 | 2,250573348 | 5,0917E-05 |
| CDC424 | 1,287724091 | 5,6945E-05 |
| SERPINA1 | 1,958293806 | 7,6274E-05 |
| SH3BGRL35 | 1,240468441 | 8,3784E-05 |

El gen JUN es un proto-oncogen involucrado en la proliferación, apoptosis y diferenciación celular. Además, se ha reportado su participación en vías como la vía MAPK, importante en la respuesta a factores de crecimiento y estrés celular. Diversos estudios han identificado la sobreexpresión del gen *JUN* como causante de la resistencia al tratamiento en ciertos tipos de cáncer, lo que implica que puede ser un gen candidato para la terapia (Wilczyński et al., 2023).

Otro gen de interés es *SERPINA1*. Este gen codifica para una proteína cuya función se ve implicada en la protección contra la degradación enzimática. Ciertos estudios han mencionado al gen *SERPINA1* junto a otros de la familia involucrados en procesos relacionados con la proliferación tumoral (B. Wang et al., 2022).

MT1F3 y *MT2A5* forman parte de la familia de las metalotioneínas, que participan en la homeostasis de metales pesados y en la protección frente al estrés oxidativo. Estos genes han sido asociados con resistencia a la quimioterapia y mecanismos relacionados con la supervivencia tumoral (Qi et al., 2023).

En cuanto a los genes subexpresados, encontramos genes relacionados con subunidades ribosomales. La falta de expresión de estos genes podría afectar a la traducción de las proteínas, derivando en muerte celular. Por otro lado, el gen *UBA523*, fundamental en la degradación de proteínas dañadas y cuya disminución de la expresión puede llevar a una acumulación de proteínas alteradas, dando lugar a la apoptosis.

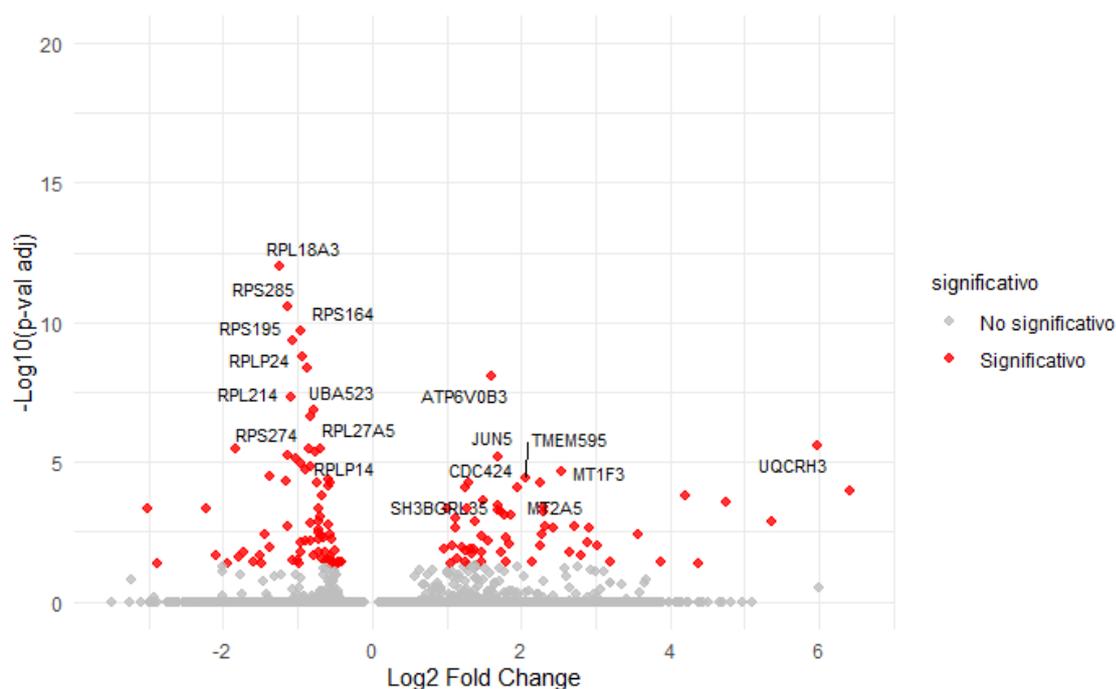


Figura 14. Volcano plot genes obtenidos en análisis de expresión diferencial. Los puntos rojos representan genes significativos ($pval < 0,05$). A la izquierda se identifican los 10 genes con patrón de extinción más significativos, y a la derecha los 10 genes en expansión.

Tabla 5. Top 10 genes más significativos en la expresión diferencial en el conjunto de subclones que se están extinguiendo. Los genes aparecen ordenados según el p valor ajustado. La columna avg_log2FC muestra el cambio de expresión génica.

| Gene | avg_log2FC | p_val_adj |
|---------|-------------|------------|
| RPL18A3 | -1,23988877 | 9,651E-13 |
| RPS285 | -1,13313747 | 2,802E-11 |
| RPS164 | -0,953004 | 1,9415E-10 |
| RPS195 | -1,05871486 | 4,3065E-10 |
| RPLP24 | -0,94106556 | 1,6156E-09 |
| UBA523 | -0,86742396 | 4,2719E-09 |
| RPL214 | -1,08717471 | 4,4105E-08 |
| RPL27A5 | -0,78628119 | 1,3326E-07 |
| RPS274 | -0,83415571 | 2,4248E-07 |
| RPLP14 | -0,70246887 | 3,1337E-06 |

5.4 Enriquecimiento funcional

Las células tumorales dependen del metabolismo celular para desarrollar una proliferación descontrolada y la capacidad de invasión a otros tejidos. La diferencia principal con las células normales es que, las células tumorales modifican estos procesos metabólicos para adaptarse al medio y sobrevivir a los cambios en el microambiente tumoral, principalmente por la quimioterapia. Por este motivo, explorar las vías metabólicas relevantes en diferentes tipos de cáncer es de gran importancia a la hora de comprender los mecanismos de resistencia o sensibilidad al tratamiento.

Un análisis complementario que puede dar una gran cantidad de información sobre los procesos celulares asociados a los mecanismos de resistencia o sensibilidad es el enriquecimiento funcional. Tras identificar los genes diferencialmente expresados, el siguiente paso fue el de observar qué rutas biológicas están implicadas en estos mecanismos.

Los resultados mostraron, por lo general, una tendencia a la expansión de genes que participan en vías relacionadas con la transcripción, regulando activamente la expresión génica, función fundamental para la proliferación celular, aspecto que explicaría la sobreexpresión a pesar del tratamiento. En cuanto a los genes que van reduciendo su actividad, se ha observado que aparecen relacionados con procesos como la biogénesis del ribosoma y la traducción.

En la Figura 15, se muestran los procesos biológicos más significativos entre los genes de los subclones que se expanden y los que se extinguen. Entre los más destacados, se observan procesos relacionados con la biogénesis de ribonucleoproteínas, la traducción de proteínas o el ensamblaje ribosómico. Estos procesos podrían estar asociados a la proliferación significativa de los subclones, lo que sería compatible con procesos relacionados con resistencia al tratamiento.

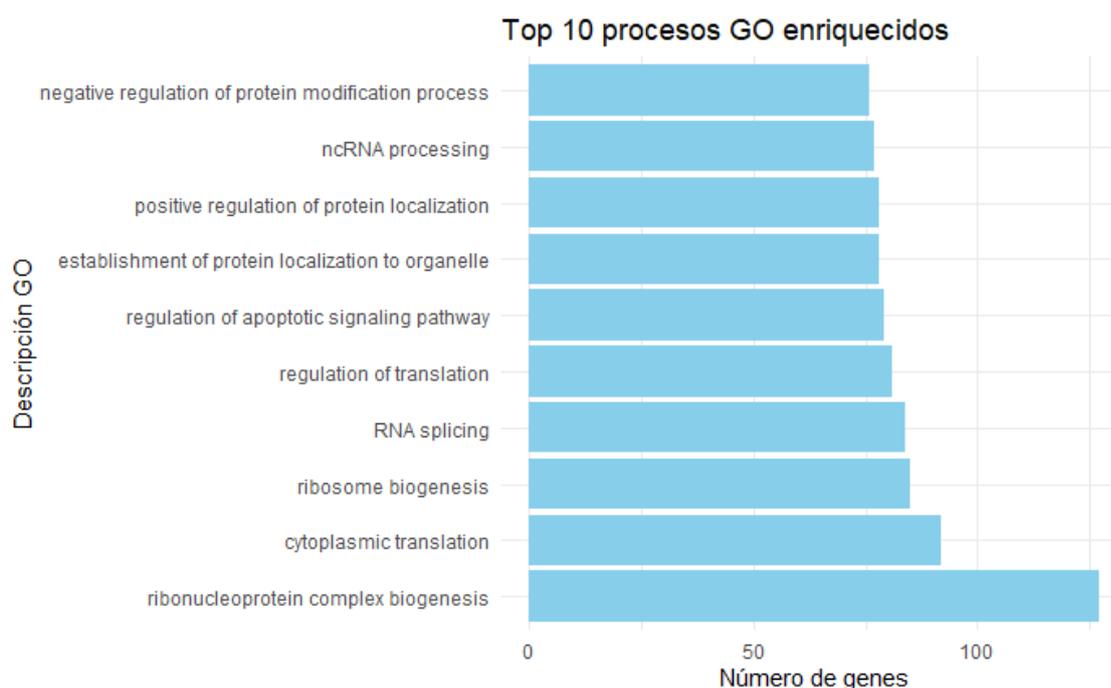


Figura 15. Top 10 procesos GO enriquecidos en células cancerosas de todos los pacientes conjuntos, según el Gene Ontology. El eje Y muestra la descripción de los procesos biológicos, y el eje X indica el número de genes de cada proceso.

La Figura 16, muestra los procesos biológicos enriquecidos en subclones de células cancerosas con tendencia a extinguirse durante el tratamiento. Se pueden identificar

procesos como la respuesta vírica, degradación proteica, regulación de la apoptosis o de la autofagia, asociados a un estrés celular como respuesta a la quimioterapia.

En cuanto a los procesos GO identificados en subclones que se expanden (Ver Figura 17), se ha observado una asociación significativa con los procesos de regulación del RNA, lo que parece influir en la resistencia al tratamiento en HGSOC (Rivera-Serrano et al., 2024).

