

# Análisis computacional de datos de scRNA-seq en cáncer de próstata: integración, caracterización y comparación de poblaciones celulares en tejido sano, PCa y CRPC.



**Silvia Arroitajauregui Avilés.**

MU Bioinf. y Bioest.

Análisis de datos Ómicos.

**Nombre Tutora de TF:**

Dra. Anna Esteve Codina.

**Profesor responsable de la asignatura:**

Dr. Ivan Pisa Dacosta.



UNIVERSITAT DE  
BARCELONA

**01/2026.**



Esta obra está sujeta a una licencia de  
Reconocimiento-NoComercial-SinObraDerivada [3.0](#)  
[España de Creative Commons](#)

**FICHA DEL TRABAJO FINAL**

|                                    |  |
|------------------------------------|--|
| <b>Título del trabajo:</b>         | <i>Análisis computacional de datos de scRNA-seq en cáncer de próstata: integración, caracterización y comparación de poblaciones celulares en tejido sano, PCa y CRPC.</i> |
| <b>Nombre del autor:</b>           | <i>Silvia Arroitajauregui Avilés.</i>  |
| <b>Nombre del director/a:</b>      | <i>Dra. Anna Esteve Codina.</i>  |
| <b>Nombre del PRA:</b>             | <i>Dr. Ivan Pisa Dacosta.</i>  |
| <b>Fecha de entrega (mm/aaaa):</b> | <i>01/2026</i>   |
| <b>Titulación o programa:</b>      | Máster Universitario en Bioinformática y Bioestadística.   |
| <b>Área del Trabajo Final:</b>     | <i>Análisis de datos Ómicos.</i>   |
| <b>Idioma del trabajo:</b>         | <i>Castellano.</i>   |
| Palabras clave                     | <i>scRNA-seq, cáncer de próstata, microambiente tumoral.</i>   |

**Resumen del Trabajo**

Este trabajo presenta un análisis computacional de datos de transcriptómica unicelular (scRNA-seq) aplicado al estudio del cáncer de próstata. Se ha llevado a cabo una integración de datos públicos de tejido prostático en tres condiciones: tejido sano, cáncer de próstata localizado (PCa) y cáncer de próstata resistente a la castración (CRPC), con el objetivo de caracterizar y comparar la composición celular y los perfiles de expresión génica en cada estado. Mediante técnicas de normalización, reducción de dimensiones, anotación celular y análisis de expresión diferencial, se identificaron cambios progresivos en múltiples compartimentos del microambiente tumoral.

Los resultados muestran una sustitución de poblaciones epiteliales benignas por subtipos tumorales específicos, así como una expansión de células estromales y proliferativas en CRPC. Se observó una pérdida sistemática de genes mitocondriales y de mantenimiento homeostático, junto con una sobreexpresión de genes asociados a estrés, plasticidad y evasión inmunitaria. En el compartimento inmune, tanto macrófagos como células T mostraron una pérdida de funciones efectoras en CRPC, sugiriendo un microambiente inmunosupresor. Estos hallazgos refuerzan el modelo de progresión prostática basado en la desdiferenciación celular y la remodelación del entorno tumoral.

El trabajo contribuye a una comprensión más detallada de la evolución celular y molecular en el cáncer de próstata, y sienta las bases para la identificación de biomarcadores o dianas terapéuticas en estadios avanzados de la enfermedad.

## **Abstract**

This work presents a computational analysis of single-cell RNA sequencing (scRNA-seq) data applied to the study of prostate cancer. Publicly available datasets of prostate tissue under three biological conditions—healthy, localized prostate cancer (PCa), and castration-resistant prostate cancer (CRPC)—were integrated with the goal of characterizing and comparing the cellular composition and gene expression profiles across disease stages. Using a standardized pipeline for data integration, normalization, dimensionality reduction, cell-type annotation, and differential expression analysis, progressive changes were identified across multiple components of the tumor microenvironment.

The results reveal a gradual replacement of benign epithelial populations by tumor-specific subtypes, as well as an expansion of stromal and proliferative cells in CRPC. A consistent loss of mitochondrial and homeostatic gene expression was observed, accompanied by increased expression of genes linked to stress response, cellular plasticity, and immune evasion. Immune cell compartments, including macrophages and T cells, showed reduced effector functions in CRPC, suggesting the establishment of an immunosuppressive environment. These findings support a model of prostate cancer progression involving dedifferentiation and microenvironmental remodeling.

This study enhances our understanding of cellular and molecular evolution in prostate cancer and provides a foundation for the future identification of stage-specific biomarkers or therapeutic targets in advanced disease.

# Índice

|        |   |    |
|--------|---|----|
| 1.     | Introducción.....   | 8  |
| 1.1.   | Contexto y justificación del Trabajo.....                     | 8  |
| 1.2.   | Objetivos del trabajo. ....                                   | 9  |
| 1.2.1. | Objetivos generales.....                                      | 9  |
| 1.2.2. | Objetivos específicos. ....                                   | 9  |
| 1.3.   | Impacto en sostenibilidad, ético-social y de diversidad. .... | 10 |
| 1.4.   | Enfoque y método seguido.....                                 | 10 |
| 1.5.   | Planificación del Trabajo. ....                               | 11 |
| 1.5.1. | Tareas.....   | 11 |
| 1.5.2. | Calendario. ....  | 12 |
| 1.5.3. | Hitos. ....   | 13 |
| 1.5.4. | Análisis de riesgos. ....                                     | 14 |
| 1.6.   | Breve sumario de productos obtenidos. ....                    | 14 |
| 1.7.   | Breve descripción de los otros capítulos de la memoria.....   | 15 |
| 2.     | Estado del arte. ....   | 16 |
| 2.1.   | Introducción al cáncer de próstata y contexto clínico. ....   | 16 |
| 2.2.   | Tecnologías de secuenciación a célula única (scRNA-seq). .... | 17 |
| 2.2.1. | Fundamentos y evolución tecnológica. ....                     | 17 |
| 2.2.2. | Procesamiento y retos analíticos. ....                        | 18 |
| 2.2.3. | Aplicaciones en cáncer y microambiente tumoral.....           | 18 |
| 2.3.   | Ánalisis computacional en scRNA-seq.....                      | 19 |
| 2.3.1. | Control de calidad y filtrado. ....                           | 19 |
| 2.3.2. | Normalización y selección de genes variables. ....            | 19 |
| 2.3.3. | Reducción dimensional y corrección de lotes. ....             | 19 |
| 2.3.4. | Construcción de grafos y clustering.....                      | 20 |
| 2.3.5. | Anotación de tipos celulares. ....                            | 20 |
| 2.3.6. | Visualización: UMAP y t-SNE. ....                             | 20 |
| 2.3.7. | Ánalisis diferencial y caracterización funcional.....         | 20 |
| 2.4.   | Aplicaciones del scRNA-seq en cáncer de próstata. ....        | 21 |
| 2.4.1. | Heterogeneidad del epitelio prostático.....                   | 21 |
| 2.4.2. | Microambiente tumoral y estroma en cáncer de próstata.....    | 21 |

|        |  |    |
|--------|--|----|
| 2.5.   | Anotación de tipos celulares.....  | 23 |
| 2.5.1. | Métodos basados en marcadores clásicos.....  | 23 |
| 2.5.2. | Métodos automatizados basados en referencia.....                                       | 24 |
| 2.5.3. | Atlas celulares como referencia.....   | 24 |
| 2.5.4. | Necesidad de curación manual en cáncer de próstata.....                                | 24 |
| 2.6.   | Avances recientes en scRNA-seq aplicado a cáncer de próstata.....                      | 25 |
| 2.6.1. | Redefinición del atlas celular prostático en tejido sano y tumor. ....                 | 25 |
| 2.6.2. | Identificación de subpoblaciones luminales asociadas a agresividad tumoral. 25         |    |
| 2.6.3. | Plasticidad celular y transición hacia fenotipos neuroendocrinos (NEPC).<br>26         |    |
| 2.6.4. | Redes de comunicación célula-célula en CRPC.....                                       | 26 |
| 2.6.5. | Aplicaciones clínicas emergentes: biomarcadores y medicina de precisión. 27            |    |
| 2.6.6. | Síntesis y relación con el presente proyecto. ....                                     | 27 |
| 3.     | Materiales y métodos.....  | 28 |
| 3.1.   | Obtención y caracterización de los datos.....  | 28 |
| 3.2.   | Procesamiento y control de calidad. ....   | 28 |
| 3.3.   | Integración de datasets, reducción dimensional y clustering. ....                      | 29 |
| 3.4.   | Identificación de genes característicos por clúster (TOP markers).....                 | 30 |
| 3.5.   | Anotación automática de poblaciones celulares. ....                                    | 31 |
| 3.6.   | Estudio de poblaciones celulares y composición por condición biológica. ....           | 32 |
| 3.7.   | Anotación manual de poblaciones celulares. ....  | 33 |
| 3.8.   | Estudio de poblaciones y composición celular por condición biológica.....              | 34 |
| 3.9.   | Análisis de expresión diferencial por tipo celular entre condiciones biológicas.<br>35 |    |
| 3.10.  | Selección de poblaciones celulares para el análisis comparativo. ....                  | 36 |
| 3.11.  | Visualización mediante volcano plots. ....   | 37 |
| 3.12.  | Visualización de genes diferencialmente expresados mediante heatmaps. ....             | 38 |
| 3.13.  | Organización y exportación de resultados de expresión diferencial. ....                | 39 |
| 4.     | Resultados. ....   | 40 |
| 4.1.   | Control de calidad (QC). ....  | 40 |
| 4.2.   | Integración y agrupamiento celular. ....   | 40 |
| 4.3.   | Identificación de genes característicos por clúster. ....                              | 43 |
| 4.4.   | Anotación automática de poblaciones celulares. ....                                    | 46 |

|        |  |    |
|--------|--|----|
| 4.5.   | Composición celular por condición biológica de los clústeres. .... | 49 |
| 4.6.   | Composición de poblaciones celulares anotadas manualmente. ....    | 49 |
| 4.7.   | Análisis de expresión diferencial por tipo celular. ....           | 54 |
| 4.7.1. | Células epiteliales luminales benignas.....                        | 55 |
| 4.7.2. | Células endoteliales.....  | 58 |
| 4.7.3. | Epiteliales luminales metabólicas.....                             | 61 |
| 4.7.4. | Macrófagos. ....   | 64 |
| 4.7.5. | Células T. ....  | 67 |
| 5.     | Discusión.....   | 70 |
| 5.1.   | Composición celular por condición. ....                            | 70 |
| 5.2.   | Expresión diferencial. ....  | 71 |
| 5.2.1. | Células epiteliales luminales benignas.....                        | 71 |
| 5.2.2. | Células endoteliales. ....   | 72 |
| 5.2.3. | Células epiteliales luminales metabólicas. ....                    | 73 |
| 5.2.4. | Macrófagos. ....   | 74 |
| 5.2.5. | Células T. ....  | 75 |
| 6.     | Conclusiones y trabajos futuros.....                               | 76 |
| 6.1.   | Limitaciones y perspectivas futuras. ....                          | 76 |
| 7.     | Glosario. ....   | 78 |
| 8.     | Bibliografía. ....   | 79 |
| 9.     | Anexos.....  | 90 |

# Lista de figuras

|  |    |
|--|----|
| Figura 1: Diagrama de Gantt del proyecto.....  | 12 |
| Figura 2: Varianza acumulada explicada. ....   | 41 |
| Figura 3: Elbow plot.....  | 41 |
| Figura 4: Representaciones UMAP tras la integración de datos.....  | 42 |
| Figura 5: Distribución de los clústeres según condición.....   | 42 |
| Figura 6: Distribución por muestra individual. ....  | 43 |
| Figura 7: Heatmap de los 10 genes más representativos del clúster 1 al 5. ....   | 44 |
| Figura 8: Heatmap de los clústeres 6 al 10.....  | 44 |
| Figura 9: Heatmap de los clústeres 11 al 15. ....  | 45 |
| Figura 10: Heatmap de los clústeres 16 al 18.....  | 45 |
| Figura 11: UMAP global con anotación automática de poblaciones celulares. ....   | 46 |
| Figura 12: UMAP anotado - Condición Healthy. ....  | 47 |
| Figura 13: UMAP anotado - Condición PCa.....   | 48 |
| Figura 14: UMAP anotado - Condición CRPC. ....   | 48 |
| Figura 15: Composición relativa de cada clúster según la condición biológica. ....   | 49 |
| Figura 16: UMAP global con anotación manual de tipos celulares. ....   | 50 |
| Figura 17: UMAP para Healthy.....  | 51 |
| Figura 18: UMAP para PCa. ....   | 51 |
| Figura 19: UMAP para CRPC. ....  | 52 |
| Figura 20: Métricas de calidad por poblaciones celulares. ....   | 52 |
| Figura 21: Composición celular por condición. ....   | 53 |
| Figura 22: Distribución por condición dentro de cada población. ....   | 54 |
| Figura 23: Número de células por tipo celular y condición y selección de poblaciones con representación suficiente en las tres condiciones. .... | 55 |
| Figura 24: Volcano plot de Benign luminal epithelial para Healthy vs. PCa. ....  | 55 |
| Figura 25: Heatmap de Benign luminal epithelial para Healthy vs. PCa. ....   | 55 |
| Figura 27: Heatmap de Benign luminal epithelial para Healthy vs. CRPC. ....  | 56 |
| Figura 26: Volcano plot de Benign luminal epithelial para Healthy vs. CRPC. ....   | 56 |
| Figura 28: Volcano plot de Benign Luminal epithelial para PCa vs. CRPC. ....   | 57 |
| Figura 29: Heatmap de Benign luminal epithelial para PCa vs. CRPC. ....  | 57 |
| Figura 30: Volcano plot de Endothelial cells para Healthy vs. PCa. ....  | 58 |
| Figura 31: Heatmap de Endothelial cells para Healthy vs. PCa. ....   | 58 |
| Figura 33: Heatmap de Endothelial cells para Healthy vs. CRPC. ....  | 59 |
| Figura 32: Volcano plot de Endothelial cells para Healthy vs. CRPC. ....   | 59 |
| Figura 35: Heatmap de Endothelial cells para PCa vs. CRPC. ....  | 60 |
| Figura 34: Volcano plot de Endothelial cells para PCa vs. CRPC. ....   | 60 |
| Figura 36: Volcano plot de Luminal methabolic epithelial para Healthy vs. PCa. ....  | 61 |
| Figura 37: Heatmap de Luminal methabolic epithelial para Healthy vs. PCa. ....   | 61 |
| Figura 38: Volcano plot de Luminal metabolic epithelial para Healthy vs. CRPC. ....  | 62 |
| Figura 39: Heatmap de Luminal metabolic epithelial para Healthy vs. CRPC. ....   | 62 |
| Figura 40: Volcano plot de Luminal metabolic epithelial para PCa vs CRPC. ....   | 63 |
| Figura 41: Heatmap de Luminal metabolic epithelial para PCa vs. CRPC. ....   | 63 |
| Figura 42: Volcano plot de Macrófagos para Healthy vs. PCa. ....   | 64 |
| Figura 43: Heatmap de Macrófagos para Healthy vs. PCa. ....  | 64 |

|   |    |
|---|----|
| Figura 44: Volcan plot de Macrófagos para Healthy vs. CRPC..... | 65 |
| Figura 45: Heatmap de Macrófagos para Healthy vs. CRPC.....     | 65 |
| Figura 46: Volcano plot de Macrófagos para PCa vs. CRPC.....    | 66 |
| Figura 47: Heatmap de Macrófagos para PCa vs. CRPC. ....        | 66 |
| Figura 48: Volcano plot de Células T para Healthy vs. PCa. .... | 67 |
| Figura 49: Heatmap de Células T para Healthy vs. PCa. ....      | 67 |
| Figura 50: Volcano plot de Células T para Healthy vs. CRPC..... | 68 |
| Figura 51: Heatmap de Células T para Healthy vs. CRPC.....      | 68 |
| Figura 52: Volcano plot de Células T para PCa vs. CRPC. ....    | 69 |
| Figura 53: Heatmap de Células T para PCa vs. CRPC. ....         | 69 |

# 1. Introducción.

## 1.1. Contexto y justificación del Trabajo.

El cáncer de próstata es el segundo cáncer más frecuentemente diagnosticado en los hombres (Rawla, 2019), y el cuarto más diagnosticado a nivel mundial, representando aproximadamente el 7,3% de casos de cáncer en el año 2022 (Bray Bsc et al., 2024). Además, constituye la quinta causa de muerte por cáncer en el mundo (Rawla, 2019).

El crecimiento y mantenimiento del epitelio prostático dependen de niveles fisiológicos de andrógenos, que estimulan la proliferación e inhiben la apoptosis celular. En consecuencia, las células cancerosas de próstata presentan una fuerte dependencia de los andrógenos, por lo que la terapia de privación androgénica (ADT) ha sido el tratamiento principal para los pacientes con enfermedad avanzada. Pese a que casi todos los pacientes con cáncer de próstata avanzado responden inicialmente a dicha terapia, prácticamente todos pacientes desarrollan resistencia y progresan hacia el cáncer de próstata resistente a la castración (CRPC), una fase clínica de peor pronóstico (Dutt & Gao, 2009).

Si bien algunos factores de riesgo pueden atenuarse limitando el consumo de alimentos ricos en grasa, aumentando la ingesta de verduras y frutas y promoviendo la actividad física, todavía no existen pruebas concluyentes sobre cómo prevenir el cáncer de próstata (Rawla, 2019), el cual es conocido por su marcada heterogeneidad molecular. Debido a que distintas zonas de un mismo tumor pueden mostrar perfiles genómicos y transcriptómicos diferentes, el diagnóstico y el tratamiento de esta enfermedad se complican. Es por ello por lo que resulta fundamental desarrollar nuevas metodologías para caracterizar la heterogeneidad intratumoral y su impacto en la progresión de la enfermedad (Haffner et al., 2020).

En este contexto, la heterogeneidad transcriptómica de los tumores de próstata y de su microambiente tumoral (TME) es poco conocida. Estudios recientes han mostrado la activación de múltiples programas de expresión génica en el TME asociados a la progresión tumoral y la resistencia terapéutica (Chen et al., 2021). No obstante, la integración comparativa de la expresión génica entre tejido prostático sano, tejido cáncer de próstata localizado (PCA) y CRPC sigue siendo un reto analítico y biológico. Este trabajo busca aportar una caracterización comparativa reproducible desde la bioinformática dentro del área del análisis de datos ómicos.

Las técnicas tradicionales de análisis transcriptómico, como el bulk RNA-seq, enmascaran la heterogeneidad celular existente dentro del tejido tumoral (Stuart & Satija, 2019a). No obstante, la transcriptómica unicelular (single cell RNA sequencing, scRNA-seq) permite cuantificar la expresión génica a nivel individual, identificando subpoblaciones celulares, trayectorias de diferenciación y estados funcionales asociados con la carcinogénesis prostática (Song et al., 2022), lo que facilita comprender los mecanismos moleculares de progresión hacia el estado resistente a la castración.

Desde la perspectiva de la bioinformática y el análisis de datos ómicos, el estudio de datos scRNA-seq plantea desafíos debido a la alta dimensionalidad, el ruido técnico y la necesidad de integrar datos provenientes de diferentes condiciones biológicas.

Herramientas como Seurat permiten el procesamiento, normalización y análisis de datos unicelulares (Stuart et al., 2019a), mientras que métodos de integración como Harmony facilitan la corrección de efectos de lote y la combinación de datasets heterogéneos (Korsunsky et al., 2019a).

Así pues, el presente trabajo se enmarca en el área de Omics Data Analysis aplicada a la transcriptómica unicelular, con el objetivo de comparar las poblaciones celulares y sus perfiles de expresión génica entre tejido prostático sano, PCa y CRPC. A través de un pipeline bioinformático reproducible, se busca caracterizar las diferencias moleculares y proporcionales entre los distintos estados de la enfermedad, contribuyendo así a una mejor comprensión de la heterogeneidad tumoral y a la generación de conocimiento transferible para la investigación biomédica.

## 1.2. Objetivos del trabajo.

### 1.2.1. Objetivos generales.

1. Desarrollar un flujo de análisis bioinformático reproducible para el estudio de datos de transcriptómica unicelular (*scRNA-seq*) aplicado al cáncer de próstata.
2. Caracterizar y comparar las poblaciones celulares y los perfiles de expresión génica entre tres condiciones biológicas: tejido prostático sano, PCa y CRPC, con el fin de identificar patrones moleculares asociados a la progresión tumoral.

### 1.2.2. Objetivos específicos.

1. Implementar un pipeline de análisis de datos *scRNA-seq* basado en las herramientas Seurat y Harmony, asegurando la reproducibilidad y trazabilidad de cada etapa.
2. Realizar el preprocesamiento y control de calidad (QC) de los datos, filtrando células de baja calidad y genes poco expresados.
3. Aplicar técnicas de normalización e integración para combinar muestras provenientes de diferentes condiciones biológicas.
4. Ejecutar análisis de reducción de dimensionalidad y clustering (PCA, UMAP) para identificar subpoblaciones celulares presentes en cada condición.
5. Anotar las poblaciones celulares basándose en marcadores génicos específicos y bases de datos de referencia.
6. Comparar las poblaciones relativas de las distintas poblaciones celulares entre tejido sano, PCa y CRPC.
7. Detectar genes diferencialmente expresados (DEGs) entre los grupos de estudio y caracterizar sus posibles funciones biológicas.
8. Generar visualizaciones (UMAPs, mapas de calor, gráficos de expresión) que permitan interpretar los resultados de forma clara y comprensible.

### 1.3. Impacto en sostenibilidad, ético-social y de diversidad.

El presente trabajo se enmarca en la investigación biomédica computacional, contribuyendo a los Objetivos de Desarrollo Sostenible (ODS), destacando en el ODS 3: Salud y bienestar, ya que promueve el conocimiento de los mecanismos moleculares implicados en la progresión del cáncer de próstata. Ampliar conocimientos sobre la heterogeneidad celular y transcriptómica en este tipo de tumores puede ayudar al desarrollo de estrategias terapéuticas más eficaces y personalizadas, fomentando un uso más racional y sostenible de los recursos sanitarios.

Desde un punto de vista ético, este proyecto se basa en el uso exclusivo de datos públicos y anonimizados disponibles en repositorios de acceso abierto, siguiendo las recomendaciones internacionales sobre el tratamiento responsable de la información biomédica. No se emplean muestras biológicas propias ni datos clínicos identificables, garantizando la privacidad de los pacientes y evitando conflictos en materia de protección de datos personales o consentimiento informado.

Referente al plano social, el estudio investiga una enfermedad con un alto impacto en la salud masculina y en el envejecimiento poblacional. Además, promueve la democratización de los análisis ómicos mediante el uso de herramientas abiertas y reproducibles, acercando el conocimiento científico equitativo y reduciendo las barreras tecnológicas en investigación.

Finalmente, el proyecto se alinea con los principios de diversidad y exclusión en la ciencia, ya que se aplican metodologías extensibles a otros contextos biológicos y poblacionales, pudiendo reproducirse en distintos tipos celulares, orígenes genéticos y condiciones clínicas.

### 1.4. Enfoque y método seguido.

El estudio de la heterogeneidad celular en el cáncer de próstata puede abordarse mediante diversas estrategias experimentales y computacionales. Entre las principales aproximaciones disponibles se encuentran los análisis de transcriptómica a granel (*bulk RNA-seq*), la citometría de flujo o la microscopía de tejidos multiplexada, que permiten obtener información sobre el perfil molecular global o la composición celular aproximada del tejido. No obstante, estas metodologías presentan limitaciones importantes, ya que promedian la señal de expresión de miles de células, enmascarando así la variabilidad y los subtipos celulares presentes en el microambiente tumoral (Stuart & Satija, 2019a).

Frente a estas aproximaciones, la transcriptómica unicelular (*scRNA-seq*) ofrece una resolución sin precedentes al cuantificar la expresión génica en cada célula individual. Esta técnica permite identificar poblaciones celulares raras, detectar transiciones de estado y caracterizar la heterogeneidad intra- e intertumoral asociada a la progresión del cáncer de próstata (Song et al., 2022a). Por tanto, constituye la estrategia más adecuada para cumplir los objetivos de este trabajo, centrados en comparar las poblaciones celulares y los perfiles de expresión génica entre tejido sano, PCa y CRPC.

En el plano computacional, existen múltiples frameworks de análisis de datos unicelulares, como Scanpy (Python), Monocle (R) o Cell Ranger (10x Genomics). Sin embargo, se ha seleccionado el entorno Seurat (R) como eje central del pipeline debido a su madurez, documentación, compatibilidad y amplia adopción en la comunidad científica (Stuart et al., 2019a). Además, el uso del método de integración Harmony (Korsunsky et al., 2019a) permite corregir efectos de lote y diferencias entre muestras, garantizando comparaciones fiables entre condiciones biológicas.

El flujo metodológico propuesto se estructura en seis etapas principales:

1. Preprocesamiento y control de calidad (QC) de las matrices de expresión (Stuart et al., 2019a).
2. Normalización e integración de datasets mediante Seurat y Harmony (Korsunsky et al., 2019a; Stuart et al., 2019a).
3. Reducción de dimensionalidad y clustering para la identificación de subpoblaciones celulares (Stuart et al., 2019a).
4. Visualización y reporte de resultados mediante gráficos UMAP y mapas de calor, garantizando la reproducibilidad del análisis (Stuart et al., 2019a).
5. Anotación de tipos celulares a partir de marcadores conocidos y bases de datos de referencia (Stuart & Satija, 2019a).
6. Comparación de proporciones y análisis diferencial de expresión génica (DEG) entre los grupos sano, PCa y CRPC, seguido de visualización mediante volcano plots (Stuart et al., 2019a).

Este flujo sigue las recomendaciones metodológicas descritas en Stuart & Satija (2019) para el análisis de datos scRNA-seq e incorpora el método Harmony para la integración de muestras y corrección de efectos de lote (Korsunsky et al., 2019a), combinando así la robustez estadística con un enfoque reproducible y escalable que facilita la interpretación biológica de los resultados.

## 1.5. Planificación del Trabajo.

### 1.5.1. Tareas.

En la siguiente tabla se detallan las tareas planificadas según las fases del proyecto, su relación con los objetivos y la duración estimada para cada una:

*Tabla 1: Planificación de tareas según los objetivos específicos del proyecto.*

| Bloque de trabajo.                     | Tarea.   | Duración estimada. | Objetivo específico. |
|--|--|--------------------|----------------------|
| Definición del TFM y revisión teórica. | Definición del tema y elaboración del plan de trabajo. | 10 días.           | OE1.                 |
|  | Revisión bibliográfica.                                | 6 días.            | OE1, OE2.            |
| Procesamiento e integración de datos.  | Preprocesamiento y control de calidad (QC).            | 14 días.           | OE2.                 |
|  | Integración y normalización de datos.                  | 10 días.           | OE3.                 |

|  |   |          |                                  |
|--|---|----------|----------------------------------|
| Análisis exploratorio y clustering.          | Clustering y reducción de dimensionalidad.                | 7 días.  | OE4.                             |
|  | Anotación de poblaciones celulares.                       | 7 días.  | OE5.                             |
| Comparativa entre grupos y expresión génica. | Comparación de proporciones celulares.                    | 5 días.  | OE6.                             |
|  | Análisis diferencial de expresión génica ( <i>DEGs</i> ). | 10 días. | OE7.                             |
| Resultados, documentación y entrega final.   | Visualización y generación de resultados.                 | 17 días. | Todos los objetivos específicos. |
|  | Introducción de resultados y conclusiones en la memoria.  | 21 días. |                                  |
|  | Preparación de la presentación oral.                      | 15 días. | OE1, OE2.                        |

### 1.5.2. Calendario.

El cronograma de trabajo se representa en la Figura 1, donde se detallan las fases del proyecto, las tareas planificadas y su distribución temporal desde septiembre de 2025 hasta enero de 2026. La planificación se ha elaborado mediante GanttProject, una herramienta de gestión de proyectos que permite visualizar las dependencias entre tareas y el grado de solapamiento entre fases.

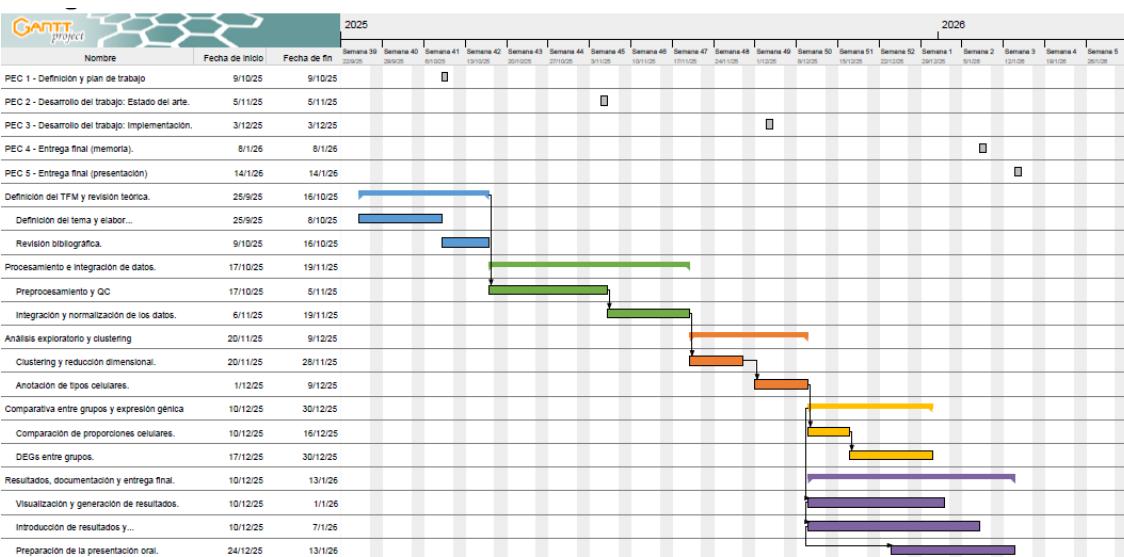


Figura 1: Diagrama de Gantt del proyecto.

El desarrollo del proyecto se estructura en cinco fases principales:

1. Definición del TFM y revisión teórica (25 septiembre 2025 - 16 octubre 2025): incluye la definición del tema, el planteamiento de objetivos y la revisión bibliográfica.
2. Procesamiento e integración de datos (17 octubre 2025 – 19 noviembre 2025): contempla el preprocesamiento, control de calidad e integración de las muestras scRNA-seq.