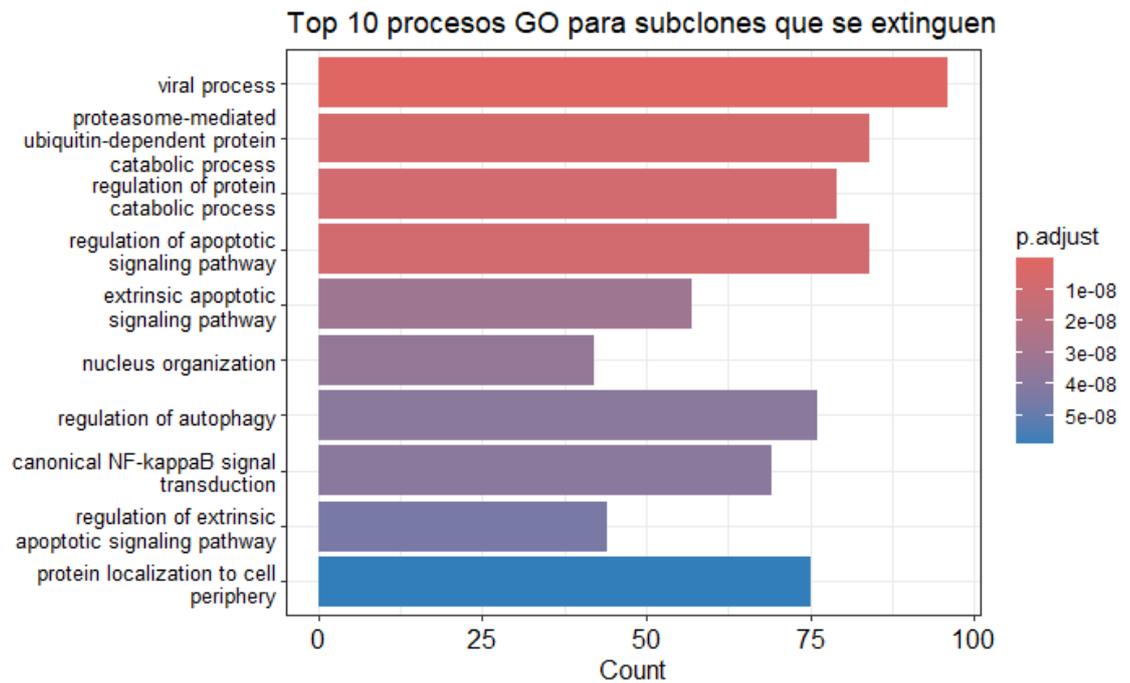


Figura 16. Top 10 procesos biológicos enriquecidos en subclones que se expanden en el conjunto de pacientes. La intensidad del color refleja el valor de p ajustado, que indica la significancia estadística de los mecanismos. El eje Y muestra la descripción de los procesos biológicos y el eje X el número de genes en cada proceso.

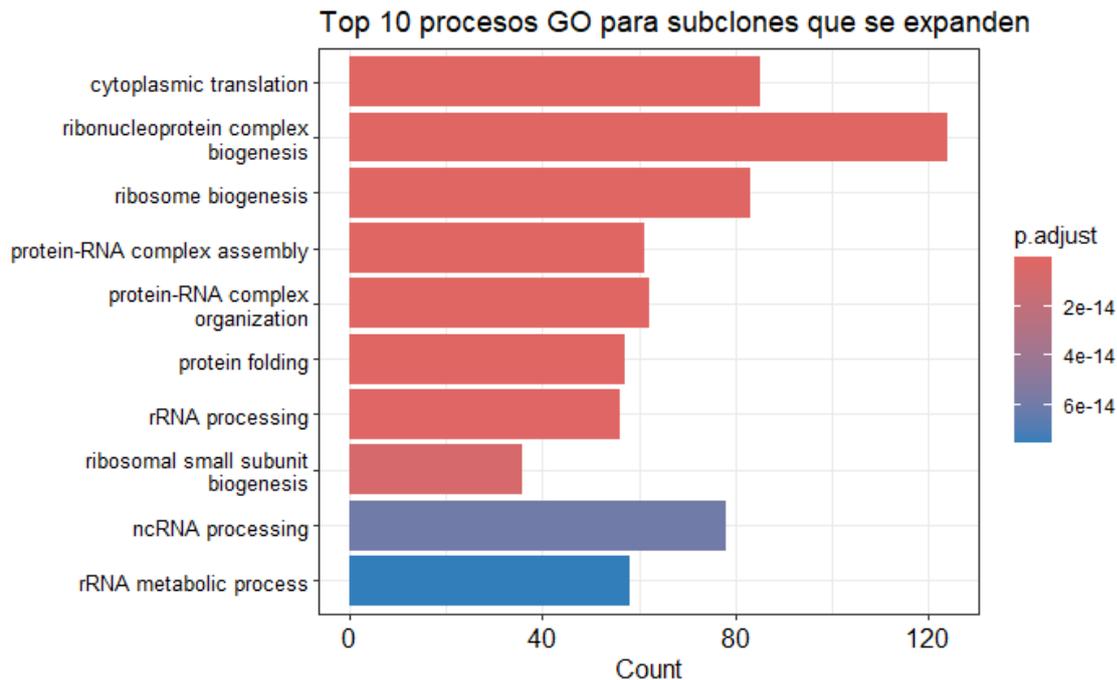


Figura 17. Top 10 procesos biológicos enriquecidos en subclones que se expanden en el conjunto de pacientes. La intensidad del color refleja el valor de *p* ajustado, que indica la significancia estadística de los mecanismos. El eje Y muestra la descripción de los

5. DISCUSIÓN Y CONCLUSIONES

Dada la importancia que tiene hoy en día reducir la tasa de mortalidad en los distintos tipos de cáncer, surgen innumerables vías de investigación que buscan reducir el mal pronóstico y dirigir las estrategias terapéuticas a un camino más personalizado. En este caso, el HGSOE se trata de un tipo de tumor que implica un gran desafío tanto para los clínicos como para los investigadores. Realizando trabajos de investigación como el que hemos elaborado, se pretende comprender mejor los mecanismos que subyacen esta enfermedad, con el fin de caracterizar a los pacientes y dirigir los tratamientos individualmente, optando a una mayor tasa de supervivencia y pronóstico.

Los análisis que realizamos sobre los 11 pacientes nos mostraron una gran variabilidad entre ellos a nivel clonal y celular. Uno de los principales desafíos al que nos enfrentamos es que, inferCNV es un método de inferencia con ciertas limitaciones. InferCNV necesita una combinación de genes para poder hacer una aproximación

segura, por lo que, los eventos clonales de menor tamaño no son detectados con esta técnica. Por este motivo, es posible que se omitan ciertos eventos relevantes para el tratamiento de la enfermedad. Asimismo, la inferencia realizada sobre estudios de RNA, es decir, de expresión, son menos precisos que los métodos basados en DNA. No obstante, hoy en día, surgen opciones de combinación entre scRNAseq y scDNAseq en las mismas células como es el DNTR-seq (por sus siglas en inglés, *direct nuclear tagmentation and RNA sequencing*), que permite obtener una información sobre las células mucho más completa con el objetivo final de conseguir una información sobre la evolución clonal más precisa (Zachariadis et al., 2020).

No obstante, inferCNV nos proporcionó información ciertamente relevante en cuanto a las regiones con ganancias y pérdidas. Se pudo asociar ciertas CNVs comunes en estos tipos de cáncer como la trisomía 19. Sin embargo, la ventaja del uso de esta herramienta fue la capacidad para poder extraer la información clonal de estos resultados.

En pacientes como EOC03, se pudo observar diferencia de comportamiento entre los clones seleccionados. Mientras los subclones s1 y s2 mostraron una clara tendencia a expandirse tras el tratamiento, otros subclones como s10 o s8 evidenciaron una pérdida en el porcentaje de células cancerosas tras el tratamiento.

La heterogeneidad celular en muestras de HGSOC es bien conocida, nuestros resultados fueron compatibles con estas características. Observamos una respuesta a la quimioterapia muy heterogénea entre los 11 pacientes. No se encontraron patrones similares entre los pacientes estudiados en cuanto a número de células y diversidad clonal. Se observó algún paciente como EOC136, en el que se identificaron 26 subclones analizables, de los cuales más del 95% (25/26) manifestaron una clara tendencia a la expansión, mientras que solo uno de los subclones evidenció una reducción significativa, lo que sugiere que en este paciente pueden existir mecanismos de resistencia que estarían posicionando a los subclones en una ventaja evolutiva frente al resto. Por otro lado, se identificaron pacientes como EOC540 o EOC733 con gran mayoría de los subclones en extinción, lo que se podría interpretar como un efecto favorable del tratamiento sobre estos pacientes.

Los clones y los subclones son agrupaciones celulares, en este caso tumorales, que comparten unas características concretas. Los tumores surgen de una célula única alterada que acumula mutaciones adicionales creciendo y dividiéndose, hasta formar subpoblaciones de células tumorales con diferentes características. Los subclones van apareciendo con el cúmulo de alteraciones que van surgiendo durante el crecimiento del tumor. Por lo que, cada subclón está formado por un clúster de genes con unas características, y, en consecuencia, un comportamiento distinto al de otros subclones. Un ejemplo claro es la resistencia a la quimioterapia. Durante el proceso, pueden surgir subclones que aumenten su expresión durante el tratamiento respecto a otros subclones por una mejor adaptación al medio. Esta diferencia se puede dar por el comportamiento de esta agrupación celular, que puede caracterizarse, por ejemplo, por realizar una macroautofagia que permita eliminar las proteínas dañadas, autoreciclándose, manteniendo así su viabilidad y capacidad de proliferación (Zhan et al., 2016). Por lo tanto, la heterogeneidad intratumoral que caracteriza este tipo de tumor da lugar a una ventaja evolutiva de ciertos subclones respecto a otros, dando lugar a una expansión clonal que domina el tumor.

A nivel funcional, la diferenciación entre las células cancerosas y las células normales o sanas, se da comparando diversas señas de identidad como la resistencia a la apoptosis, la inmortalidad replicativa, angiogénesis, capacidad de mantener la señalización proliferativa, la capacidad de activación de invasión y metástasis, evasión inmunológica, plasticidad fenotípica, senescencia celular, inestabilidad genómica, reprogramación epigenética y otras más (Hanahan, 2022; Morden et al., 2021).

Los subgrupos clonales pueden presentar diferentes mecanismos de respuesta a la acción de la terapia. Distintos estudios han mostrado la dificultad en el análisis de este comportamiento y cómo otros factores pueden verse envueltos en los mecanismos de resistencia. El cambio de características en el microambiente tumoral, las diferencias de presiones hidrostáticas o la calidad de los vasos sanguíneos del tumor, pueden dificultar la acción de los fármacos sobre las células cancerosas (Bai et al., 2021). Sumado a esto, el microambiente tumoral puede proteger a las células cancerosas mediante la secreción de factores, lo que supone un aspecto de interés en la investigación sobre la resistencia a fármacos. En el caso del OC, la caracterización de la heterogeneidad

tumoral nos provee una oportunidad clave para determinar las características moleculares, y así observar los comportamientos y el pronóstico.