3. Análisis exploratorio y clustering (20 noviembre 2025 – 09 diciembre 2025): comprende la reducción de dimensionalidad, el agrupamiento celular y la anotación de poblaciones.
4. Comparativa entre grupos y expresión génica (10 diciembre 2025 – 30 diciembre 2025): centrada en la comparación de proporciones celulares y el análisis diferencial de expresión entre tejido sano, PCa y CRPC.
5. Resultados, documentación y entrega final (10 diciembre 2025 – 13 enero 2026): incluye la generación de visualizaciones, redacción de la memoria y preparación de la presentación oral.

### 1.5.3. Hitos.

Los hitos definidos permiten estructurar el proyecto en etapas verificables y coherentes con los plazos académicos establecidos. Cada entrega (PEC) implica la obtención de un conjunto de resultados parciales —ya sea la validación metodológica, los resultados biológicos o la documentación final— que aseguran la trazabilidad del proceso y la progresión ordenada hacia los objetivos generales del trabajo.

*Tabla 2: Hitos del proyecto.*

| Hito.                                  | Descripción.   | Fecha.             | Resultado esperado.   |
|--|--|--------------------|---|
| Definición del tema y plan de trabajo. | Formulación de objetivos, diseño del cronograma y planteamiento metodológico.                        | 08 octubre 2025.   | Entrega PEC 1: plan de trabajo.   |
| Desarrollo y seguimiento del proyecto. | Finalización de la revisión bibliográfica y primeras etapas de procesamiento e integración de datos. | 05 noviembre 2025. | Entrega PEC 2: datos procesados y normalizados listos para análisis exploratorio. |
| Desarrollo y seguimiento del proyecto. | Ejecución de análisis clustering, reducción de dimensionalidad y anotación de poblaciones celulares. | 03 diciembre 2025. | Entrega PEC 3: identificación y caracterización de subpoblaciones celulares.      |
| Cierre de la memoria.                  | Comparativa entre grupos (sano, PCa y CRPC), análisis diferencial de expresión y                     | 07 enero 2026.     | Entrega PEC 4: revisión final del análisis y memoria completa.                    |

|                             |   |                |   |
|-----------------------------|---|----------------|---|
|                             | visualización de resultados.                                |                |   |
| Presentación final del TFM. | Preparación de las diapositivas y defensa oral del trabajo. | 13 enero 2026. | Entrega PEC 5: presentación y evaluación final. |

#### 1.5.4. Análisis de riesgos.

Durante el desarrollo del presente Trabajo de Fin de Máster se han identificado varios factores que podrían afectar al cumplimiento del plan de trabajo y a la consecución de los objetivos. Los principales riesgos se relacionan con el alcance del proyecto, las limitaciones temporales y las dificultades técnicas asociadas al análisis de datos scRNA-seq.

En primer lugar, el alcance excesivo del proyecto constituye un riesgo relevante. El carácter exploratorio del análisis unicelular podría llevar a incluir comparaciones adicionales o análisis complementarios que excedan el tiempo disponible. Para mitigar este riesgo, se mantendrá el enfoque centrado en los objetivos definidos, limitando el estudio a la comparación entre tejido sano, PCa y CRPC.

Otro factor crítico son las limitaciones temporales derivadas del calendario académico. La estructura del proyecto deja un margen reducido para imprevistos, por lo que se realizará un seguimiento continuo de los hitos y se reservará tiempo específico para validación y redacción final.

Entre los riesgos técnicos, destaca la limitación computacional derivada del uso de un equipo de trabajo con 16 GB de memoria RAM y 10 núcleos, lo que puede restringir la carga y el procesamiento simultáneo de grandes matrices de expresión unicelular. Este riesgo se mitigará mediante la ejecución de análisis por subconjuntos, la optimización del código y el uso de versiones comprimidas de los objetos de Seurat.

Adicionalmente, la integración de muestras heterogéneas podría generar efectos de lote o sesgos, que se abordarán mediante el uso del método Harmony y evaluaciones intermedias de calidad. Finalmente, se consideran posibles dificultades en la interpretación biológica y en la reproducibilidad computacional, mitigadas mediante la documentación exhaustiva del pipeline y el uso de bases de datos y referencias actualizadas.

En conjunto, se espera que las medidas preventivas aplicadas —basadas en una planificación realista, el uso de herramientas consolidadas y la documentación rigurosa del proceso— aseguren la viabilidad técnica, la reproducibilidad y la finalización exitosa del proyecto dentro del marco temporal establecido.

#### 1.6. Breve sumario de productos obtenidos.

1. Memoria del trabajo final de máster: informe final que recoge los resultados, discusión y conclusiones del estudio.

2. Pipeline y scripts del análisis: conjunto de archivos reproducibles asociados al producto técnico del trabajo.
3. Presentación final (diapositivas y vídeo): material audiovisual destinado a la defensa final del trabajo final de máster.

## 1.7. Breve descripción de los otros capítulos de la memoria.

Este apartado ofrece una visión general de los distintos capítulos que componen la presente memoria, con el objetivo de contextualizar su estructura y facilitar su lectura.

- Capítulo 2. Introducción: Describe el contexto biológico y clínico del cáncer de próstata, con especial énfasis en los mecanismos moleculares implicados en su progresión y resistencia. Se revisa también el estado actual del conocimiento sobre el microambiente tumoral y la aplicación de transcriptómica unicelular en oncología.
- Capítulo 3. Objetivos: Expone los objetivos generales y específicos del estudio, centrados en caracterizar los cambios celulares y moleculares asociados a la progresión prostática desde tejido sano hasta CRPC mediante análisis de scRNA-seq.
- Capítulo 4. Materiales y Métodos: Detalla los datos utilizados, el pipeline bioinformático empleado para el procesamiento, integración, anotación y análisis diferencial de datos scRNA-seq, así como los criterios para la interpretación de los resultados.
- Capítulo 5. Resultados: Presenta de forma estructurada los principales hallazgos del estudio, incluyendo la identificación de tipos celulares, análisis de composición por condición, y análisis de expresión diferencial por tipo celular.
- Capítulo 6. Discusión: Interpreta los resultados en relación con el estado del arte, explorando sus implicaciones funcionales y clínicas. Se comentan patrones de desregulación molecular y posibles mecanismos asociados a la progresión tumoral.
- Capítulo 7. Conclusiones: Resume los hallazgos principales, señala las limitaciones del estudio y propone líneas futuras de investigación.
- Capítulo 8. Bibliografía: Recoge todas las referencias bibliográficas utilizadas a lo largo del trabajo.
- Capítulo 9. Anexos: Incluye el repositorio al código en R, figuras complementarias, tablas y documentación técnica relevante para reproducibilidad y transparencia del análisis.

## 2. Estado del arte.

### 2.1. Introducción al cáncer de próstata y contexto clínico.

El cáncer de próstata es el segundo cáncer más frecuentemente diagnosticado en los hombres (Rawla, 2019), y el cuarto más diagnosticado a nivel mundial, representando aproximadamente el 7,3% de casos de cáncer en el año 2022 (Bray Bsc et al., 2024). Además, constituye la quinta causa de muerte por cáncer en el mundo (Rawla, 2019). Su incidencia aumenta con la edad, diagnosticándose en su mayoría en hombres mayores de 60 años.

El crecimiento y mantenimiento del epitelio prostático dependen de niveles fisiológicos de andrógenos, que estimulan la proliferación e inhiben la apoptosis celular. En consecuencia, las células cancerosas de próstata presentan una fuerte dependencia de los andrógenos, por lo que la terapia de privación androgénica (ADT) ha sido el tratamiento principal para los pacientes con enfermedad avanzada. Pese a que casi todos los pacientes con cáncer de próstata avanzado responden inicialmente a dicha terapia, prácticamente todos pacientes desarrollan resistencia y progresan hacia el cáncer de próstata resistente a la castración (CRPC), una fase clínica de peor pronóstico (Dutt & Gao, 2009).

Así pues, la enfermedad puede dividirse en tres estados:

1. Tejido prostático sano, con una organización celular bien definida.
2. PCa localizado, donde aparecen alteraciones en la diferenciación del comportamiento epitelial, proliferación aumentada y remodelación del microambiente tumoral.
3. Cáncer de próstata resistente a la castración (CRPC), asociado con una biología tumoral más agresiva y fenotipos epiteliales aberrantes, pudiendo llegar a la transdiferenciación hacia estados neuroendocrinos (Bluemn et al., 2017; Y. Wang & Wang, 2025).

La transición hacia CRPC constituye un gran desafío clínico. La presión inducida por ADT favorece la selección de clones con activación independiente del receptor de andrógenos (AR), amplificaciones del gen AR, mutaciones, pérdida de PTEN o adquisición de programas transcripcionales alternativos vinculados a resistencia(Watson et al., 2015). En los casos más agresivos, la enfermedad puede evolucionar hacia un fenotipo endocrino, caracterizado por perdida de AR, expresión de marcadores como CHGA y SYP, y una marcada plasticidad celular (Beltran et al., 2016).

Tradicionalmente, los estudios transcriptómicos se han basado en tecnologías bulk RNA-seq, las cuales permiten obtener una medida promedio de la expresión génica en un tejido. No obstante, esta aproximación oculta la heterogeneidad celular, un aspecto clave en PCa, ya que coexisten distintos subtipos epiteliales, células inmunes infiltradas, fibroblastos asociados al cáncer (CAFs) y poblaciones vasculares en diferentes proporciones (Hirz et al., 2023; Pan et al., 2023). Esta limitación incapacita la identificación de biomarcadores y la comprensión de los mecanismos de progresión tumoral.

En este contexto, la tecnología single-cell RNA sequencing (scRNA-seq) es útil para caracterizar tejidos complejos. Su capacidad para medir la expresión génica de miles de células individualmente permite describir con precisión los linajes celulares, estados transcripcionales y mecanismos de resistencia terapéutica, demostrando ser especialmente útil en PCa para:

- Diferenciar poblaciones luminales, basales y transicionales.
- Identificar subpoblaciones epiteliales asociadas a agresividad.
- Analizar la interacción entre el tumor y su microambiente inmuno-estromal.
- Describir la plasticidad celular ligada a la transición a CRPC (Chen et al., 2021).

Así pues, el uso de scRNA-seq ha permitido redefinir el mapa celular prostático en condiciones normales y patológicas. Esta ofrece una visión mucho más detallada de cómo evoluciona la enfermedad, qué tipos celulares participan en cada estadio y qué programas transcripcionales podrían construir nuevas dianas terapéuticas.

## 2.2. Tecnologías de secuenciación a célula única (scRNA-seq).

La secuenciación de RNA a célula única (scRNA-seq) constituye uno de los avances más importantes en la biología molecular moderna. Esta tecnología permite obtener el transcriptoma completo de células individuales, superando las limitaciones de los estudios bulk, que solo capturan valores promedios (F. Tang et al., 2009). El desarrollo de scRNA-seq ha transformado la investigación biomédica al hacer posible la caracterización de heterogeneidad celular, la identificación de poblaciones raras y el estudio de transiciones dinámicas de estados celulares, aspectos muy importantes en cáncer.

### 2.2.1. Fundamentos y evolución tecnológica.

Los primeros protocolos scRNA-seq surgieron alrededor de 2009-2012 con métodos basados en amplificación mediante transcriptasa inversa y PCR (Ramsköld et al., 2012; F. Tang et al., 2009). Estos sistemas permitieron secuencias unas pocas decenas de células, pero presentaban limitaciones importantes en sensibilidad, reproducibilidad y escalabilidad.

El desarrollo completo apareció con los métodos basados en Unique Molecular Identifiers (UMIs) y barcoding masivo, especialmente a partir de la plataforma 10x Genomics Chromium (Zheng et al., 2017). La incorporación de UMIs redujo el sesgo de amplificación y mejoró la cuantificación de expresión, mientras que los microgeles emulsionados (GEMs) permitieron procesar miles de células de manera simultánea. Este enfoque inició la era de los transcriptomas masivos, permitiendo analizar decenas de miles de células en un único experimento.

Otros avances relevantes incluyen:

- Métodos basados en microplacas (Smart-seq2), que permiten capturar transcriptomas más completos, especialmente útiles para la detección de isoformas, aunque con menor escalabilidad (Picelli et al., 2014).

- Protocolos basados en droplet microfluidics (Drop-seq, inDrop), precursores conceptuales de 10x, que facilitaron la miniaturización de reacciones en miles de gotas independientes (Macosko et al., 2015).
- Tecnologías multi-ómicas a célula única, como CITE-seq o scATAC-seq, que combinan transcriptómica con información epigenética o proteómica (Stoeckius et al., 2017).

Gracias a estos desarrollos, scRNA-seq se ha consolidado como una herramienta estándar en estudios de cáncer, neurociencia, desarrollo embrionario y biomedicina traslacional.

### **2.2.2. Procesamiento y retos analíticos.**

El análisis de datos de scRNA-seq implica una cadena metodológica compleja que incluye filtrado, normalización, corrección de efectos de lote, detección de subpoblaciones y anotación. Aunque existen diferentes frameworks, uno de los más populares es Seurat, que proporciona funciones para todas las etapas del pipeline (Hao et al., 2021).

Entre los principales retos al procesamiento se destacan:

- Ruido técnico y dropout evento: el escaso contenido de ARN en células individuales genera tasas elevadas de genes no detectador (dropouts). Esto puede llevar a errores en la identificación de subpoblaciones si no se aplican métodos adecuados de normalización y modelado (Luecken & Theis, 2019).
- Variabilidad entre muestras (batch effect): la integración de múltiples muestras requiere de métodos de corrección robustos. Actualmente destacan enfoques como Harmony (Korsunsky et al., 2019b), Seurat V4 integration o LIGER, los cuales preservan la señal biológica evitando sobre correcciones.
- Escalabilidad computacional: los datasets actuales pueden superar las 500.000 células, lo que demanda algoritmos optimizados en tiempo y memoria. Métodos como SCTransform introducen modelos basados en regresión regularizada para estabilizar la varianza y mejorar el rendimiento (Hafemeister & Satija, 2019).

### **2.2.3. Aplicaciones en cáncer y microambiente tumoral.**

La aplicación de scRNA-seq ha sido muy importante en oncología, permitiendo:

- Descubrir subpoblaciones epiteliales raras asociadas a agresividad tumoral.
- Caracterizar estados celulares transicionales presentes durante la progresión y la resistencia terapéutica.
- Mapear la interacción entre tumor, sistema inmune y estroma.
- Identificar programas de plasticidad tumoral, un rasgo central en cáncer de próstata avanzado (Beltran et al., 2016; Chen et al., 2021).

Actualmente, scRNA-seq constituye el estándar para generar atlas celulares y estudiar la evolución tumoral desde una perspectiva integrativa, combinando genómica, transcriptómica y fenómenos regulatorios.

## 2.3. Análisis computacional en scRNA-seq.

En análisis de datos de scRNA-seq es necesario una cadena metodológica debido al elevado nivel de ruido, la dispersión técnica y la complejidad estructural de los transcriptomas a célula única. Por ello, en los últimos años se han desarrollado pipelines estandarizados que permiten abordar desde las fases iniciales de control de calidad hasta la integración, anotación y análisis diferencial. Entre los más destacados encontramos Seurat, Scanpy y Bioconductor, cada uno con implementaciones específicas, pero con un flujo conceptual común (Luecken & Theis, 2019; Stuart et al., 2019b; Stuart & Satija, 2019b).

### 2.3.1. Control de calidad y filtrado.

El primer paso consiste en identificar células de baja calidad y artefactos técnicos. Se emplean métricas como:

- Número de genes detectados por célula (nFeature\_RNA),
- Número total de UMI o lecturas (nCount\_RNA),
- Proporción de genes mitocondriales (indicador de estrés o apoptosis),
- Proporción de genes ribosomales (estado biosintético),
- Expresión de marcadores nucleares como MALAT1, útil para identificar núcleos incompletos o rotos (Clarke & Bader, 2024).

Las guías recientes recomiendan aplicar umbrales adaptativos por muestra, en lugar de cortes fijos, ya que las distribuciones pueden variar entre tejidos, plataformas y condiciones experimentales (Clarke & Bader, 2024).

### 2.3.2. Normalización y selección de genes variables.

Una vez filtrados los datos, se procede a la normalización. Los métodos clásicos basados en log-normalización dependen del tamaño de la biblioteca, por lo que se pueden introducir sesos. Una alternativa muy aceptada es SCTransform, un enfoque basado en regresión negativa binomial regularizada que estabiliza la varianza y reduce el impacto de la profundidad de secuenciación (Hafemeister & Satija, 2019).

En paralelo, se identifican genes altamente variables (HVGs), esenciales para capturar la señal biológica en los pasos posteriores de reducción dimensional y clustering (Hao et al., 2021).

### 2.3.3. Reducción dimensional y corrección de lotes.

La reducción dimensional se realiza comúnmente mediante PCA, seleccionando las primeras componentes que capturen la mayor variabilidad biológica. Sin embargo, cuando se integran muestras es necesario corregir el efecto de lote. Existen para ello diversas maneras:

- Harmony, que ajusta de forma iterativa los embeddings preservando la estructura biológica (Korsunsky et al., 2019b).
- Seurat Integration, basado en anclajes entre datasets (Stuart & Satija, 2019b).
- BBKNN o Scanorama, usados en Python (Hie et al., 2024).

Esta es un paso muy crítico del pipeline, ya que una sobre corrección puede eliminar diferencias reales entre condiciones.

#### 2.3.4. Construcción de grafos y clustering.

La identificación de subpoblaciones celulares se basa en grafos de vecinos más cercanos (KNN) construidos sobre las PCs seleccionadas. Sobre este grafo se aplica un algoritmo de clustering, habitualmente:

- Louvain o
- Leiden, una versión mejorada más robusta y que evita particiones mal definidas (Traag et al., 2019).

Este también es un parámetro crítico, ya que valores bajos generan agrupaciones amplias y valores altos producen subclústeres exagerados. Se recomienda evaluar distintas resoluciones y contrastar la estabilidad de las particiones con métricas como ARI y NMI (Luecken & Theis, 2019).

#### 2.3.5. Anotación de tipos celulares.

Este apartado puede abordarse mediante:

- Métodos basados en marcadores conocidos, como EPCAM para epitelio, CD3D para células T, etc.
- Métodos automatizados, como ScType (Aran et al., 2019; Domínguez Conde et al., 2022; Ianevski et al., 2022), SingleR o CellTypist.
- Referencias de atlas, como Human Cell Atlas o Tabula Sapiens (Consortium, 2019; Jones et al., 2022; Regev et al., 2017).

El proceso se completa con una curación manual, especialmente importante en tejidos heterogéneos como el de próstata, donde existen subpoblaciones epiteliales y estromales de gran complejidad (Henry et al., 2018; Karthaus et al., 2020).

#### 2.3.6. Visualización: UMAP y t-SNE.

Las técnicas de reducción no lineal, como UMAP, se han convertido en el estándar para visualizar relaciones locales entre celular y la estructura global del dataset, superando en estabilidad a t-SNE (Becht et al., 2018).

#### 2.3.7. Análisis diferencial y caracterización funcional.

El análisis de genes marcadores mediante métodos como FindAllMarkers o modelado estadístico avanzado permite identificar:

- Genes diferenciales por clúster.
- Firmas transcriptómicas de estados funcionales.
- Rutas señalizadoras activas mediante GSEA, PROGENy o AUCell (Aibar et al., 2017; Schubert et al., 2018; Subramanian et al., 2005).

## 2.4. Aplicaciones del scRNA-seq en cáncer de próstata.

### 2.4.1. Heterogeneidad del epitelio prostático.

El epitelio prostático humano constituye un tejido altamente organizado compuesto principalmente por tres compartimentos celulares: células luminales, células basales y células neuroendocrinas. Tradicionalmente, los estudios basados en bulk RNA-seq o inmunohistoquímica describían este epitelio como relativamente estable, sin embargo, la incorporación de scRNA-seq ha mostrado una diversidad celular mucho mayor de la que se sospechaba inicialmente (Henry et al., 2018; Karthaus et al., 2020).

En tejido sano, las células luminales representan el compartimento mayoritario y se caracterizan por la expresión de marcadores como KLK3, KLK2, ACPP y la actividad del receptor de andrógenos. No obstante, los estudios de scRNA-seq han demostrado que este compartimento no es homogéneo, si no que incluye subpoblaciones con funciones diferenciadas, como luminales metabólicas, secretoras, inflamatorias o mucinosas (Crowley et al., 2020). Esta estratificación fina es clave para comprender la vulnerabilidad diferencial de cada subtipo frente a la transformación tumoral.

El compartimiento basal, tradicionalmente asociado al mantenimiento del epitelio, también presenta mayor diversidad de la esperada. Aunque estas células expresan KRT5, KRT14 y TP63, estudios recientes han identificado estados progenitores intermedios con plasticidad hacia linajes luminales, especialmente relevantes durante la carcinogénesis (Guo et al., 2020; Karthaus et al., 2020).

En fase tumoral, la heterogeneidad se amplifica. Diferentes estudios de scRNA-seq han evidenciado la presencia de subpoblaciones luminales tumorales con perfiles diferenciados, incluyendo células CRISP3-high, KLK11+, KLK12+, células con alta expresión de PCA3, estados luminales inflamatorios y células luminales alteradas con programas de estrés o proliferación aumentada (Berglund et al., 2018; Chen et al., 2021). Esta heterogeneidad está fuertemente asociada con la progresión clínica, la respuesta al tratamiento y el desarrollo de resistencia hormonal.

En conjunto, se ha transformado la comprensión de la biología del epitelio prostático, mostrando que tanto en condiciones fisiológicas como en cáncer existe una compleja diversidad transcripcional que solo puede capturarse adecuadamente mediante análisis a resolución unicelular. Este conocimiento es clave para contextualizar los resultados del presente proyecto, donde se identifican múltiples subpoblaciones luminales, basales y epiteliales alterados a lo largo de Healty → PCa → CRPC.

### 2.4.2. Microambiente tumoral y estroma en cáncer de próstata.

El microambiente tumoral (TME) desempeña un papel central en la progresión del cáncer de próstata y en el desarrollo de resistencia terapéutica. La introducción del scRNA-seq ha permitido descomponer este ecosistema y caracterizar de forma detallada los distintos compartimentos estromales, inmunitarios y vasculares que rodean al tumor (Chen et al., 2021; Hirz et al., 2023).

### *Fibroblastos asociados al cáncer (CAFs).*

Los CAFs constituyen una de las poblaciones más influyentes en el TME prostático. Estudios unicelulares han identificado múltiples subtipos, como CAFs inflamatorios, miofibroblásticos y perivasculares, cada uno con funciones específicas en la remodelación de la matriz extracelular, la secreción de factores paracrinos y la promoción de fenotipos invasivos en las células epiteliales tumorales (Pan et al., 2023; Song et al., 2022b). En particular, los CAFs se expanden drásticamente en PCa avanzado y CRPC, donde contribuyen a la resistencia al tratamiento hormonal a través de la secreción de IL-6, TGF- $\beta$  y factores de activación del AR independiente de andrógenos (Chen et al., 2021; Kalluri, 2016).

### *Células inmunitarias del microambiente.*

El infiltrado inmune en cáncer de próstata suele caracterizarse por una marcada inmunosupresión, en contraste con otros tumores sólidos más inmunogénicos. A nivel unicelular, se han observado:

- Macrófagos M2 con programas inmunosupresores dominantes (CSF1R+, MRC1+) asociados a peor pronóstico (Chen et al., 2021)
- Linfocitos T CD8+ disfuncionales, con expresión elevada de genes de agotamiento como PDCD1, LAG3 y HAVCR2 (Petitprez et al., 2019).
- Células T reguladoras (Tregs) enriquecidas en estados avanzados, contribuyendo a la evasión inmune.
- Una infiltración muy limitada de células NK y ausencia de programas citotóxicos robustos, lo que explica la respuesta pobre a las inmunoterapias estándar.

Estos hallazgos han impulsado nuevas terapias basadas en reprogramar el microambiente inmunitario, como la inhibición de CSF1R o el uso de combinaciones anti-AR + anti-PD1- (Chesner et al., 2025; Ge et al., 2022)

### *Vasculatura tumoral: células endoteliales y pericitos.*

La angiogénesis en cáncer de próstata, aunque más moderada que en otros tumores, constituye un componente clave del TME. Los análisis de scRNA-seq han revelado:

- Subtipos endoteliales activados con perfiles pro-angiogénicos (FLT1+, KDR+).
- Pericitos con programas contráctiles y de remodelación vascular (ACTA2+, TAGLN+).
- Estados de transición endotelio-mesénquima (endMT), vinculados a metástasis y resistencia a antiandrógenos (Pan et al., 2023; Song et al., 2022b).

### *Interacciones tumor-estroma a nivel unicelular.*

Los estudios recientes han demostrado que la progresión PCa → CRPC se acompaña de una reconfiguración profunda del microambiente, donde:

- Los CAFs inducen plasticidad epitelial y fenotipos EMT-like.
- Los macrófagos M2 y Tregs inhiben la activación citotóxica del infiltrado T.
- Los pericitos y células endoteliales favorecen la creación de nichos vasculares favorables a la diseminación tumoral.

- Se producen comunicaciones específicas entre clústeres tumorales y estromales, mediadas por ligando-receptor (CXCL12/CXRC4, TGFB1/TGFBR1, IL6/IL6R), fácilmente identificables mediante scRNA-seq.

Todo esto ha posicionado el estudio del TME como un pilar fundamental para comprender la progresión tumoral y para desarrollar nuevas estrategias terapéuticas dirigidas.

## 2.5. Anotación de tipos celulares.

La anotación de tipos celulares constituye una de las etapas más importantes en el análisis de datos de scRNA-seq, ya que permite interpretar biológicamente los clústeres obtenidos tras los procesos de normalización, integración y clustering. Su objetivo es asignar una identidad celular a cada grupo basándose en patrones de expresión génica, firmas transcriptómicas y modelos de referencia. Debido a la elevada complejidad de los tejidos humanos y a la presencia de estados transicionales, la anotación celular requiere un equilibrio entre métodos computacionales y conocimiento biológico (Luecken & Theis, 2019).

### 2.5.1. Métodos basados en marcadores clásicos.

El enfoque tradicional consiste en asignar identidades celulares a partir de la expresión de genes marcadores conocidos. Este método ha sido la base de la inmunohistoquímica y de los estudios bulk, y continúa siéndolo para scRNA-seq para validar las anotaciones automáticas. Ejemplos comunes incluyen:

- Epitelio prostático:
  - o Luminal: KLK3, KLK2, ACPP, KRT8, KRT18.
  - o Basal: KRT5, KRT14, TP63.
  - o Neuroendocrino: CHGA, SYP, ENO2.
- Sistema immune:
  - o Células T: CD3D, CD3E.
  - o Células B: MS4A1 (CD20).
  - o Macrófagos: CD68, MRC1.
- Estroma:
  - o Fibroblastos/CAFs: COL1A1, ACTA2, TAGLN.
  - o Endoteliales: PECAM1 (CD31), VWF.
  - o Pericitos: RDG5, PDGFRB.

Estos marcadores se encuentran recogidos en bases de datos como CellMarker 2.0 (Hu et al., 2023a) o PanglaoDB (Franzén et al., 2019), que constituyen recursos fundamentales para la anotación inicial.

Este método incluye una serie de ventajas, como es la simplicidad y la alta interpretabilidad, así como es adecuado para tejidos estudiados de manera extensiva. No obstante, es incapaz de capturar estados transicionales o subtipos inéditos, así como tiene una dependencia directa del conocimiento previo, lo que puede inducir a sesgos.

### 2.5.2. Métodos automatizados basados en referencia.

La necesidad de manejar datasets cada vez más grandes ha impulsado el desarrollo de métodos de anotación automatizada. Estos algoritmos asignan tipos celulares comparando la expresión génica de cada célula o clúster con atlas de referencia o con firmas transcriptómicas previamente curadas.

Entre los métodos más destacados destacan:

- ScType: asignación basada en firmas de marcadores positivos y negativos específicos por tejido. Permite identificar subtipos finos y minimiza asignaciones erróneas por ruido técnico (Ianevski et al., 2022).
- SingleR: utiliza correcciones entre el transcriptoma de cada célula y bases de datos de referencia anotadas manualmente (Aran et al., 2019).
- CellTypist: modelo de machine learning con clasificadores entrenados en grandes atlas celulares humanos (Domínguez Conde et al., 2022).
- Garnett: utiliza clasificadores jerárquicos basados en marcadores definidos por el usuario (Pliner et al., 2019).

Este método supone una serie de ventajas: la escabilidad para cientos de miles de células, la reproducibilidad y reducción de subjetividad y la integración con referencias homogéneas. No obstante, presenta un listado de limitaciones, como el riesgo de asignaciones incorrectas si el tipo celular no está presente en la referencia, la dificultad para diferenciar subestados o fenotipos aberrantes tumorales y que a menudo requieren de curación manual posterior, especialmente en tejidos complejos como el prostático.

### 2.5.3. Atlas celulares como referencia.

El desarrollo de atlas celulares ha redefinido el estándar de anotación:

- Human Cell Atlas (HCA) (Regev et al., 2017).
- Tabula Sapiens (Jones et al., 2022).
- Human Biomolecular Atlas Program (HuBMAP) (Consortium, 2019).

Estos recursos permiten comparar transcriptomas individuales con miles de tipos celulares bien clasificados, proporcionando una base sólida para anotaciones iniciales.

Sin embargo, los atlas son generalmente elaborados a partir de tejidos sanos, y por tanto, no reflejan la plasticidad tumoral, especialmente relevante en cáncer de próstata avanzado, fenotipos luminales aberrantes, estados neuroendocrinos inducidos por terapia y transiciones epitelio-mesénquima (EMT-like).

Por ello, los métodos basados exclusivamente en atlas suelen requerir ajustes manuales.

### 2.5.4. Necesidad de curación manual en cáncer de próstata.

Como se ha comentado, el tejido prostático presenta una heterogeneidad celular muy pronunciada, con subpoblaciones luminales, basales y secretoras difícilmente clasificables mediante algoritmos genéricos. Estudios recientes han demostrado:

- Presencia de subtipos luminales muy específicos (KLK11+, KLK12+, PCA3-high, CRISP3-high) (Chen et al., 2021).
- Estados transicionales con plasticidad epitelial relevantes en la progresión hacia CRPC (Karthaus et al., 2020).
- Fenotipos neuroendocrinos emergentes bajo presión terapéutica (Beltran et al., 2016).
- CAFs y macrófagos tumorales con subestructura funcional heterogénea (Kumar et al., 2016).

Ningún algoritmo captura completamente esta complejidad, razón por la cual la curación manual basada en expresión de marcadores + literatura especializada es indispensable.

En el contexto de un proyecto de scRNA-seq en próstata, el enfoque combinado es el más eficiente:

Anotación final = (ScType / SingleR / CellTypist) + Curación Manual basada en marcadores y bibliografía.