De acuerdo con diversos estudios, la expansión de los subclones tras el tratamiento se ha relacionado con mecanismos de resistencia, mientras que la extinción de estos subclones podría estar asociada a la eficacia del tratamiento.

El análisis de expresión diferencial nos dio la posibilidad de identificar un número de genes expresados significativamente en las muestras pre y post tratamiento. Se realizaron búsquedas bibliográficas sobre los genes hallados como diferencialmente expresados para relacionarlos con los posibles mecanismos de resistencia o sensibilidad al tratamiento. Entre los genes estudiados, nos encontramos genes como *JUN*, *SERPINA1*, *MT1F3* o *MT2A5*, relacionados con subclones en expansión y asociados a proliferación tumoral y mecanismos de resistencia a quimioterapia en la literatura científica. Estos genes junto a otros han sido relacionados con activación de vías de señalización que hacen que las células puedan sobrevivir durante el tratamiento. Del otro lado, se identificaron genes subexpresados en subclones en extinción, sobre todo, relacionados con complejos ribosomales, cuya falta de expresión puede derivar en apoptosis celular.

El enriquecimiento funcional nos permitió interpretar los resultados obtenidos en la expresión diferencial identificando vías moleculares y rutas biológicas significativamente afectadas durante el tratamiento. El estudio reveló que los genes que tienen a expandirse tras el tratamiento están involucrados, por lo general, en biogénesis de ribonucleoproteínas, la traducción de proteínas, la proliferación celular y procesos de mantenimiento del DNA. El crecimiento de las células tumorales y la supervivencia de estas depende de rutas biológicas como estas. Por otro lado, los genes representativos de los clones que se extinguían mostraron una involucración en vías asociadas con la apoptosis, la respuesta al estrés y procesos virales, lo que indicaría que estos genes pertenecen a subclones que son más susceptibles al tratamiento (Stewart, 2007).

Los procesos virales y catabólicos de proteínas dependientes de ubiquitina están asociados a la degradación de proteínas dañadas, mediante el sistema ubiquitina-

proteasoma. Diversos estudios han mostrado que estos procesos juegan un papel fundamental en la supervivencia de las células tumorales, dado que depende de estos para eliminar las proteínas dañadas que se acumulan debido al estrés del microambiente tumoral durante el tratamiento. Cuando estos procesos sufren modificaciones, la acumulación de proteínas deterioradas, puede inducir la apoptosis de las células tumorales. Por este motivo, se ha descrito como posible estrategia terapéutica la inhibición del proteasoma, con el fin de inducir la muerte de las células tumorales en HGSOE (Manasanch & Orlowski, 2017).

Por otro lado, el enriquecimiento de la regulación de la apoptosis indica que estas células se ven sometidas a un estrés significativo que activa esas vías. La quimioterapia neoadyuvante induce apoptosis en células tumorales sensibles al tratamiento y en aquellos subclones que no logran adaptarse desarrollando la resistencia que otros sí.

En el HGSOE, las células tumorales están sometidas a un alto grado de estrés por el microambiente tumoral hostil que presentan y los tratamientos con quimioterapia. En los subclones susceptibles al tratamiento, este estrés prolongado deriva en una apoptosis dado que la autofagia no es suficiente para manejarlo. Este proceso se ha observado en subclones que no son capaces de adaptarse al microambiente tumoral tras el tratamiento, provocando su extinción. En la literatura científica, se pueden encontrar publicaciones que demuestran como la autofagia y la inhibición de las rutas relacionadas con la supervivencia celular, pueden reducir la proliferación, instando a la apoptosis de estos subclones (Hu et al., 2017).

Las células que no se adaptan a estas condiciones de estrés, activan las vías de apoptosis y los procesos de autofagia excesiva. No obstante, la regulación de la autofagia puede tener un papel favorable en los subclones que se expanden, ya que una activación controlada permite a estos subclones liberarse de las proteínas dañadas. Este proceso de adaptación permite, al reciclar los componentes de las proteínas eliminadas, mantener su viabilidad y seguir proliferando. Diversos estudios han mostrado cómo las células tumorales usan la autofagia como mecanismo de resistencia a los tratamientos, y uno de los ejemplos son las células cancerosas de HGSOE (Peracchio et al., 2012; Phadte et al., 2024).

Otros procesos involucrados en la expansión clonal durante el tratamiento son la traducción citoplasmática y la biogénesis ribosomal. Estos procesos son de gran importancia en la síntesis proteica, permitiendo a las células cancerosas seguir proliferando. El aumento en la biogénesis ribosomal y proteica se ha relacionado directamente con un aumento de la proliferación de células tumorales en HGSOC.

Además, la matriz extracelular también juega un papel fundamental en la proliferación de células cancerosas y su capacidad de invasión y supervivencia (Hastings et al., 2019). Las vías implicadas en la matriz extracelular se han asociado a metástasis, progresión de la enfermedad y a resistencia al tratamiento (Pietilä et al., 2021). Los procesos identificados mediante el enriquecimiento funcional indican una mayor actividad de ensamblaje proteico, lo que se vincula con una capacidad para alterar su entorno. La remodelación de la matriz extracelular es clave para permitir la migración y expansión de las células cancerosas.

Cabe destacar las limitaciones halladas en la realización de este trabajo. Una de las principales restricciones de este estudio fue el tamaño muestral. Aunque pudimos observar tendencias a expansión o extinción clonal en algunos pacientes, la variabilidad observada entre estos pacientes implica una falta de muestras que aclaren estos resultados. Igualmente, la proporción clonal puede variar según la localización de la muestra obtenida, por este motivo, se extrajeron muestras únicamente del mesenterio, epiplón y peritoneo antes del tratamiento, y en el epiplón, peritoneo y tumor tras el tratamiento.

Además, aunque se han observado genes diferencialmente expresados tanto en subclones en expansión como en extinción, no se ha realizado un análisis exhaustivo de las funciones de dichos genes y su implicación o no en la resistencia al tratamiento. Asimismo, la identificación de rutas biológicas significativamente relevantes en subclones que podrían estar generando mecanismos de resistencia tampoco han sido íntegramente estudiadas. No obstante, esto sugiere una aproximación a menor escala de la caracterización clonal en muestras de HGSOC. Todo ello muestra la creciente necesidad de ampliar los estudios en estos ámbitos con el fin de esclarecer la

importancia de estas características en el pronóstico y tratamiento de las pacientes con OC.

La caracterización clonal en muestras tumorales y la identificación tanto de genes como rutas biológicas relevantes son una herramienta de gran valor para obtener perfiles individualizados de los pacientes, guiando decisiones terapéuticas más personalizadas y, por lo tanto, mejorando la respuesta a tratamiento y aumentando las tasas de supervivencia.

En conclusión, este trabajo ha conseguido dar una visión sobre la dinámica clonal en el OC, concretamente en muestras de HGSOC, destacando la relevancia de la heterogeneidad intratumoral y el desafío que implica. Además, hemos tratado con gran interés el comportamiento de los subclones tras la quimioterapia neoadyuvante, identificando genes y rutas biológicas que pueden ser potenciales biomarcadores para predecir la evolución clonal y desarrollar nuevas terapias personalizadas. Por último, se ha demostrado como la scRNAseq puede ser una gran herramienta para esclarecer la complejidad genética del cáncer. No obstante, sería interesante realizar estudios similares con un mayor número de pacientes para observar con mayor precisión genes y mecanismos significativos. A su vez, el uso de DNTR-seq podría dar una información más completa con el fin de realizar una caracterización más precisa.

8. DECLARACIÓN DEL USO DE HERRAMIENTAS DE IA

El autor y tutor del presente trabajo declaran haber usado tecnologías asistidas por IA en el proceso de redacción. Durante el proceso práctico en RStudio se utilizó ChatGPT-4o para solucionar errores en el código. Además, se utilizó ChatGPT-4o en la redacción del texto con el fin de corregir faltas de ortografía, coherencia y traducción. Tras el uso de esta herramienta, los autores revisaron y editaron el contenido obtenido según fue necesario y, por lo tanto, asumen toda la responsabilidad por el contenido publicado.

9. BIBLIOGRAFÍA

9.1 Artículos y capítulos de libros

Al-Sukhni, W., Joe, S., Lionel, A. C., Zwingerman, N., Zogopoulos, G., Marshall, C. R., Borgida, A., Holter, S., Gropper, A., Moore, S., Bondy, M., Klein, A. P., Petersen, G. M., Rabe, K. G., Schwartz, A. G., Syngal, S., Scherer, S. W., & Gallinger, S. (2012). Identification of germline genomic copy number variation in familial pancreatic cancer. *Human Genetics*, *131*(9), 1481-1494. <https://doi.org/10.1007/s00439-012-1183-1>

Bai, X., Li, Y., Zeng, X., Zhao, Q., & Zhang, Z. (2021). Single-cell sequencing technology in tumor research. *Clinica Chimica Acta*, *518*, 101-109. <https://doi.org/10.1016/j.cca.2021.03.013>

Bowtell, D. D., Böhm, S., Ahmed, A. A., Aspúria, P.-J., Bast, R. C., Beral, V., Berek, J. S., Birrer, M. J., Blagden, S., Bookman, M. A., Brenton, J. D., Chiappinelli, K. B., Martins, F. C., Coukos, G., Drapkin, R., Edmondson, R., Fotopoulou, C., Gabra, H., Galon, J., ... Balkwill, F. R. (2015). Rethinking ovarian cancer II: Reducing mortality from high-grade serous ovarian cancer. *Nature Reviews Cancer*, *15*(11), 668-679. <https://doi.org/10.1038/nrc4019>

Chai, C., Liang, L., Mikkelsen, N. S., Wang, W., Zhao, W., Sun, C., Bak, R. O., Li, H., Lin, L., Wang, F., & Luo, Y. (2024). Single-cell transcriptome analysis of epithelial, immune, and stromal signatures and interactions in human ovarian cancer. *Communications Biology*, *7*(1), 131. <https://doi.org/10.1038/s42003-024-05826-1>

Denisenko, E., De Kock, L., Tan, A., Beasley, A. B., Beilin, M., Jones, M. E., Hou, R., Muirí, D. Ó., Bilic, S., Mohan, G. R. K. A., Salfinger, S., Fox, S., Hmon, K. P. W., Yeow, Y., Kim, Y., John, R., Gilderman, T. S., Killingbeck, E., Gray, E. S., ... Forrest, A. R. R. (2024). Spatial transcriptomics reveals discrete tumour microenvironments and autocrine loops within ovarian cancer subclones. *Nature Communications*, *15*(1), 2860. <https://doi.org/10.1038/s41467-024-47271-y>

Dimova, I., Orsetti, B., Negre, V., Rouge, C., Ursule, L., Lasorsa, L., Dimitrov, R., Doganov, N., Toncheva, D., & Theillet, C. (2009). Genomic Markers for Ovarian Cancer at Chromosomes 1, 8 and 17 Revealed by Array CGH Analysis. *Tumori Journal*, *95*(3), 357-366. <https://doi.org/10.1177/030089160909500315>