Este enfoque es el estándar en publicaciones recientes de referencia y el recomendado por revisiones metodológicas actuales (Clarke & Bader, 2024; Luecken & Theis, 2019).

## 2.6. Avances recientes en scRNA-seq aplicado a cáncer de próstata.

En los últimos años, el uso de scRNA-seq en cáncer de próstata ha permitido generar nuevos modelos de progresión tumoral, redefinir la heterogeneidad epitelial e identificar mecanismos de resistencia terapéutica. Para conocer estos, se realiza un resumen de los avances más importantes producidos entre 2020 y 2024.

### 2.6.1. Redefinición del atlas celular prostático en tejido sano y tumor.

Los estudios unicelulares ya habían mostrado una diversidad mayor de la esperada en el epitelio prostático (Henry et al., 2018; Karthaus et al., 2020), pero esto ha sido ampliado.

Uno de los avances más importantes fue la caracterización exhaustiva de subpoblaciones luminales y basales en tejido sano y PCa utilizando scRNA-seq combinado con validación histológica (Crowley et al., 2020). Este estudio demostró que el epitelio luminal incluye múltiples estados transcripcionales diferenciados (secretorios, metabólicos, inflamatorios y mucinosos), sentando las bases para comprender como cada subtipo responde de manera distinta a la carcinogénesis.

### 2.6.2. Identificación de subpoblaciones luminales asociadas a agresividad tumoral.

Varios trabajos han identificado perfiles luminales alterados vinculados a agresividad clínica:

- Células CRISP3-high, asociadas a invasividad y progresión hacia estados resistentes.
- Subpoblaciones PCA3-high, con firmas transcripcionales aberrantes características del epitelio tumoral alterado.
- Estados luminales inflamatorios con activación de NF-kB.
- Células proliferativas transversales presentes tanto en epitelio como en estroma.

Estos hallazgos aparecen de forma reiterada en estudios como Berglund et al. (2018) y Chen et al. (2021), que combinaron scRNA-seq de tumores primarios y metastásicos con análisis funcionales, demostrando la relevancia de estas poblaciones en la progresión PCa → CRPC.

### 2.6.3. Plasticidad celular y transición hacia fenotipos neuroendocrinos (NEPC).

Se ha producido un gran avance en la caracterización de la transición a NEPC (prostate cancer neuroendocrine), una forma altamente agresiva de la enfermedad que surge como mecanismo de escape a la terapia antiandrogénica.

Estudios como Beltran et al. (2016) y Bluemn et al. (2017) establecieron la base molecular de esta transición, no obstante, estudios recientes han propuesto modelos celulares detallados gracias a scRNA-seq:

- Guo et al. (2020) demostraron que células luminares AR+ pueden reprogramarse hacia estados NEPC a través de programas transcripcionales mediados por SOX2, EXH2 y MYCN.
- Estudios de transcriptómica unicelular y análisis evolutivos han descrito estados transicionales con rasgos NE-like y EMT-like durante la progresión hacia CRPC y neuroendocrine prostate cancer (Beltran et al., 2016; Bluemn et al., 2017; Song et al., 2022b)

Así pues, la plasticidad celular es un mecanismo adaptativo clave en CRPC resistente a terapias hormonales.

### 2.6.4. Redes de comunicación célula-célula en CRPC.

Este es uno de los avances más relevantes para entender la contribución del microambiente tumoral (TME) en la progresión.

Trabajos como Kumar et al. (2016) han demostrado que:

- Los CAFs (fibroblastos asociados a cáncer) desarrollan una señalización aumentada vía IL-6, TGF- $\beta$  y CXCL12, que índice plasticidad epitelial y resistencia a antiandrógenos.
- Los macrófagos M2 se comunican con células epiteliales tumorales mediante vías como TGFB1/TGFRB1 y CSF1/CSF1R, promoviendo un ambiente inmunsupresor.
- Se identifican circuitos específicos luminal-CAFs-Tregs mediante análisis de ligando-receptor, clave en CRPC.

Esto ha sido posible gracias a herramientas como CellPhoneDB, NicheNet o CellChat (Browaeys et al., 2019; Efremova et al., 2020; Jin et al., 2021).

## 2.6.5. Aplicaciones clínicas emergentes: biomarcadores y medicina de precisión.

El scRNA-seq ha comenzado a generar implicaciones clínicas directas:

- Identificación de biomarcadores epiteliales de resistencia a AE (KLK11+, KLK12+, CRISP3-high).
- Detección de subpoblaciones con potencial metastásico.
- Identificación de dianas terapéuticas en el estroma (CSF1R, TGFVR2, IL6R).
- Desarrollo de modelos predictivos basados en firmas transcriptómicas específicas de células raras.

El uso combinado de organoides derivados de pacientes y tecnologías ómicas, incluyendo scRNA-seq, se ha propuesto como una vía prometedora para predecir respuesta a terapias antiandrogénicas y diseñar estrategias de medicina personalizada (Gao et al., 2014; Rago et al., 2023).

## 2.6.6. Síntesis y relación con el presente proyecto.

Los avances recientes demuestran que:

1. El epitelio prostático es altamente heterogéneo y dinámico.
2. La transición hacia CRPC implica estados celulares intermedios detectables solo mediante scRNA-seq.
3. El microambiente tumoral reconfigura profundamente la progresión del PCa avanzado.
4. La identificación de subpoblaciones luminales específicas es clave para desarrollar nuevos biomarcadores y terapias.

El presente proyecto se enmarca directamente dentro de esta línea de investigación, ya que:

- Integra múltiples muestras para caracterizar la heterogeneidad epitelial.
- Identifica clústeres luminales, basales y estados alterados como CRIP3-high o PCA3-high.
- Describe como estas poblaciones cambian en Healthy → PCa → CRPC.
- Analiza la composición celular y sus posibles implicaciones biológicas.

Por ello, este trabajo representa una aplicación directa de los avances metodológicos y conceptuales desarrollados durante los últimos años.

### 3. Materiales y métodos.

#### 3.1. Obtención y caracterización de los datos.

Todos los datos provienen de repositorios públicos de NCBI GEO, generados mediante tecnología 10x Genomics Chromium y secuenciados con Illumina NovaSeq 6000, quedando garantizada la compatibilidad técnica entre los estudios.

- Healthy (GSE141445, Crowley et al., 2020). Este estudio proporciona muestras de tejido prostático benigno, cada una con entre 1500 y 3000 por muestra. Se seleccionaron tres muestras representativas para mantener un tamaño equilibrado con el resto de las condiciones.
- PCA (GSE150692, Chen et al., 2021). Se selecciona la matriz integrada GSM4203181\_data.matrix.txt.gz, que contiene 13 muestras de tejido tumoral primario de pacientes con adenocarcinoma prostático. Aunque la plataforma original asociada al estudio ([www.pradcellatlar.com](http://www.pradcellatlar.com)) ya no está disponible en la fecha de la consulta (octubre de 2025), el dataset se conserva íntegro en GEO. Dada la ausencia de metadatos por célula, se trató la matriz como una muestra única representativa del tejido tumoral localizado, empleando un subconjunto de 5000 células por limitaciones de memoria (16 GB de RAM).
- CRPC (GSE210358, Chan et al., 2022). Este estudio tiene 14 muestras de cáncer de próstata resistente a castración. Cada una dispone de metadatos detallados (condición, tipo celular y origen anatómico). Para equilibrar las condiciones y optimizar el uso de la memoria, se seleccionaron 3 muestras representativas.

En conjunto, los datos finales integran 31.606 células distribuidas en siete muestras, reflejando las tres etapas de progresión de la enfermedad.

#### 3.2. Procesamiento y control de calidad.

Cada dataset fue importado y transformado en un objeto Seurat independiente (Butler et al., 2018).

Para ello, se diseñaron funciones específicas en R adaptadas a cada formato de origen, garantizando una importación reproducible y estructurada.

Una vez creados los objetos Seurat, se verificaron mediante:

- Dimensión de la matriz (genes x células).
- Metadatos coherentes (sample\_id, condition, percent.mt).
- Comprobación de nombres génicos únicos.

El cálculo del porcentaje de genes mitocondriales (percent.mt) se realiza mediante la función PercentageFeatureSet() de Seurat, que identifica los genes con prefijo “MT-“. En el caso de PCa, los valores resultantes fueron 0%, ya que los genes mitocondriales habían sido previamente filtrados por los autores para reducir el ruido técnico (Luecken & Theis, 2019).

El filtrado se efectuó muestra por muestra, en lugar de hacerlo tras la integración, para reducir carga de memoria del ordenador.

Sobre cada objeto Seurat se calculan métricas estándar de calidad (nFeature\_RNA, nCount\_RNA, percent.mt, percent.ribo, expresión de MALAT1). A partir de estas, se generan visualizaciones específicas (boxplots e histogramas con densidad recortados entre los percentiles 1 y 99) y se aplican filtros adaptativos por muestra (Clarke & Bader, 2024; Luecken & Theis, 2019; McCarthy et al., 2017; Montserrat-Ayuso & Esteve-Codina, 2024).

La normalización y detección de genes variables se realizó con SCTransform (Hafemeister & Satija, 2019), que modela la expresión génica mediante regresión negativa binomial regularizada, corrigiendo el efecto de profundidad de secuenciación. Se seleccionaron las 3000 características génicas más variables, siguiendo la configuración recomendada en Seurat v5 (Stuart et al., 2019b).

### 3.3. Integración de datasets, reducción dimensional y clustering.

Tras el control de calidad y la normalización individual de cada muestra, los distintos objetos Seurat correspondientes a tejido sano (Healthy), cáncer de próstata localizado (PCa) y cáncer de próstata resistente a la castración (CRPC) se integraron en un único objeto para permitir comparaciones directas entre condiciones biológicas.

La integración se realizó tras la normalización mediante SCTransform, utilizando las características variables previamente identificadas en cada muestra. Este enfoque permite reducir el impacto de variabilidad técnica asociada a la profundidad de secuenciación y a diferencias globales entre experimentos, preservando al mismo tiempo la variabilidad biológica relevante<sup>681921</sup> (Hafemeister & Satija, 2019; Stuart et al., 2019c).

Dado que los datos proceden de estudios independientes y de múltiples muestras, se aplicó una corrección explícita de efectos de lote mediante el algoritmo Harmony (Korsunsky et al., 2019b). Harmony ajusta las componentes principales obtenidas a partir de la PCA inicial, alineando las distribuciones de las células en el espacio reducido en función de una variable de lote definida. En este trabajo, la variable empleada fue sample\_id, lo que permite corregir diferencias sistemáticas entre muestras individuales sin forzar la mezcla artificial de poblaciones biológicamente distintas. La calidad de la integración se evaluó mediante la inspección visual del embedding resultante y la comprobación de que las poblaciones celulares no se segregaban exclusivamente por origen muestral.

Como paso previo al clustering, se realizó un análisis de componentes principales (PCA) sobre el objeto integrado, empleando las características normalizadas mediante SCTransform. El número de componentes principales a retener se seleccionó de forma empírica, combinando la inspección del elbow plot con la evaluación de la estabilidad de los clústeres obtenidos al variar el número de componentes. Finalmente, se seleccionaron 30 componentes principales, un valor que proporciona un compromiso adecuado entre reducción de ruido y preservación de la estructura biológica, y que es coherente con recomendaciones habituales en análisis de scRNA-seq de tamaño similar.

A partir del espacio reducido (Harmony), se construyó un grafo de vecinos más cercanos mediante la función FindNeighbors, y se aplicó un algoritmo de clustering basado en grafos (Leiden) para identificar grupos celulares transcripcionalmente homogéneos. La resolución del clustering se exploró de manera sistemática, evaluando distintos valores y seleccionando aquella que producía una partición biológicamente interpretable, evitando tanto la sobrefragmentación de poblaciones conocidas como la fusión artificial de tipos celulares distintos. La resolución final se fijó en 0.5, tras comprobar la estabilidad de los clústeres y su coherencia con marcadores celulares conocidos.

Para la visualización de la estructura global de los datos, se utilizó UMAP (Uniform Manifold Approximation and Projection), calculado a partir del espacio corregido por Harmony. Este embedding bidimensional se empleó como referencia visual en las etapas posteriores de anotación celular y análisis comparativos entre condiciones, permitiendo evaluar tanto la separación entre poblaciones celulares como su grado de solapamiento entre estados de la enfermedad.

### 3.4. Identificación de genes característicos por clúster (TOP markers).

Con el objetivo de apoyar la interpretación biológica de los clústeres obtenidos tras la integración y el clustering, se llevó a cabo la identificación de los genes marcadores característicos de cada clúster mediante un análisis de expresión diferencial entre clústeres. Este análisis proporciona la base transcriptómica necesaria para evaluar la coherencia del clustering y para sustentar los procesos posteriores de anotación automática y manual.

La detección de genes marcadores se realizó utilizando la función FindAllMarkers() de Seurat, aplicando un enfoque de tipo uno contra todos, en el que cada clúster se compara frente al conjunto del resto de células. El análisis se restringió a genes sobreexpresados en el clúster de interés, aplicando criterios estándar ampliamente utilizados en estudios de scRNA-seq (Hao et al., 2021; Stuart et al., 2019):

- Se consideraron únicamente genes con expresión positiva en el clúster (only.pos = TRUE).
- Se exigió que el gen estuviese expresado en al menos el 25% de las células del clúster (min.pct = 0.25).
- Se aplicó un umbral mínimo de cambio de expresión ( $\text{avg\_log2FC} \geq 0.25$  en escala log2), con el fin de priorizar señales con relevancia biológica y reducir falsos positivos (Hao et al., 2021; Stuart & Satija, 2019b).

El resultado principal de este análisis fue una tabla completa de genes marcadores por clúster, que se exportó en formato Excel para su inspección y uso posterior en las etapas de anotación celular.

Con el fin de facilitar la interpretación y visualización de los resultados, se generaron además dos subconjuntos derivados de la tabla completa:

- TOP2, que recoge los dos genes con mayor avg\_log2FC por clúster, proporcionando una visión sintética de la identidad dominante de cada agrupación celular.
- TOP10, que incluye los diez genes más representativos por clúster y se empleó para la generación de heatmaps de expresión génica.

Dado el elevado número de clústeres identificados, la visualización del heatmap global se realizó también en formato por bloques de clústeres, representando únicamente sus TOP10 genes característicos. Este enfoque permitió mejorar la claridad visual y facilitar la comparación entre firmas transcriptómicas sin perder información relevante.

Este análisis de genes característicos constituye un componente central del pipeline analítico, ya que proporciona la evidencia transcriptómica necesaria para (i) validar la coherencia del clustering, (ii) apoyar la anotación automática de poblaciones celulares y (iii) realizar una anotación manual informada y consistente con marcadores conocidos y con la literatura especializada.

### 3.5. Anotación automática de poblaciones celulares.

Una vez definidos los clústeres y caracterizados a partir de sus genes marcadores, se realizó una anotación automática preliminar de las poblaciones celulares con el objetivo de obtener una primera asignación objetiva y reproducible de identidades celulares, que sirviera como punto de partida para la posterior anotación manual.

Para ello se empleó ScType, una herramienta de anotación basada en firmas génicas específicas de tipo celular, diseñada para su integración con objetos Seurat y adaptada al análisis de datos de scRNA-seq. ScType asigna puntuaciones a cada célula en función del solapamiento entre su perfil de expresión y conjuntos de genes marcadores previamente definidos para distintos tipos celulares.

En este trabajo se utilizó una base de datos de marcadores específica de tejido prostático (Ianevski et al., 2022), cargada desde un fichero externo en formato Excel, que incluye genes característicos de poblaciones epiteliales, estromales, inmunes y vasculares descritas en próstata tanto en condiciones fisiológicas como patológicas. Esta personalización resulta especialmente relevante en tejidos con alta heterogeneidad celular, como la próstata, donde las firmas transcriptómicas pueden variar de forma significativa según el estado de diferenciación o progresión tumoral.

El procedimiento seguido fue el siguiente:

- Se calculó, para cada célula, un score de concordancia con cada tipo celular definido en la base de datos de ScType.
- A nivel de clúster, se agregaron los scores de todas las células pertenecientes al mismo clúster para obtener una puntuación global por tipo celular.
- Para cada clúster se seleccionaron las identidades con mayor puntuación, generando una tabla resumen con las diez etiquetas mejor puntuadas por clúster.
- Como criterio de control de calidad, cuando la puntuación máxima obtenida para un clúster fue inferior a una cuarta parte del número de células del clúster, la anotación automática se consideró poco fiable y se asignó la etiqueta “Unknown”.

La tabla final de anotaciones automáticas por clúster se exportó en formato Excel, permitiendo su inspección manual y su comparación directa con los genes marcadores identificados previamente. Posteriormente, la anotación automática seleccionada se incorporó al objeto Seurat como una nueva columna del metadato (CellType), asociando a cada célula la identidad asignada por ScType.

A partir de este objeto anotado automáticamente se generaron distintas representaciones UMAP:

- Un UMAP global, mostrando la distribución espacial de todas las poblaciones celulares.
- UMAPs estratificados por condición biológica (Healthy, PCa y CRPC), lo que permitió evaluar la presencia, distribución y posibles expansiones o contracciones de determinadas poblaciones a lo largo de la progresión tumoral.

Esta anotación automática proporcionó una visión inicial consistente de la organización celular del dataset integrado, aunque, dada la complejidad del tejido prostático y la plasticidad celular asociada al cáncer, se consideró necesaria una validación y refinamiento posterior mediante anotación manual basada en marcadores, firmas transcriptómicas y literatura especializada. Este proceso sentó las bases para la anotación manual definitiva de las poblaciones celulares, descrita en el apartado siguiente.

### 3.6. Estudio de poblaciones celulares y composición por condición biológica.

Con el objetivo de caracterizar la organización celular del tejido prostático a lo largo de la progresión tumoral, se realizó un análisis descriptivo de la composición de poblaciones celulares en función de la condición biológica (Healthy, PCa y CRPC). Este análisis permite evaluar la presencia relativa de cada población celular y detectar posibles expansiones, contracciones o pérdidas asociadas a los distintos estados de la enfermedad.

Para ello se empleó el objeto integrado y anotado, utilizando los metadatos correspondientes a la identidad celular asignada y a la condición biológica de cada célula. A partir de esta información se construyó una tabla resumen en la que, para cada combinación población–condición, se contabilizó el número absoluto de células y se calculó el porcentaje que representan dentro de cada población celular.

Estas tablas permiten evaluar qué fracción de cada población se encuentra asociada a tejido sano, tumor primario o enfermedad resistente a la castración, proporcionando una visión global de los cambios celulares que acompañan a la progresión del cáncer de próstata. Adicionalmente, se generó una tabla en formato ancho que incluye, para cada población celular, el número de células y el porcentaje correspondiente a cada condición, así como el número total de células por población, lo que facilita la comparación entre poblaciones y la identificación de aquellas más abundantes o específicas de una condición concreta.

Como complemento visual, se construyó un gráfico de barras apiladas, en el que cada barra representa una población celular y su altura corresponde al 100% de las células de dicha población, desglosadas por condición biológica. Este tipo de representación permite visualizar de forma intuitiva la contribución relativa de cada condición a cada población celular y detectar patrones de enriquecimiento o depleción asociados al estado tumoral.

Este análisis de composición celular constituye un paso clave previo al análisis de expresión diferencial, ya que permite contextualizar los resultados transcriptómicos posteriores, identificar poblaciones presentes en todas las condiciones y detectar aquellas restringidas o enriquecidas en estados específicos, aspectos fundamentales para la correcta interpretación biológica de los cambios moleculares observados.

### 3.7. Anotación manual de poblaciones celulares.

Tras la anotación automática preliminar mediante ScType, se llevó a cabo un proceso de anotación manual de los clústeres, con el objetivo de validar, refinar y consolidar la asignación de identidades celulares. Esta fase combina la evidencia computacional con la interpretación biológica de los perfiles de expresión génica, permitiendo corregir asignaciones ambiguas y definir subpoblaciones celulares de forma más precisa, especialmente en un tejido altamente heterogéneo como la próstata.

Para ello se partió del objeto integrado y anotado automáticamente y se construyó un diccionario de mapeo clúster-tipo celular, en el que a cada clúster se le asignó una identidad celular basada en la integración de múltiples fuentes de información:

- (i) los genes marcadores característicos obtenidos mediante el análisis de expresión diferencial por clúster (FindAllMarkers) (Stuart et al., 2019c);
- (ii) las firmas transcriptómicas propuestas por ScType (Ianevski et al., 2022); y
- (iii) la literatura especializada sobre poblaciones epiteliales, inmunes y estromales del tejido prostático en condiciones fisiológicas y tumorales (Chen et al., 2021; Henry et al., 2018; Karthaus et al., 2020).

Este mapeo permitió identificar distintos subtipos de células luminales (benignas, metabólicas, secretoras, mucinosas, CRISP3-high y PCA3-high), así como poblaciones basales, células inmunes (linfocitos T y B, macrófagos y mastocitos), células vasculares (endoteliales y pericitos), fibroblastos asociados a cáncer (CAFs), células proliferativas y una población con características neuroendocrinas/EMT-like, descrita previamente como relevante en la progresión hacia el cáncer de próstata resistente a la castración.

El diccionario de anotación manual se aplicó al metadato del objeto Seurat, añadiendo una nueva columna que asigna a cada célula su identidad celular manual en función del clúster al que pertenece. Para garantizar la completitud del proceso, cualquier clúster que no pudiera asignarse de forma fiable se etiquetó como Unknown.

A partir del objeto con anotación manual se generaron representaciones UMAP (Becht et al., 2018), tanto globales como estratificadas por condición biológica (Healthy, PCa y CRPC). Estas visualizaciones permitieron evaluar la organización espacial de las

poblaciones celulares, la coherencia interna de los clústeres y su distribución relativa entre condiciones, así como explorar la diversidad del compartimento epitelial y la presencia de poblaciones asociadas a estados tumorales avanzados (Wu et al., 2021).

### 3.8. Estudio de poblaciones y composición celular por condición biológica.

Con el objetivo de caracterizar los cambios en la composición celular asociados a la progresión del cáncer de próstata, se llevó a cabo un estudio descriptivo de las poblaciones celulares identificadas, analizando su distribución relativa entre las distintas condiciones biológicas consideradas (Healthy, PCa y CRPC).

Este análisis se realizó sobre el objeto integrado y anotado, utilizando tanto la información de clúster como la anotación manual de tipos celulares (CellType\_Manual). A partir del metadato de cada célula se seleccionaron las variables correspondientes al tipo celular y a la condición biológica, construyéndose tablas resumen que permiten cuantificar la abundancia y proporción relativa de cada población.

En una primera aproximación, se calculó para cada combinación tipo celular–condición el número absoluto de células y el porcentaje que representan dentro de cada tipo celular. Este enfoque permite responder a la pregunta de qué fracción de una población celular concreta procede de tejido sano, tumor primario o cáncer resistente a la castración, y facilita la identificación de poblaciones enriquecidas o depletadas a lo largo de la progresión tumoral.

Adicionalmente, se generaron tablas en formato ancho que incluyen, para cada tipo celular, el número de células y el porcentaje correspondiente a cada condición, así como el número total de células por población. Estas tablas se exportaron en formato Excel para su inspección detallada y su uso posterior en la interpretación biológica de los resultados.

Para facilitar la visualización de estos patrones, se construyeron gráficos de barras apiladas con dos enfoques complementarios:

- Distribución por condición dentro de cada tipo celular, mostrando el porcentaje de células de cada condición biológica que contribuyen a una población concreta.
- Composición celular por condición, representando el porcentaje que cada tipo celular aporta al conjunto de células de cada condición biológica.

Estos gráficos permiten evaluar de forma intuitiva los cambios en la arquitectura celular del tejido prostático a lo largo de la progresión desde tejido sano hasta cáncer resistente a la castración, y constituyen una base fundamental para contextualizar los análisis posteriores de expresión diferencial.

### 3.9. Análisis de expresión diferencial por tipo celular entre condiciones biológicas.

Con el objetivo de caracterizar los cambios transcriptómicos asociados a la progresión tumoral en cada población celular, se realizó un análisis de expresión diferencial (DEGs) estratificado por tipo celular, comparando las distintas condiciones biológicas: tejido sano (Healthy), cáncer de próstata localizado (PCa) y cáncer de próstata resistente a la castración (CRPC).

El análisis se llevó a cabo sobre el objeto integrado con anotación manual, utilizando la identidad celular definida en el metadato CellType\_Manual para asegurar que las comparaciones se realizasen exclusivamente entre células del mismo tipo. De este modo, se evita la confusión entre diferencias de composición celular y cambios de expresión genuinos dentro de una misma población.

Para cada tipo celular, se realizaron las siguientes comparaciones pareadas:

- Healthy vs PCa.
- Healthy vs CRPC.
- PCa vs CRPC.

El análisis se efectuó empleando el assay SCT del objeto Seurat, que proporciona valores de expresión corregidos por profundidad de secuenciación y otros efectos técnicos mediante un modelo de regresión binomial negativa regularizada. Previamente al cálculo de DEGs, el assay SCT se preparó para análisis diferencial mediante la función PrepSCTFindMarkers(), con el fin de armonizar los modelos internos y garantizar la comparabilidad entre células procedentes de distintas muestras.

La identificación de genes diferencialmente expresados se realizó utilizando la función FindMarkers() de Seurat, aplicando el test no paramétrico de Wilcoxon rank-sum, ampliamente utilizado en análisis de scRNA-seq. Para garantizar la robustez de los resultados, se aplicaron los siguientes criterios de filtrado:

- Se exigió un número mínimo de células por condición dentro de cada tipo celular ( $\geq 25$  células), descartando automáticamente aquellas comparaciones con tamaño muestral insuficiente.
- Se consideraron únicamente genes expresados en al menos un 10% de las células del grupo comparado (min.pct = 0.10).
- Se aplicó un umbral mínimo de cambio de expresión de  $|\log_{2}FC| \geq 0.25$ , priorizando genes con relevancia biológica.
- Se utilizaron valores de p ajustados por múltiples comparaciones (p\_val\_adj) para la evaluación de significación estadística.

En aquellos casos en los que el assay SCT contenía múltiples modelos internos con tamaños de librería desiguales —una situación habitual en objetos integrados— se desactivó la recorrección de UMIs (recorrect\_umi = FALSE), siguiendo las recomendaciones del propio framework Seurat para evitar errores técnicos sin comprometer la validez del contraste estadístico.

Los resultados del análisis de expresión diferencial se organizaron en tablas independientes por comparación, exportadas en formato Excel, incluyendo una hoja por tipo celular. Cada tabla contiene, para cada gen identificado, el valor de cambio de expresión promedio (avg\_log2FC), la proporción de células expresoras en cada grupo, y los valores de significación estadística ajustados.

Este enfoque permitió identificar firmas transcriptómicas específicas asociadas a la progresión tumoral dentro de cada población celular, proporcionando una base sólida para la interpretación biológica de los cambios observados entre tejido sano, cáncer localizado y enfermedad avanzada resistente a la castración.

### 3.10. Selección de poblaciones celulares para el análisis comparativo.

El análisis de expresión diferencial en datos de scRNA-seq se realizó siguiendo las buenas prácticas recomendadas en la bibliografía, incluyendo una normalización adecuada, el control de calidad previo y la selección de poblaciones celulares con representación suficiente en las condiciones comparadas. Estos pasos son considerados fundamentales para garantizar resultados robustos y reproducibles en estudios unicelulares (Luecken & Theis, 2019). Asimismo, trabajos recientes destacan la importancia de minimizar falsos positivos y de considerar adecuadamente la variabilidad biológica y técnica inherente a los datos de scRNA-seq (Heumos et al., 2023; Squair et al., 2021).

Dado que el análisis de expresión diferencial entre condiciones biológicas requiere la comparación de poblaciones celulares equivalentes, se llevó a cabo una selección previa de los tipos celulares aptos para el análisis comparativo. Para ello, se evaluó la abundancia de cada población celular anotada manualmente en las tres condiciones estudiadas (Healthy, PCa y CRPC).

Se estableció un umbral mínimo de células por condición, de modo que únicamente se consideraron aquellas poblaciones que presentaban una representación suficiente y equilibrada en las tres condiciones biológicas. Este criterio permitió garantizar una potencia estadística adecuada y evitar comparaciones sesgadas o poco robustas derivadas de tamaños muestrales reducidos, tal y como recomiendan estudios metodológicos en análisis de scRNA-seq y diseño experimental en RNA-seq (Heumos et al., 2023; Luecken & Theis, 2019; Schurch et al., 2016; Squair et al., 2021).

Como resultado de este filtrado, el análisis de expresión diferencial se restringió a un subconjunto de poblaciones celulares presentes de forma consistente a lo largo de toda la progresión tumoral, incluyendo poblaciones epiteliales luminas y componentes clave del microambiente tumoral, como macrófagos, células T y células endoteliales.

Las poblaciones ausentes o escasamente representadas en alguna de las condiciones (por ejemplo, fibroblastos asociados a cáncer, poblaciones luminas altamente especializadas o subpoblaciones características de estadios avanzados) no se incluyeron en el análisis de expresión diferencial comparativo. En estos casos, su relevancia biológica se abordó mediante el análisis de composición celular y la descripción de su aparición o

expansión específica en determinadas condiciones, en lugar de mediante comparaciones transcriptómicas directas.

Este enfoque permite diferenciar de forma clara entre cambios asociados a la emergencia o desaparición de poblaciones celulares y aquellos debidos a alteraciones transcriptómicas dentro de una misma identidad celular, siguiendo las recomendaciones actuales para el análisis riguroso de datos de scRNA-seq.

### 3.11. Visualización mediante volcano plots.

El análisis de expresión diferencial entre condiciones biológicas se realizó a nivel de tipo celular, comparando exclusivamente poblaciones celulares equivalentes y con representación suficiente en todas las condiciones estudiadas (Healthy, PCa y CRPC). Esta selección previa permitió garantizar una potencia estadística adecuada y evitar comparaciones sesgadas derivadas de tamaños muestrales reducidos, siguiendo las recomendaciones actuales para el análisis de datos de scRNA-seq (Heumos et al., 2023; Luecken & Theis, 2019; Squair et al., 2021).

Para cada población celular seleccionada, se llevaron a cabo comparaciones pareadas entre condiciones (Healthy vs PCa, Healthy vs CRPC y PCa vs CRPC) mediante análisis de expresión diferencial, obteniendo listas de genes diferencialmente expresados caracterizadas por su cambio medio de expresión (avg\_log2FC) y su significación estadística ajustada por múltiples contrastes (p\_val\_adj), de acuerdo con el flujo de trabajo estándar propuesto en Seurat (Stuart & Satija, 2019b).