Fan, J., Lee, H.-O., Lee, S., Ryu, D., Lee, S., Xue, C., Kim, S. J., Kim, K., Barkas, N., Park, P. J., Park, W.-Y., & Kharchenko, P. V. (2018). Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Research*, *28*(8), 1217-1227. <https://doi.org/10.1101/gr.228080.117>

Gao, R., Bai, S., Henderson, Y. C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A., Wang, F., Shaitelman, S. F., Wang, J. R., Chen, K., Moulder, S., Lai, S. Y., & Navin, N. E. (2021). Delineating copy number and clonal substructure in human tumors from

single-cell transcriptomes. *Nature Biotechnology*, 39(5), 599-608.
<https://doi.org/10.1038/s41587-020-00795-2>

Glasgow, M. A., Argenta, P., Abrahante, J. E., Shetty, M., Talukdar, S., Croonquist, P. A., Khalifa, M. A., & Starr, T. K. (2019). Biological Insights into Chemotherapy Resistance in Ovarian Cancer. *International Journal of Molecular Sciences*, 20(9), 2131.
<https://doi.org/10.3390/ijms20092131>

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1), 31-46. <https://doi.org/10.1158/2159-8290.CD-21-1059>

Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, 9(1), 75. <https://doi.org/10.1186/s13073-017-0467-4>

Havasi, A., Cainap, S. S., Havasi, A. T., & Cainap, C. (2023). Ovarian Cancer—Insights into Platinum Resistance and Overcoming It. *Medicina*, 59(3), 544. <https://doi.org/10.3390/medicina59030544>

Hu, J.-L., Hu, X.-L., Guo, A.-Y., Wang, C.-J., Wen, Y.-Y., & Cang, S.-D. (2017). Endoplasmic reticulum stress promotes autophagy and apoptosis and reverses chemoresistance in human ovarian cancer cells. *Oncotarget*, 8(30), 49380-49394. <https://doi.org/10.18632/oncotarget.17673>

Jayson, G. C., Kohn, E. C., Kitchener, H. C., & Ledermann, J. A. (2014). Ovarian cancer. *The Lancet*, 384(9951), 1376-1388. [https://doi.org/10.1016/S0140-6736\(13\)62146-7](https://doi.org/10.1016/S0140-6736(13)62146-7)

Ju, W., Yoo, B. C., Kim, I.-J., Kim, J. W., Kim, S. C., & Lee, H. P. (2009). Identification of Genes With Differential Expression in Chemoresistant Epithelial Ovarian Cancer Using High-Density Oligonucleotide Microarrays. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*, 18(2), 47-56. <https://doi.org/10.3727/096504009789954672>

Keyvani, V., Farshchian, M., Esmaeili, S.-A., Yari, H., Moghbeli, M., Nezhad, S.-R. K., & Abbaszadegan, M. R. (2019). Ovarian cancer stem cells and targeted therapy. *Journal of Ovarian Research*, 12(1), 120. <https://doi.org/10.1186/s13048-019-0588-z>

Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A., & Marioni, J. C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nature Communications*, 6(1), 8687. <https://doi.org/10.1038/ncomms9687>

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, 58(4), 610-620. <https://doi.org/10.1016/j.molcel.2015.04.005>

Li, G.-M., Xiao, G.-Z., Qin, P.-F., Wan, X.-Y., Fu, Y.-J., Zheng, Y.-H., Luo, M.-Y., Ren, D.-L., Liu, S.-P., Chen, H.-X., & Lin, H.-C. (2022). Single-Cell RNA Sequencing Reveals

Heterogeneity in the Tumor Microenvironment between Young-Onset and Old-Onset Colorectal Cancer. *Biomolecules*, 12(12), 1860. <https://doi.org/10.3390/biom12121860>

Manasanch, E. E., & Orłowski, R. Z. (2017). Proteasome inhibitors in cancer therapy. *Nature Reviews Clinical Oncology*, 14(7), 417-433. <https://doi.org/10.1038/nrclinonc.2016.206>

Mark, H. F. L., Afify, A. M., Werness, B. A., Das, S., Mark, S., & Samy, M. (1999). Trisomy 8 in Stage I and Stage III Ovarian Cancer Detected by Fluorescence in Situ Hybridization. *Experimental and Molecular Pathology*, 66(1), 76-81. <https://doi.org/10.1006/exmp.1999.2241>

Mayr, D., & Diebold, J. (2003). Chromosomal changes in ovarian cancer. *Current Diagnostic Pathology*, 9(5), 328-332. [https://doi.org/10.1016/S0968-6053\(03\)00050-4](https://doi.org/10.1016/S0968-6053(03)00050-4)

McNeish, I. A., Ledermann, J. A., Webber, L., James, L., Kaye, S. B., Hall, M., Hall, G., Clapp, A., Earl, H., Banerjee, S., Kristeleit, R., Raja, F., Feeney, A., Lawrence, C., Dawson-Athey, L., Persic, M., & Khan, I. (2014). A randomised, placebo-controlled trial of weekly paclitaxel and saracatinib (AZD0530) in platinum-resistant ovarian, fallopian tube or primary peritoneal cancer. *Annals of Oncology*, 25(10), 1988-1995. <https://doi.org/10.1093/annonc/mdu363>

Morden, C. R., Farrell, A. C., Sliwowski, M., Lichtensztein, Z., Altman, A. D., Nachtigal, M. W., & McManus, K. J. (2021). Chromosome instability is prevalent and dynamic in high-grade serous ovarian cancer patient samples. *Gynecologic Oncology*, 161(3), 769-778. <https://doi.org/10.1016/j.ygyno.2021.02.038>

Olbrecht, S., Busschaert, P., Qian, J., Vanderstichele, A., Loverix, L., Van Gorp, T., Van Nieuwenhuysen, E., Han, S., Van Den Broeck, A., Coosemans, A., Van Rompuy, A.-S., Lambrechts, D., & Vergote, I. (2021). High-grade serous tubo-ovarian cancer refined with single-cell RNA sequencing: Specific cell subtypes influence survival and determine molecular subtype classification. *Genome Medicine*, 13(1), 111. <https://doi.org/10.1186/s13073-021-00922-x>

Ortiz, M., Wabel, E., Mitchell, K., & Horibata, S. (2022). Mechanisms of chemotherapy resistance in ovarian cancer. *Cancer Drug Resistance*. <https://doi.org/10.20517/cdr.2021.147>

Patch, A.-M., Christie, E. L., Etemadmoghadam, D., Garsed, D. W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P. J., Kassahn, K. S., Newell, F., Quinn, M. C. J., Kazakoff, S., Quek, K., Wilhelm-Benartzi, C., Curry, E., Leong, H. S., The Australian Ovarian Cancer Study Group, ... Bowtell, D. D. L. (2015). Whole-genome characterization of chemoresistant ovarian cancer. *Nature*, 521(7553), 489-494. <https://doi.org/10.1038/nature14410>

Peracchio, C., Alabiso, O., Valente, G., & Isidoro, C. (2012). Involvement of autophagy in ovarian cancer: A working hypothesis. *Journal of Ovarian Research*, 5(1), 22. <https://doi.org/10.1186/1757-2215-5-22>

Phadte, P., Bishnu, A., Dey, P., M, M., Mehrotra, M., Singh, P., Chakrabarty, S., Majumdar, R., Rekhi, B., Patra, M., De, A., & Ray, P. (2024). Autophagy-mediated ID1 turnover dictates chemo-resistant fate in ovarian cancer stem cells. *Journal of Experimental & Clinical Cancer Research*, *43*(1), 222. <https://doi.org/10.1186/s13046-024-03147-z>

Pietilä, E. A., Gonzalez-Molina, J., Moyano-Galceran, L., Jamalzadeh, S., Zhang, K., Lehtinen, L., Turunen, S. P., Martins, T. A., Gultekin, O., Lamminen, T., Kaipio, K., Joneborg, U., Hynninen, J., Hietanen, S., Grénman, S., Lehtonen, R., Hautaniemi, S., Carpén, O., Carlson, J. W., & Lehti, K. (2021). Co-evolution of matrisome and adaptive adhesion dynamics drives ovarian cancer chemoresistance. *Nature Communications*, *12*(1), 3904. <https://doi.org/10.1038/s41467-021-24009-8>

Qi, Y.-N., Liu, Z., Hong, L.-L., Li, P., & Ling, Z.-Q. (2023). Methyltransferase-like proteins in cancer biology and potential therapeutic targeting. *Journal of Hematology & Oncology*, *16*(1), 89. <https://doi.org/10.1186/s13045-023-01477-7>

Rivera-Serrano, M., Flores, M., Rolon, J., Reyes, V., Valiyeva, F., & Vivas-Mejia, P. E. (2024). Abstract 2989: MicroRNA profiling in human samples and cell lines uncovers eight key players in the cisplatin resistance of HGSOC. *Cancer Research*, *84*(6_Supplement), 2989-2989. <https://doi.org/10.1158/1538-7445.AM2024-2989>

Schwarz, R. F., Ng, C. K. Y., Cooke, S. L., Newman, S., Temple, J., Piskorz, A. M., Gale, D., Sayal, K., Murtaza, M., Baldwin, P. J., Rosenfeld, N., Earl, H. M., Sala, E., Jimenez-Linan, M., Parkinson, C. A., Markowitz, F., & Brenton, J. D. (2015). Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis. *PLOS Medicine*, *12*(2), e1001789. <https://doi.org/10.1371/journal.pmed.1001789>

Senkowski, W., Gall-Mas, L., Falco, M. M., Li, Y., Lavikka, K., Kriegbaum, M. C., Oikkonen, J., Bulanova, D., Pietras, E. J., Voßgröne, K., Chen, Y.-J., Erkan, E. P., Dai, J., Lundgren, A., Grønning Høg, M. K., Larsen, I. M., Lamminen, T., Kaipio, K., Huvila, J., ... Wennerberg, K. (2023). A platform for efficient establishment and drug-response profiling of high-grade serous ovarian cancer organoids. *Developmental Cell*, *58*(12), 1106-1121.e7. <https://doi.org/10.1016/j.devcel.2023.04.012>

Serin Harmanci, A., Harmanci, A. O., & Zhou, X. (2020). CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nature Communications*, *11*(1), 89. <https://doi.org/10.1038/s41467-019-13779-x>

Smebye, M. L., Sveen, A., Haugom, L., Davidson, B., Tropé, C. G., Lothe, R. A., Heim, S., Skotheim, R. I., & Micci, F. (2014). Chromosome 19 rearrangements in ovarian carcinomas: Zinc finger genes are particularly targeted. *Genes, Chromosomes and Cancer*, *53*(7), 558-567. <https://doi.org/10.1002/gcc.22166>