Con el objetivo de facilitar la interpretación visual de los resultados, los genes diferencialmente expresados se representaron mediante volcano plots, en los que el eje X corresponde al cambio de expresión en escala log2 (avg\_log2FC) y el eje Y a la significación estadística ajustada expresada como  $-\log_{10}(p\_val\_adj)$ . Este tipo de representación permite identificar de forma simultánea genes con cambios de expresión relevantes y alta significación estadística, constituyendo una herramienta ampliamente utilizada en estudios de transcriptómica unicelular y bulk RNA-seq (Luecken & Theis, 2019).

Se aplicaron umbrales estrictos para definir genes diferencialmente expresados significativos, considerando únicamente aquellos con  $p\_val\_adj < 0.05$  y un cambio absoluto de expresión  $\geq 0.25$  en escala log2, en línea con criterios habituales en análisis de scRNA-seq (Squair et al., 2021; Stuart & Satija, 2019b). Los genes que cumplían simultáneamente ambos criterios se destacaron en los volcano plots, y se seleccionaron los genes más significativos para su anotación directa en la figura, facilitando la identificación de firmas transcriptómicas relevantes.

Este enfoque de visualización permitió comparar de manera coherente los cambios transcriptómicos asociados a la progresión tumoral dentro de un mismo tipo celular, diferenciando alteraciones específicas de la identidad celular de aquellas debidas a cambios en la composición poblacional. La interpretación de los resultados se realizó teniendo en cuenta tanto la significación estadística como la plausibilidad biológica de los genes identificados, siguiendo buenas prácticas para minimizar falsos positivos y

sobreinterpretaciones en datos de scRNA-seq. Dado el elevado número de comparaciones y poblaciones celulares analizadas, no se generaron heatmaps globales de DEGs, priorizando una interpretación dirigida a genes relevantes desde el punto de vista biológico (Heumos et al., 2023; Luecken & Theis, 2019; Squair et al., 2021).

### 3.12. Visualización de genes diferencialmente expresados mediante heatmaps.

Con el objetivo de profundizar en la interpretación de los cambios transcriptómicos detectados mediante el análisis de expresión diferencial, se realizó una visualización adicional de genes diferencialmente expresados seleccionados mediante mapas de calor (heatmaps). Este tipo de representación permite evaluar patrones de expresión relativa entre condiciones dentro de un mismo tipo celular, facilitando la identificación de firmas génicas coherentes y de posibles subprogramas transcripcionales asociados a la progresión tumoral.

Dado que la generación de heatmaps globales con todos los genes diferencialmente expresados puede resultar difícil de interpretar y propensa a artefactos en datasets unicelulares de gran tamaño, se adoptó un enfoque dirigido, siguiendo recomendaciones metodológicas recientes (Heumos et al., 2023; Luecken & Theis, 2019). Para cada comparación y tipo celular seleccionado, se extrajo un subconjunto de genes diferencialmente expresados significativos, priorizando aquellos con mayor magnitud de cambio de expresión y relevancia estadística.

En concreto, para cada población celular analizada y para cada comparación entre condiciones (Healthy vs PCa, Healthy vs CRPC y PCa vs CRPC), se seleccionaron:

- Los 15 genes más sobreexpresados (upregulated) en una condición respecto a la otra.
- Los 15 genes más infraexpresados (downregulated).

La selección se realizó ordenando los genes diferencialmente expresados en función del valor absoluto del cambio de expresión promedio ( $|avg\_log2FC|$ ), aplicando los mismos umbrales de significación estadística definidos previamente ( $p\_val\_adj < 0.05$  y  $|log2FC| \geq 0.25$ ). Este criterio permite centrar la visualización en genes con efectos biológicamente relevantes, reduciendo el ruido asociado a cambios marginales.

Los heatmaps se construyeron utilizando valores de expresión normalizados del assay SCT, escalados por gen para resaltar diferencias relativas entre células y condiciones. La visualización se realizó a nivel de célula o agregada por condición según el tipo celular, con el fin de preservar la estructura interna de la población analizada. Este enfoque permite identificar patrones de coexpresión y contrastar visualmente las diferencias transcriptómicas entre estados biológicos dentro de una misma identidad celular.

La generación de heatmaps dirigidos complementa los volcano plots, proporcionando una visión más integrada de los cambios transcriptómicos y facilitando la interpretación funcional de los genes diferencialmente expresados más relevantes.

### 3.13. Organización y exportación de resultados de expresión diferencial.

Con el fin de garantizar la trazabilidad, reproducibilidad y facilidad de exploración de los resultados obtenidos, los datos derivados del análisis de expresión diferencial se organizaron y exportaron sistemáticamente en formato tabular.

Para cada comparación entre condiciones biológicas (Healthy vs PCa, Healthy vs CRPC y PCa vs CRPC), se generó un archivo Excel independiente que contiene una hoja por tipo celular analizado. En cada hoja se incluyó la lista completa de genes evaluados en la comparación correspondiente, junto con las métricas estándar del análisis de expresión diferencial, incluyendo:

- Cambio medio de expresión (avg\_log2FC).
- Proporción de células expresoras en cada grupo comparado.
- Valor de significación estadística sin ajustar y ajustado por múltiples comparaciones (p\_val y p\_val\_adj).

Adicionalmente, se generaron tablas resumen derivadas que recogen, para cada tipo celular y comparación:

- El subconjunto de genes diferencialmente expresados significativos.
- La clasificación de genes como sobreexpresados o infraexpresados.
- La selección de los genes más relevantes (top upregulated y downregulated) utilizados en las visualizaciones mediante heatmaps.

En aquellos casos en los que una población celular no presentó genes diferencialmente expresados que cumpliesen los criterios de significación establecidos, se generó igualmente una hoja específica indicando explícitamente la ausencia de DEGs significativos. Este procedimiento evita ambigüedades en la interpretación de resultados y permite distinguir entre ausencia de señal biológica y ausencia de análisis.

La exportación sistemática de resultados en formato Excel facilita su inspección manual, su integración con herramientas externas de análisis funcional (por ejemplo, enriquecimiento de rutas o anotación funcional) y su reutilización en fases posteriores del trabajo o en estudios futuros. Este enfoque está alineado con las buenas prácticas recomendadas para el análisis reproducible de datos de scRNA-seq (Heumos et al., 2023; Luecken & Theis, 2019).

El conjunto de métodos descritos establece un pipeline reproducible y coherente para el análisis de datos de scRNA-seq, permitiendo abordar tanto cambios en la composición celular como alteraciones transcriptómicas específicas asociadas a la progresión del cáncer de próstata.

## 4. Resultados.

### 4.1. Control de calidad (QC).

Antes del análisis principal, se aplicaron criterios de control de calidad para garantizar la fiabilidad de los datos transcriptómicos a nivel celular. Se evaluaron métricas estándar como el número de genes detectados por célula, el número total de cuentas (UMIs), y el porcentaje de expresión de genes mitocondriales, cuyos valores se visualizan mediante boxplots e histogramas (ver Anexo: [9.2. Control de calidad de las muestras \(QC\)](#)).

El filtrado eliminó aquellas células con valores atípicos extremos en dichas métricas, lo que permitió retener un total de 28.919 células individuales, es decir, un 91,5% del total inicial, distribuidas de la siguiente manera:

- Healthy: 6.748 células retenidas (91,7%)
- PCa: 4.904 células retenidas (98,1%)
- CRPC: 17.267 células retenidas (90,4%)

Tal como se observa en la Tabla 3, las pérdidas fueron algo mayores en la muestra CRPC\_1, posiblemente debido a artefactos de disociación o mayor fragilidad del tejido tumoral avanzado. En todos los casos, se conservaron suficientes células por muestra para permitir análisis estadísticos robustos.

Tabla 3: Resultados tras el control de calidad.

| Sample    | Cells_before | Cells_after | Percent_kept | nFeature_p1 | nFeature_p99 | max_percent_mt_used | MALAT1_p99 | Ribo_p99 |
|-----------|--------------|-------------|--------------|-------------|--------------|---------------------|------------|----------|
| HEALTHY_1 | 2825         | 2716        | 96,14        | 1258,68     | 6699         | 8,59                | 1119,04    | 51,19    |
| HEALTHY_2 | 2303         | 2228        | 96,74        | 1002,02     | 6119,04      | 15                  | 963,64     | 49,65    |
| HEALTHY_3 | 1600         | 1461        | 91,31        | 1008,96     | 6412,73      | 15                  | 1457,26    | 0        |
| PCA       | 5000         | 4904        | 98,08        | 274         | 5053,02      | 0                   | 1728,19    | 60,27    |
| CRPC_1    | 4078         | 3188        | 78,18        | 155         | 8005,92      | 15                  | 2685,3     | 39,98    |
| CRPC_2    | 8084         | 7276        | 90           | 337,83      | 5194,87      | 15                  | 1260,17    | 34,88    |
| CRPC_3    | 7716         | 7146        | 92,61        | 295,15      | 5736,4       | 15                  | 1112,7     | 40,16    |

En la muestra PCa, el porcentaje de expresión mitocondrial resultó ser del 0%, ya que estos genes fueron previamente filtrados por los autores del dataset original, con el fin de reducir el ruido técnico.

### 4.2. Integración y agrupamiento celular.

Tras el filtrado de calidad, las células retenidas fueron integradas en un único objeto Seurat para su análisis conjunto. La integración permitió armonizar los datos procedentes de distintas condiciones (Healthy, PCa y CRPC) y minimizar efectos de batch. Posteriormente, se aplicaron técnicas de reducción dimensional para representar la variabilidad global de los datos.

La selección del número óptimo de componentes principales (PCs) se basó en dos enfoques complementarios: la varianza acumulada explicada y el elbow plot de la desviación estándar. Las primeras 20 PCs capturaron más del 80% de la varianza total

(Figura 2), y el elbow plot mostró una inflexión clara en ese mismo punto, tras el cual el aporte marginal de nuevas componentes disminuye notablemente (Figura 3).

Estas 20 PCs fueron utilizadas para realizar el agrupamiento de las células mediante algoritmos de clustering, y su posterior representación en dos dimensiones con UMAP, que permitió identificar visualmente distintas poblaciones celulares.

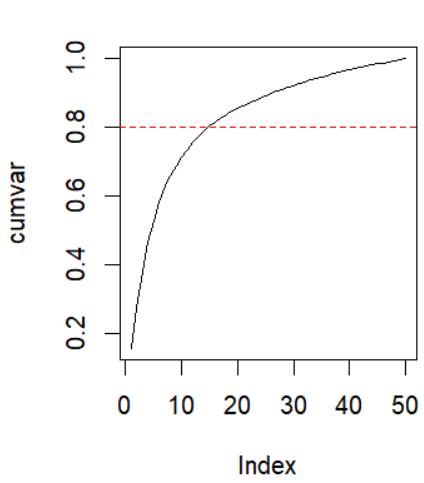


Figura 2: Varianza acumulada explicada.

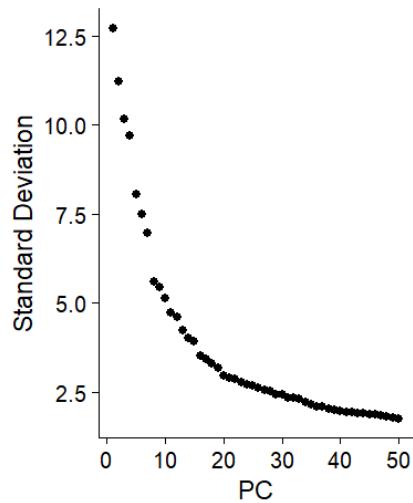
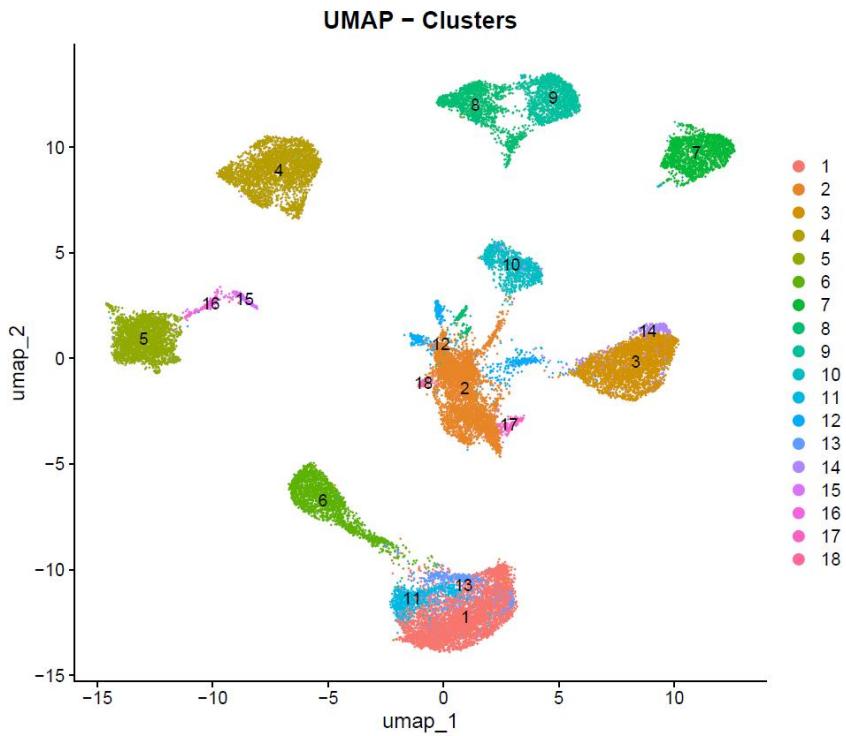


Figura 3: Elbow plot.

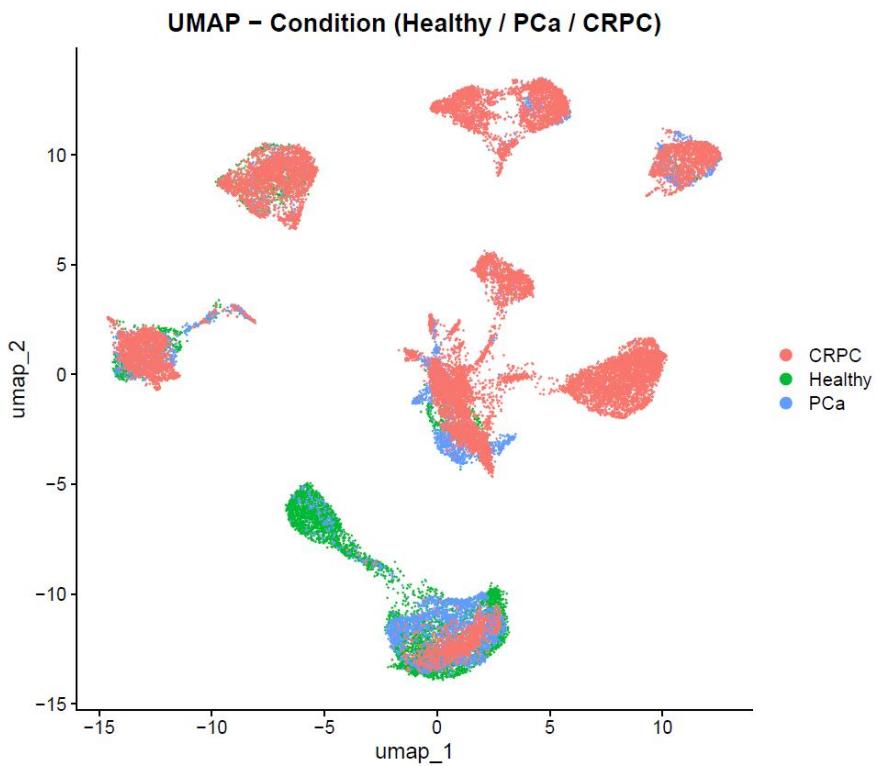
Se evaluaron diferentes valores de resolución para el algoritmo de clustering Louvain (0.2, 0.3 y 0.5). Se seleccionó una resolución de 0.5 por ofrecer un equilibrio óptimo entre granularidad biológica y estabilidad estructural de los clústeres (ver Anexo: [9.3. Decisión de resolución de clústeres](#)). A esta resolución, el análisis UMAP permitió identificar 18 clústeres celulares bien definidos (Figura 3).