Stegle, O., Teichmann, S. A., & Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, *16*(3), 133-145. <https://doi.org/10.1038/nrg3833>

Stewart, D. J. (2007). Mechanisms of resistance to cisplatin and carboplatin. *Critical Reviews in Oncology/Hematology*, 63(1), 12-31. <https://doi.org/10.1016/j.critrevonc.2007.02.001>

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5), 377-382. <https://doi.org/10.1038/nmeth.1315>

Torre, L. A., Trabert, B., DeSantis, C. E., Miller, K. D., Samimi, G., Runowicz, C. D., Gaudet, M. M., Jemal, A., & Siegel, R. L. (2018). Ovarian cancer statistics, 2018. *CA: A Cancer Journal for Clinicians*, 68(4), 284-296. <https://doi.org/10.3322/caac.21456>

Wang, B., Chao, S., & Guo, B. (2022). Integrated weighted gene co-expression network analysis reveals biomarkers associated with prognosis of high-grade serous ovarian cancer. *Journal of Clinical Laboratory Analysis*, 36(2), e24165. <https://doi.org/10.1002/jcla.24165>

Wang, L., Wang, X., Zhu, X., Zhong, L., Jiang, Q., Wang, Y., Tang, Q., Li, Q., Zhang, C., Wang, H., & Zou, D. (2024). Drug resistance in ovarian cancer: From mechanism to clinical trial. *Molecular Cancer*, 23(1), 66. <https://doi.org/10.1186/s12943-024-01967-3>

Wilczyński, J., Paradowska, E., Wilczyńska, J., & Wilczyński, M. (2023). Prediction of Chemoresistance—How Preclinical Data Could Help to Modify Therapeutic Strategy in High-Grade Serous Ovarian Cancer. *Current Oncology*, 31(1), 229-249. <https://doi.org/10.3390/curroncol31010015>

Wu, M., Sun, Y., Wu, J., & Liu, G. (2020). Identification of Hub Genes in High-Grade Serous Ovarian Cancer Using Weighted Gene Co-Expression Network Analysis. *Medical Science Monitor*, 26. <https://doi.org/10.12659/MSM.922107>

Wu, Y., Xia, L., Guo, Q., Zhu, J., Deng, Y., & Wu, X. (2020). Identification of Chemoresistance-Associated Key Genes and Pathways in High-Grade Serous Ovarian Cancer by Bioinformatics Analyses. *Cancer Management and Research*, Volume 12, 5213-5223. <https://doi.org/10.2147/CMAR.S251622>

Yin, F., Yi, S., Wei, L., Zhao, B., Li, J., Cai, X., Dong, C., & Liu, X. (2019). Microarray-based identification of genes associated with prognosis and drug resistance in ovarian cancer. *Journal of Cellular Biochemistry*, 120(4), 6057-6070. <https://doi.org/10.1002/jcb.27892>

Yuan, Q., Lv, N., Chen, Q., Shen, S., Wang, Y., & Tong, J. (2024). Application of single cell sequencing technology in ovarian cancer research (review). *Functional & Integrative Genomics*, 24(5), 144. <https://doi.org/10.1007/s10142-024-01432-w>

Zachariadis, V., Cheng, H., Andrews, N., & Enge, M. (2020). A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells. *Molecular Cell*, 80(3), 541-553.e5. <https://doi.org/10.1016/j.molcel.2020.09.025>

Zhan, L., Zhang, Y., Wang, W., Song, E., Fan, Y., Li, J., & Wei, B. (2016). Autophagy as an emerging therapy target for ovarian carcinoma. *Oncotarget*, 7(50), 83476-83487. <https://doi.org/10.18632/oncotarget.13080>

Zhang, K., Erkan, E. P., Jamalzadeh, S., Dai, J., Andersson, N., Kaipio, K., Lamminen, T., Mansuri, N., Huhtinen, K., Carpén, O., Hietanen, S., Oikkonen, J., Hynninen, J., Virtanen, A., Häkkinen, A., Hautaniemi, S., & Vähärautio, A. (2022). Longitudinal single-cell RNA-seq analysis reveals stress-promoted chemoresistance in metastatic ovarian cancer. *Science Advances*, 8(8), eabm1831. <https://doi.org/10.1126/sciadv.abm1831>

Zhou, J., Kang, Y., Chen, L., Wang, H., Liu, J., Zeng, S., & Yu, L. (2020). The Drug-Resistance Mechanisms of Five Platinum-Based Antitumor Agents. *Frontiers in Pharmacology*, 11, 343. <https://doi.org/10.3389/fphar.2020.00343>

9.2 Bases de datos, herramientas bioinformáticas y páginas web.

American Cancer Society (ACS). (2024). Ovarian cancer statistics. <https://www.cancer.org>

Google. (n.d.). Gemini. <https://gemini.google.com/app>

inferCNV of the Trinity CTAT Project. <https://github.com/broadinstitute/inferCNV>

International Agency for Research on Cancer (IARC). (2022). Global cancer statistics 2022. <https://www.iarc.who.int>

OpenAI. (2024). ChatGPT (Octubre 2024 versión). <https://chat.openai.com/>

Proyecto DECIDER: Innovative personalized treatments for ovarian cancer. Recuperado de <https://www.deciderproject.eu/>

Reid, C. M., Salomon, M. P., Miyara, S., Lu, Z., Tsai, S., Bilal, E., ... & McGregor, K. A. (2021). Longitudinal single-cell RNA-seq data of metastatic ovarian cancer. Gene Expression Omnibus (GEO), GSE165897. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165897>

Research Rabbit. (n.d.). Research Rabbit. <https://researchrabbitapp.com/>

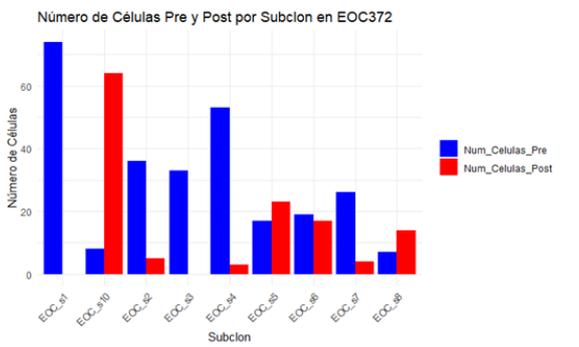
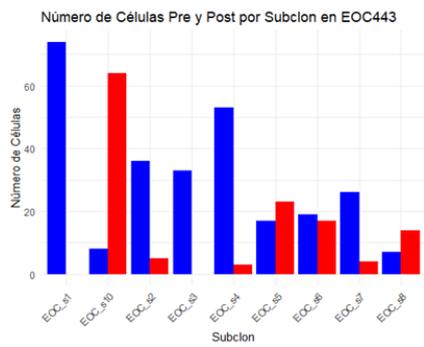
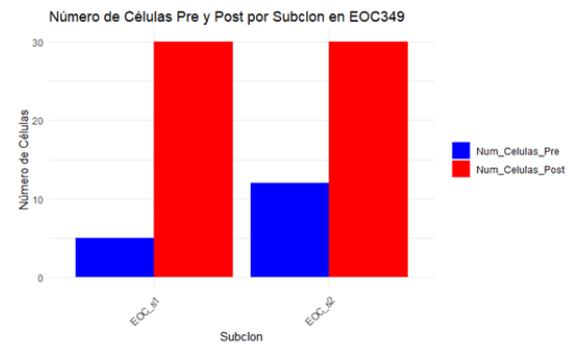
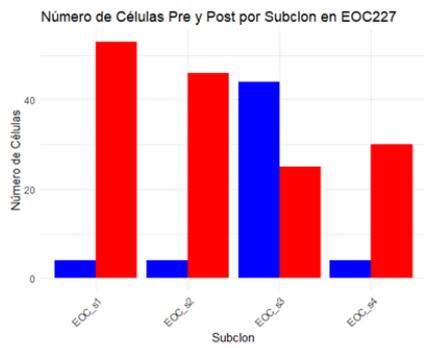
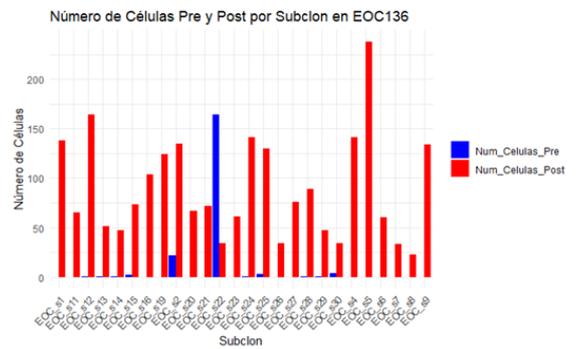
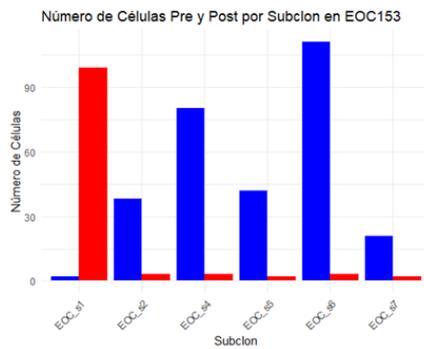
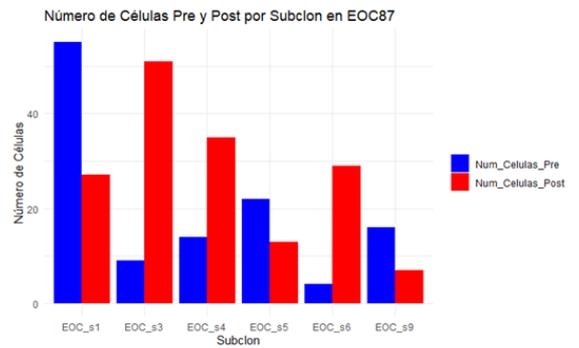
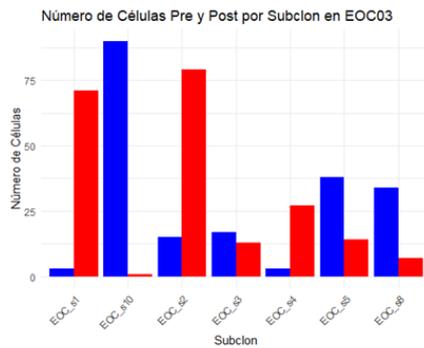
Satija Lab. (2023, October 31). Integrative analysis in Seurat v5. https://satijalab.org/seurat/articles/seurat5_integratation

SciSpace. (n.d.). SciSpace (Version 1.4.9). <https://typeset.io/>

Tickle, T., Ti, G. C., Brown, M., & Haas, B. (2019). inferCNV of the Trinity CTAT Project (Klarman Cell Observatory, Broad Institute of MIT and Harvard). Recuperado de <https://github.com/broadinstitute/inferCNV>

World Health Organization. (2018). Cancer. <https://www.who.int/news-room/fact-sheets/detail/cancer>

10. ANEXOS



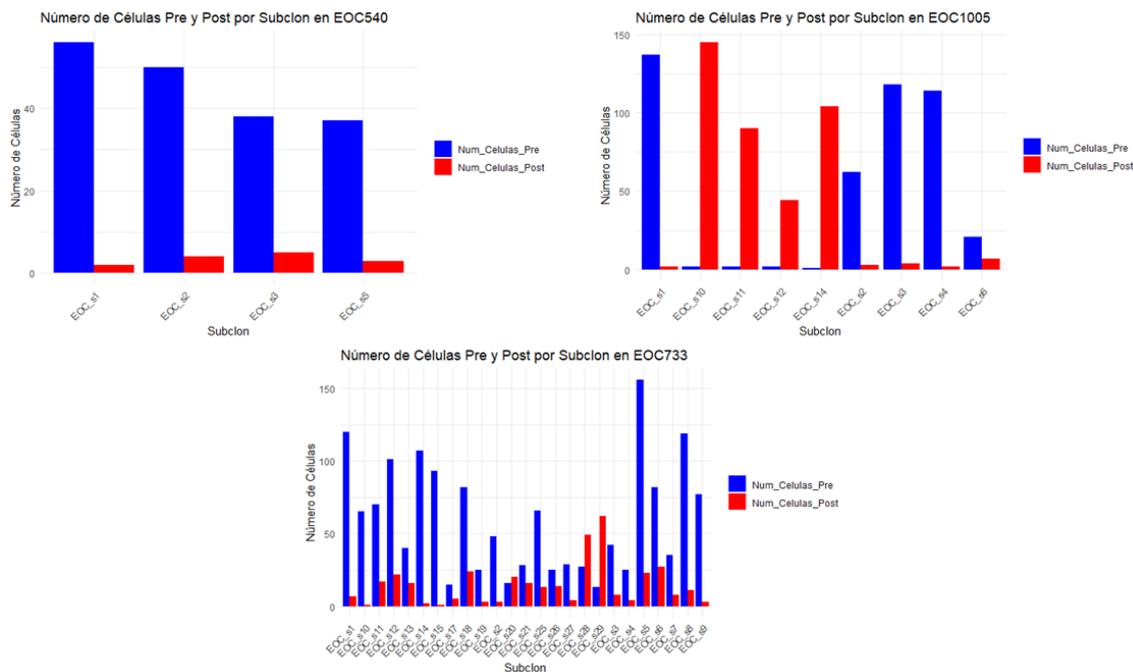


Figura S 1. Distribución de células pre y post tratamiento por subclón en los 11 pacientes. Los gráficos muestran en azul el número de células pre-tratamiento y en rojo post-tratamiento. En el eje X se pueden observar los subclones de cada paciente.

inferCNV

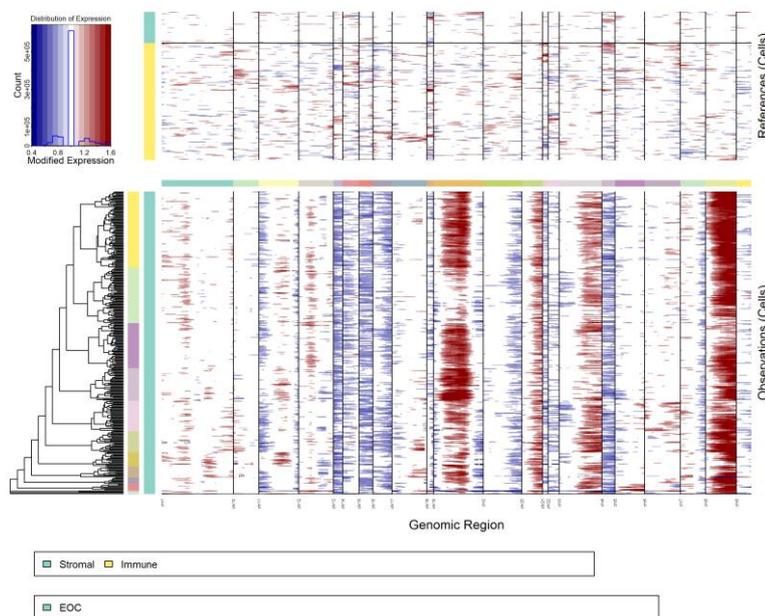


Figura S 2. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 87.

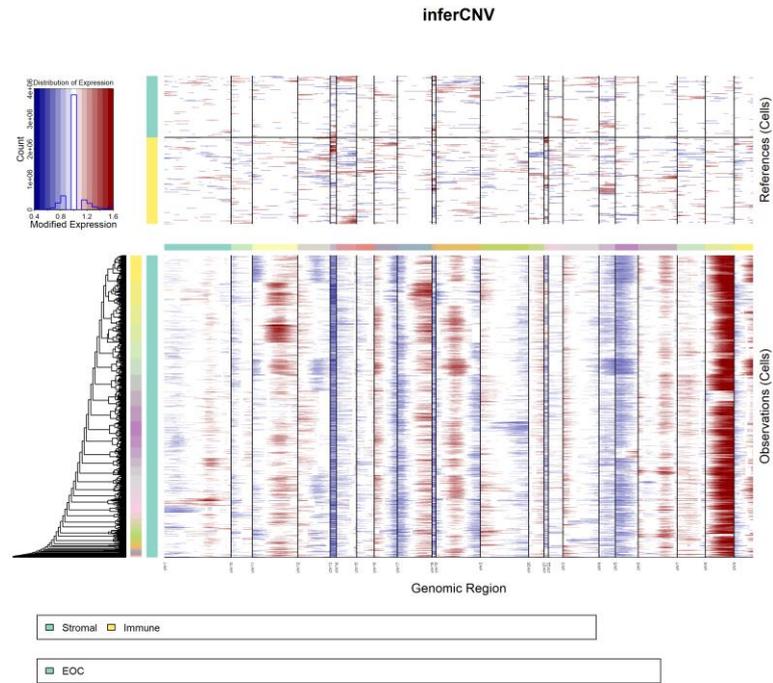


Figura S 3. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 136.

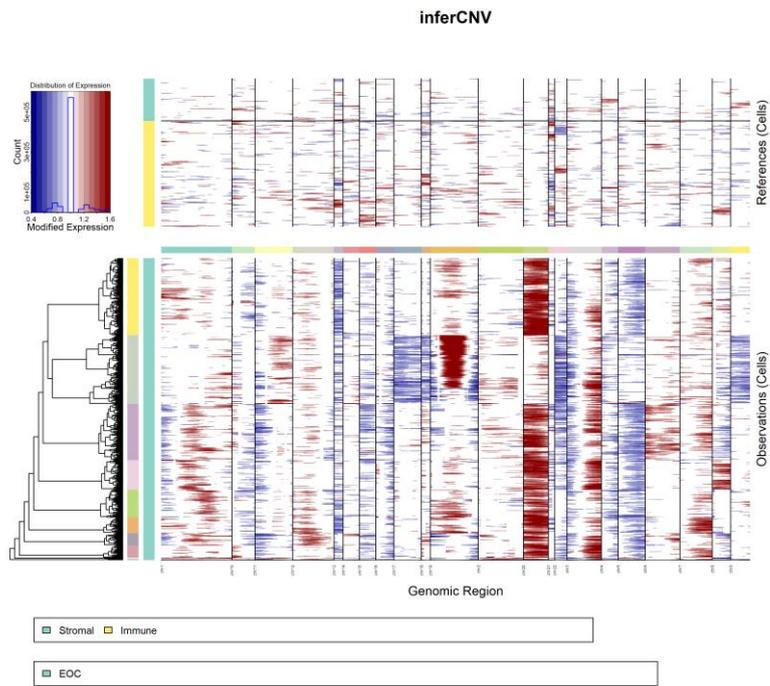


Figura S 4. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 153.

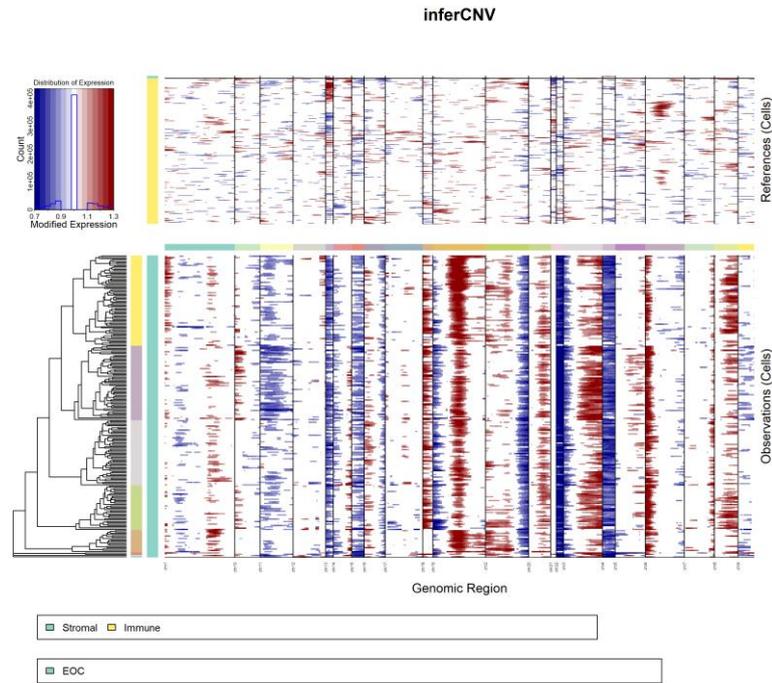


Figura S 5. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 227.

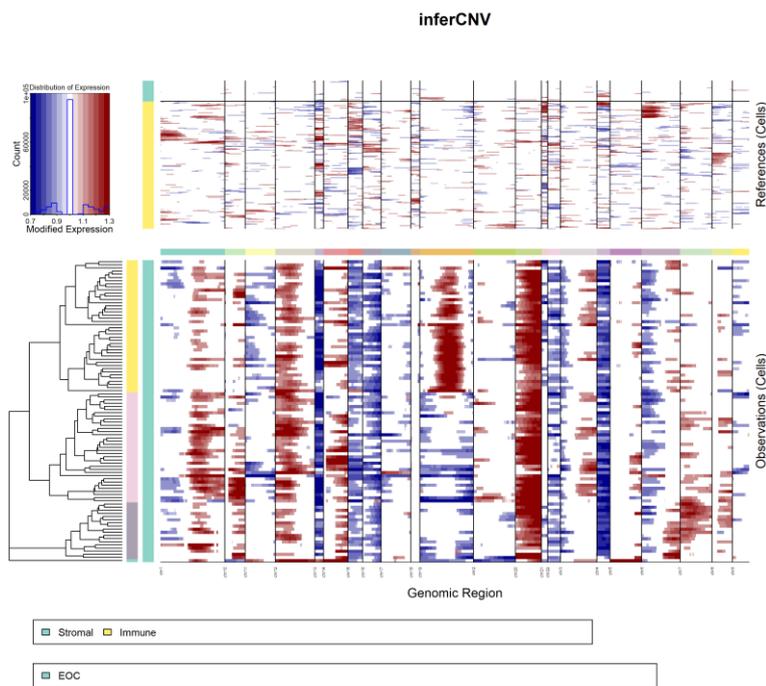


Figura S 6. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 349.

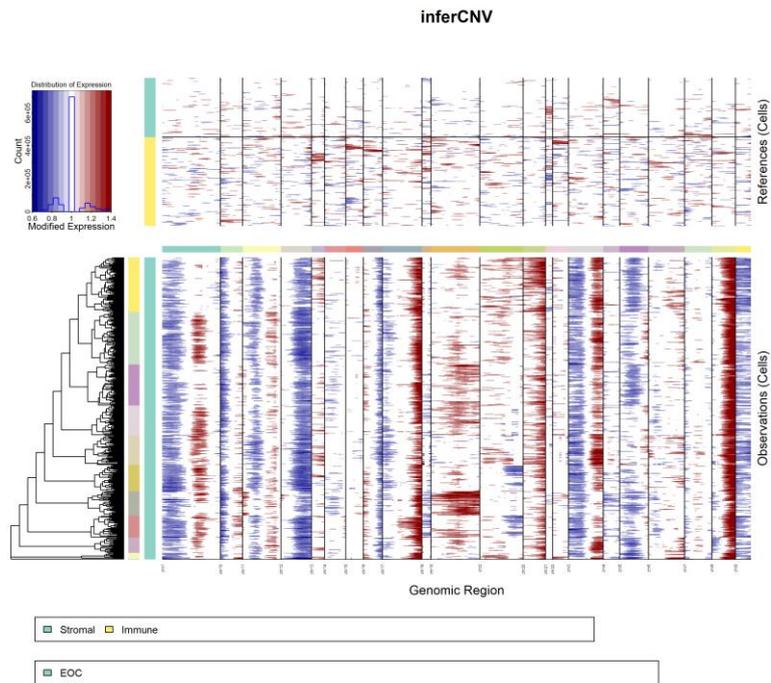


Figura S 7. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 372.

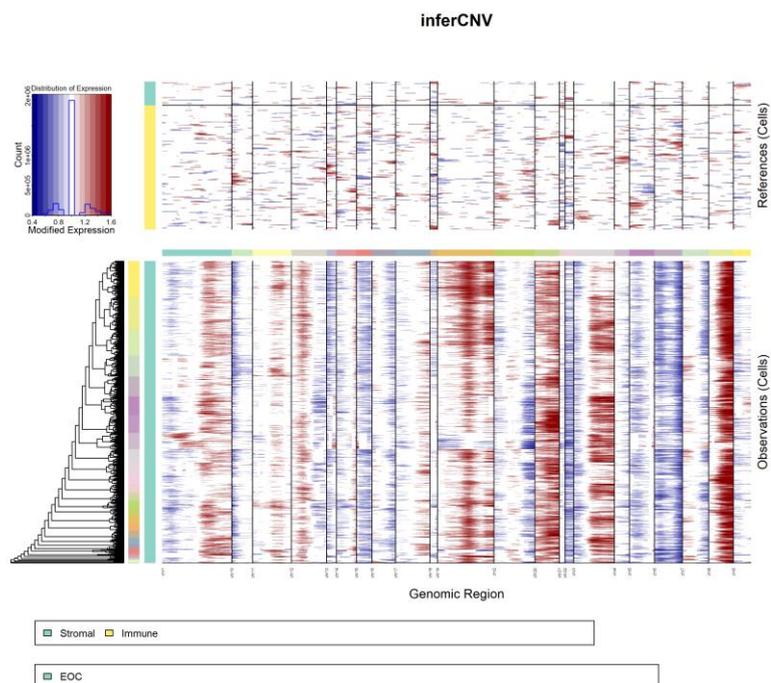


Figura S 8. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 443.

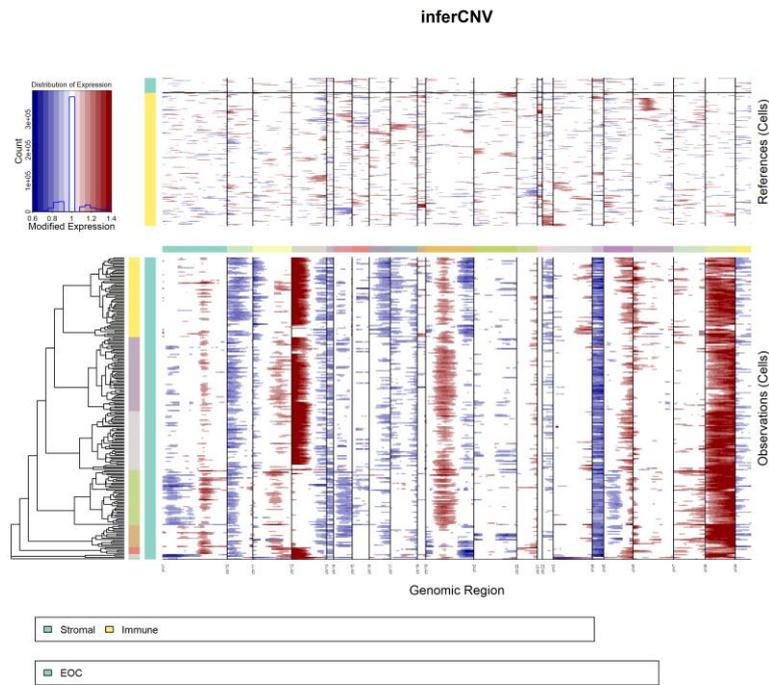


Figura S 9. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 540.

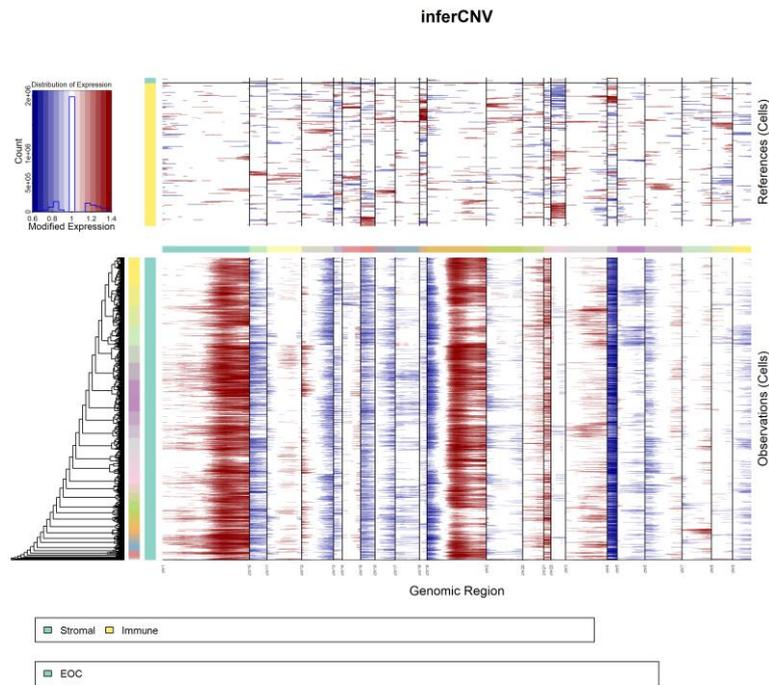


Figura S 10. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 733.

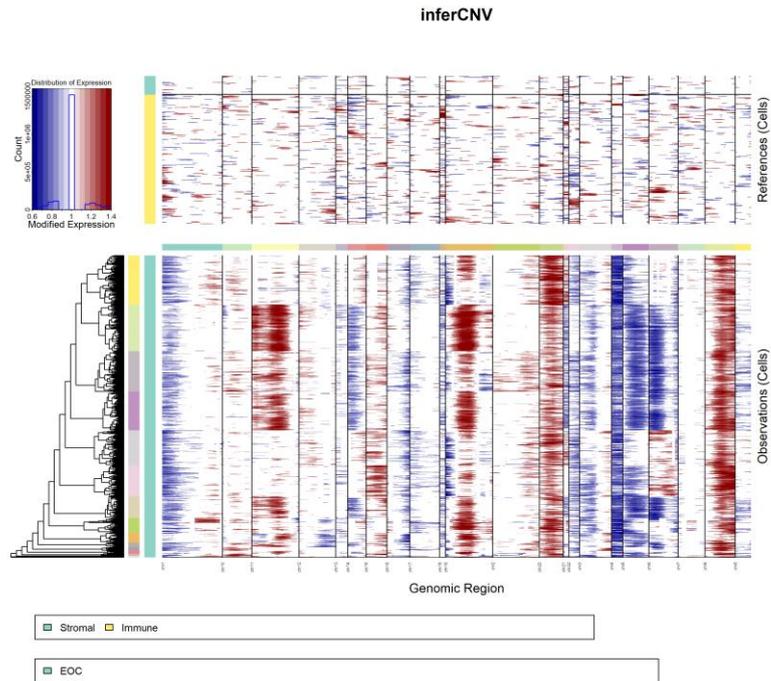


Figura S 11. Mapa de calor generado por inferCNV que muestra las ganancias (rojo) y las pérdidas (azul) en células de referencia (parte superior) y células cancerosas (parte inferior). Paciente 1005.

Tabla S 1. Genes significativos con p-valor, avg_log2FC y p-valor ajustado.

| Gene | p_val | avg_log2FC | p_val_adj |
|----------|----------|--------------|------------|
| SERPINA1 | 4,68E-09 | 1,958293806 | 7,6274E-05 |
| RPS19 | 9,78E-08 | -0,83357374 | 0,00159449 |
| FXD3 | 2,4E-07 | 2,288105694 | 0,00391029 |
| RPL6 | 4,02E-07 | -0,835131212 | 0,0065521 |
| SAT1 | 4,31E-07 | 1,568409915 | 0,0070195 |
| ZBTB16 | 4,45E-07 | 2,891798363 | 0,00725968 |
| BAZ1A | 7,14E-07 | 1,204393007 | 0,01163867 |
| HLA-DRB5 | 8,05E-07 | 1,35525527 | 0,01311812 |
| RPSA | 1,09E-06 | -0,947806018 | 0,01770213 |
| CPEB4 | 1,42E-06 | 2,807641161 | 0,02315421 |
| RPL37A | 1,84E-06 | -0,667352302 | 0,02998499 |

| Gene | p_val | avg_log2FC | p_val_adj |
|-----------|----------|--------------|------------|
| RPS26 | 2,19E-06 | -0,440503269 | 0,03565894 |
| S100A4 | 2,21E-06 | 1,809043514 | 0,03607 |
| RPL19 | 2,28E-06 | -0,462829803 | 0,03723097 |
| OPTN | 2,31E-06 | 1,478767018 | 0,03771377 |
| FAU | 2,33E-06 | -0,420056822 | 0,03804172 |
| RPS16 | 2,54E-06 | -0,494466202 | 0,04145753 |
| RPS27A | 2,89E-06 | -0,54878385 | 0,0471387 |
| RPL14 | 2,89E-06 | -0,483507725 | 0,0471387 |
| ABL24 | 2,24E-08 | 1,68627094 | 0,00036451 |
| NR3C14 | 4,49E-08 | 1,857181775 | 0,00073217 |
| MARCKSL14 | 2,29E-07 | -1,436778597 | 0,00372622 |
| WFDC23 | 6,98E-07 | -1,378723892 | 0,01137717 |
| MIF3 | 1,34E-06 | -0,774443887 | 0,02187448 |
| MDK3 | 1,52E-06 | -1,786904538 | 0,02474894 |
| CLDN64 | 2,21E-06 | -1,599310658 | 0,03596517 |
| FOSL23 | 2,99E-06 | 1,287150965 | 0,04872453 |
| GTSF11 | 2,02E-28 | 6,518555215 | 3,2971E-24 |
| RPL18A3 | 5,92E-17 | -1,239888769 | 9,651E-13 |
| RPS285 | 1,72E-15 | -1,133137468 | 2,802E-11 |
| RPS164 | 1,19E-14 | -0,953003998 | 1,9415E-10 |
| RPS195 | 2,64E-14 | -1,058714858 | 4,3065E-10 |
| RPLP24 | 9,91E-14 | -0,941065564 | 1,6156E-09 |
| UBA523 | 2,62E-13 | -0,867423957 | 4,2719E-09 |
| ATP6V0B3 | 5,08E-13 | 1,590629618 | 8,286E-09 |
| RPL214 | 2,71E-12 | -1,087174709 | 4,4105E-08 |
| RPL27A5 | 8,18E-12 | -0,786281193 | 1,3326E-07 |

| Gene | p_val | avg_log2FC | p_val_adj |
|-------------|----------|--------------|------------|
| RPS274 | 1,49E-11 | -0,83415571 | 2,4248E-07 |
| UQCRH3 | 1,53E-10 | 5,977208274 | 2,4862E-06 |
| RPLP14 | 1,92E-10 | -0,70246887 | 3,1337E-06 |
| TECR4 | 1,93E-10 | -1,836058339 | 3,1433E-06 |
| EEF1D4 | 2,04E-10 | -0,84563309 | 3,3248E-06 |
| RPL13A5 | 2,62E-10 | -0,756881757 | 4,2719E-06 |
| NDUFB74 | 3,36E-10 | -1,136246088 | 5,4809E-06 |
| JUN5 | 4,09E-10 | 1,684186307 | 6,6706E-06 |
| EIF3K3 | 4,5E-10 | -1,017014453 | 7,3343E-06 |
| EEF1B25 | 6,74E-10 | -0,953972552 | 1,099E-05 |
| COX6B14 | 8,85E-10 | -0,836792322 | 1,4429E-05 |
| RPSA4 | 1,08E-09 | -0,903105845 | 1,7613E-05 |
| MT1F3 | 1,29E-09 | 2,536165111 | 2,105E-05 |
| ALKBH72 | 2,09E-09 | -1,365372547 | 3,4019E-05 |
| TMEM595 | 2,36E-09 | 2,050867574 | 3,8532E-05 |
| RPL84 | 2,69E-09 | -0,587015985 | 4,388E-05 |
| GADD45GIP14 | 3,09E-09 | -1,153702486 | 5,0377E-05 |
| MT2A5 | 3,12E-09 | 2,250573348 | 5,0917E-05 |
| RPS55 | 3,22E-09 | -0,745367752 | 5,2535E-05 |
| RPS95 | 3,36E-09 | -0,567123817 | 5,4757E-05 |
| CDC424 | 3,49E-09 | 1,287724091 | 5,6945E-05 |
| RPL283 | 4,3E-09 | -0,591994649 | 7,0138E-05 |
| SH3BGRL35 | 5,14E-09 | 1,240468441 | 8,3784E-05 |
| RPS115 | 1,01E-08 | -0,683792919 | 0,0001647 |
| BNC1 | 1,02E-08 | 4,206978651 | 0,00016594 |
| MEAF63 | 1,47E-08 | 1,49817651 | 0,00023893 |

| Gene | p_val | avg_log2FC | p_val_adj |
|----------|----------|--------------|------------|
| KLF65 | 2,31E-08 | 2,294731114 | 0,00037684 |
| CAPZB3 | 2,72E-08 | 1,266400853 | 0,00044304 |
| PGAM2 | 2,72E-08 | -2,217975757 | 0,00044336 |
| RPS34 | 2,72E-08 | -0,720817257 | 0,00044336 |
| NPW3 | 2,73E-08 | -3,010029735 | 0,00044449 |
| IGF2BP25 | 2,78E-08 | 1,005078468 | 0,00045296 |
| DNAJC83 | 3,07E-08 | 1,694273193 | 0,00049956 |
| ITM2B5 | 3,47E-08 | 1,725476214 | 0,00056502 |
| PERP5 | 3,64E-08 | 2,312229683 | 0,00059338 |
| TM2D15 | 4,79E-08 | 1,774880179 | 0,00078053 |
| RPL7A4 | 5,18E-08 | -0,690002599 | 0,0008446 |
| CAMTA14 | 6,26E-08 | 1,119498819 | 0,00102071 |
| RPS155 | 8,6E-08 | -0,721338159 | 0,00140165 |
| MAP7D15 | 8,72E-08 | 1,375622799 | 0,00142119 |
| RPL23A4 | 1,12E-07 | -0,579137015 | 0,00182281 |
| ELF15 | 1,16E-07 | 2,32087346 | 0,00188957 |
| TSC22D34 | 1,27E-07 | 2,711690555 | 0,00207229 |
| MRPS124 | 1,28E-07 | -1,130743975 | 0,00208244 |
| EIF4G33 | 1,31E-07 | 2,425995484 | 0,00213697 |
| RSPRY14 | 1,34E-07 | 2,919012561 | 0,00218835 |
| RHEB4 | 1,34E-07 | 1,116637263 | 0,00218962 |
| RPLP05 | 1,62E-07 | -0,723750436 | 0,0026429 |
| RPL364 | 2,15E-07 | -0,728119417 | 0,00350306 |
| RPS204 | 2,43E-07 | -0,554543778 | 0,00395633 |
| MT1E4 | 2,45E-07 | 3,567579963 | 0,00399904 |
| RPS216 | 2,81E-07 | -0,674630939 | 0,00457529 |

| Gene | p_val | avg_log2FC | p_val_adj |
|-----------|----------|--------------|------------|
| HP1BP34 | 2,83E-07 | 1,4664193 | 0,00461336 |
| RPL315 | 2,91E-07 | -0,651719488 | 0,00474408 |
| ATP6V1D4 | 3E-07 | 1,791989171 | 0,00489239 |
| RPL385 | 3,4E-07 | -0,552479115 | 0,00554756 |
| C19orf534 | 3,61E-07 | -0,714861808 | 0,00589025 |
| FBL4 | 4,12E-07 | -0,901377385 | 0,00671075 |
| DBI5 | 4,34E-07 | -0,965656409 | 0,00707843 |
| CLSTN15 | 5,11E-07 | 1,842072476 | 0,00832803 |
| AHDC14 | 5,69E-07 | 2,247149166 | 0,00927502 |
| OSCP11 | 6E-07 | 3,01445822 | 0,0097827 |
| YTHDF24 | 6,43E-07 | 1,075321476 | 0,01048575 |
| AKIRIN14 | 7,73E-07 | 1,326177141 | 0,01260082 |
| RAB7A3 | 7,92E-07 | 0,960206997 | 0,01290404 |
| CDKN1A4 | 8,43E-07 | 1,24361988 | 0,01374137 |
| RPL353 | 9,27E-07 | -0,500734554 | 0,01511304 |
| RPL103 | 9,71E-07 | -0,622432924 | 0,01582881 |
| SNRPD24 | 9,94E-07 | -0,714882903 | 0,01619883 |
| TGIF15 | 1,01E-06 | 1,742058732 | 0,0164036 |
| CAB39L | 1,05E-06 | 2,651398768 | 0,01718113 |
| LUZP13 | 1,07E-06 | 1,470357963 | 0,0175162 |
| LDLR4 | 1,08E-06 | -1,722349026 | 0,01758606 |
| TMOD34 | 1,12E-06 | 1,341657483 | 0,01818889 |
| FABP63 | 1,38E-06 | -1,511639025 | 0,02255437 |
| NTHL14 | 1,38E-06 | -2,086925968 | 0,02255672 |
| RPL135 | 1,48E-06 | -0,569381907 | 0,02419349 |
| PRDX15 | 1,71E-06 | 1,143717192 | 0,02787282 |

| Gene | p_val | avg_log2FC | p_val_adj |
|----------|----------|--------------|------------|
| RPL124 | 1,97E-06 | -0,572070075 | 0,03210072 |
| RPS245 | 1,97E-06 | -0,597906386 | 0,03210072 |
| NENF4 | 2,09E-06 | -1,062571222 | 0,03408371 |
| TPM44 | 2,13E-06 | -0,993706352 | 0,03468625 |
| RPL295 | 2,25E-06 | -0,623686206 | 0,03672708 |
| ARID5B5 | 2,27E-06 | 3,187936197 | 0,03701931 |
| CAP14 | 2,31E-06 | 1,261064247 | 0,03767491 |
| MT1M2 | 2,33E-06 | 3,87272433 | 0,03798355 |
| RSRP14 | 2,34E-06 | 2,140299773 | 0,03808328 |
| CDC42EP5 | 2,56E-06 | -2,879539858 | 0,04178746 |
| PKN13 | 2,58E-06 | -1,488976533 | 0,0420635 |
| C34 | 2,59E-06 | -1,943491489 | 0,04222564 |
| PNISR5 | 2,64E-06 | 1,06242151 | 0,04306902 |
| AHCY5 | 2,78E-06 | -0,980948344 | 0,04527733 |
| SCP24 | 3,02E-06 | 1,408429834 | 0,04926671 |
| ZNF804A | 6,94E-09 | 6,422336075 | 0,00011313 |
| RASSF92 | 1,61E-08 | 4,762332703 | 0,0002625 |
| NUDT112 | 8,05E-08 | 5,365211537 | 0,00131154 |
| NNMT4 | 2,53E-06 | 4,386219041 | 0,04119307 |