La visualización por condición clínica reveló que los distintos estados (Healthy, PCa y CRPC) se distribuyen de forma diferenciada en el espacio transcripcional, aunque con solapamiento parcial en algunos clústeres (Figura 4).

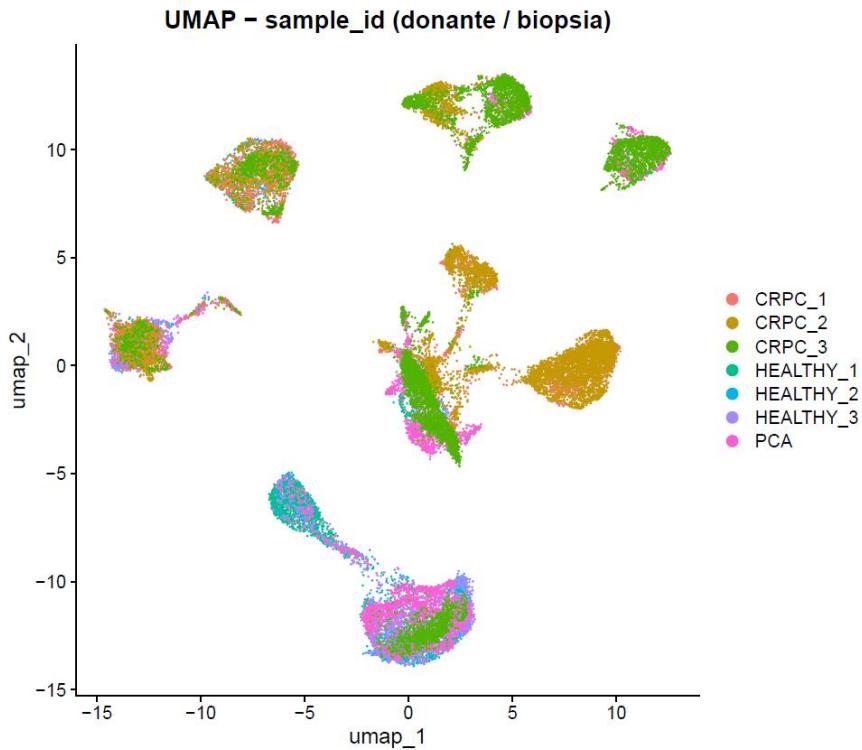
Por otro lado, el análisis por muestra individual mostró que no existen agrupamientos espurios por donante, lo que sugiere una integración efectiva y ausencia de efecto batch residual (Figura 5).



*Figura 4: Representaciones UMAP tras la integración de datos.*



*Figura 5: Distribución de los clústeres según condición.*



*Figura 6: Distribución por muestra individual.*

#### 4.3. Identificación de genes característicos por clúster.

Con el fin de caracterizar transcriptómicamente los clústeres celulares obtenidos y evaluar la coherencia biológica del agrupamiento, se analizaron los genes característicos enriquecidos en cada clúster. Este análisis permitió identificar firmas de expresión específicas asociadas a cada agrupación celular y sirvió como paso previo a la anotación de tipos celulares.

La visualización de los 10 genes más representativos por clúster mediante un heatmap (Figura 6 a Figura 10) mostró patrones de expresión claramente diferenciados entre los 18 clústeres. En esta representación, el color amarillo indica niveles elevados de expresión relativa, mientras que el lila refleja una expresión baja o ausente, permitiendo identificar de forma intuitiva la especificidad de cada firma transcriptómica.

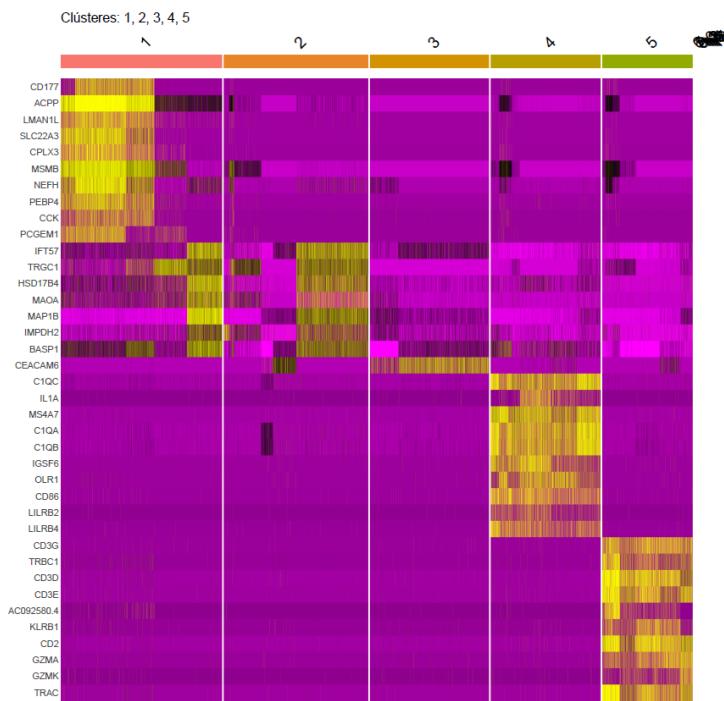


Figura 7: Heatmap de los 10 genes más representativos del clúster 1 al 5.

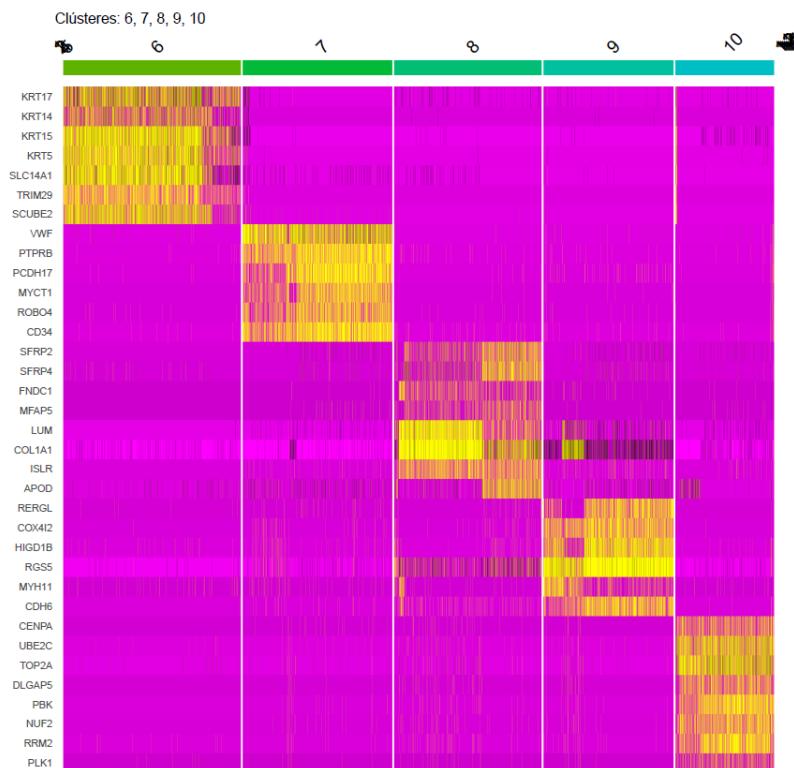


Figura 8: Heatmap de los clústeres 6 al 10.



Figura 9: Heatmap de los clústeres 11 al 15.



Figura 10: Heatmap de los clústeres 16 al 18.

Además de la obtención del heatmap, se generaron tres tablas complementarias derivadas del análisis completo:

- La tabla completa de genes marcadores por clúster.
- Un subconjunto con los 10 genes más representativos por clúster (TOP10).
- Un resumen de los 2 genes con mayor logFC por clúster (TOP2), útiles para la interpretación biológica.

La tabla correspondiente al conjunto TOP10 se incluye en el Anexo: [9.4. TOP10 genes por clúster](#), para su consulta detallada.

#### 4.4. Anotación automática de poblaciones celulares.

Se obtuvieron identidades celulares preliminares para cada clúster mediante anotación automática con ScType. Esta asignación sirvió como punto de partida para evaluar la coherencia del agrupamiento y guiar la posterior anotación manual.

El UMAP global anotado (Figura 11) reveló una diversidad de poblaciones celulares, incluyendo componentes epiteliales diferenciados (Epithelial Basal, Epithelial Hillock, Epithelial Luminal), poblaciones inmunes (CTL, CD8+ effector, B cells, Macrophages, Mast cells y PDC), poblaciones estromales (Fibroblasts y Pericytes) y vasculares (Endothelial cells). Además, se identifica una población clasificada como Tumor, junto con otras etiquetadas como Unknown, ya que no alcanzan puntuación suficiente en ninguna de las identidades propuestas.

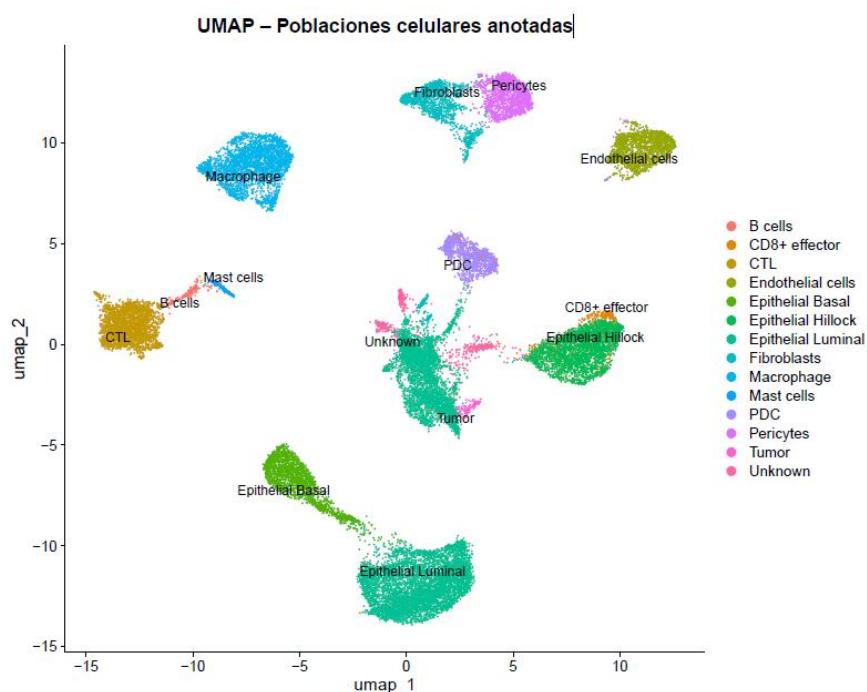


Figura 11: UMAP global con anotación automática de poblaciones celulares.

La representación UMAP por condición mostró como varía la distribución de las poblaciones celulares a lo largo de la progresión tumoral:

- En Healthy (Figura 12) se observa un predominio de las poblaciones epiteliales luminales y basales, junto con componentes inmunes (CTL, B cells, Macrophages) y algunas células endoteliales y estromales, específicamente Pericytes. La proporción de células Unknown fue baja.

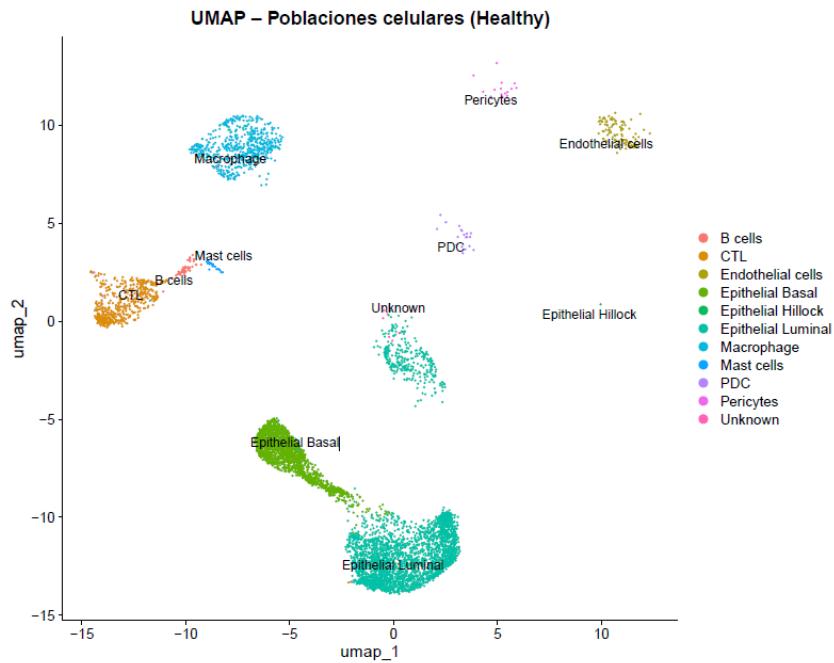


Figura 12: UMAP anotado - Condición Healthy.

- En PCa (Figura 13) aparecen poblaciones tumorales con una distribución diversa de células epiteliales, inmunes y vasculares.

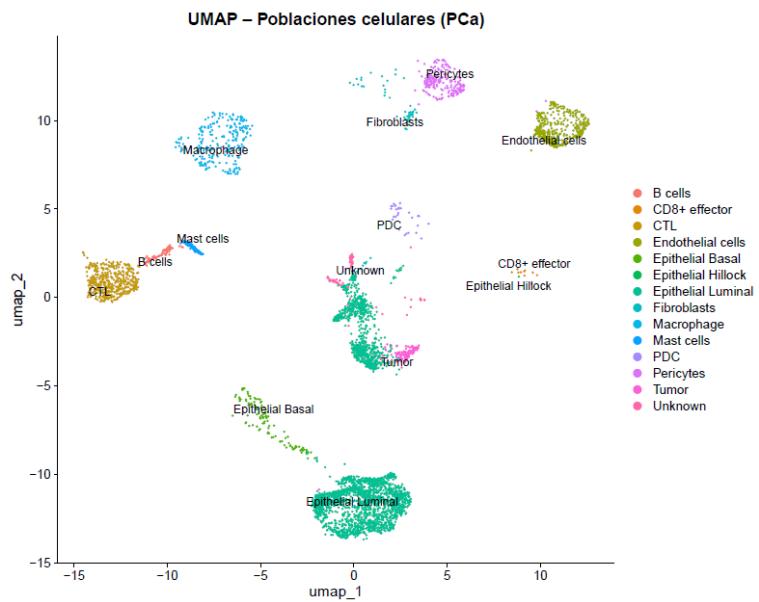


Figura 13: UMAP anotado - Condición PCa.

- En CRPC (Figura 14) se observó una ausencia de células anotadas como Tumor por ScType, con una expansión de Fibroblasts, Pericytes, CD8+ effectors y un incremento notable de células Unknown.

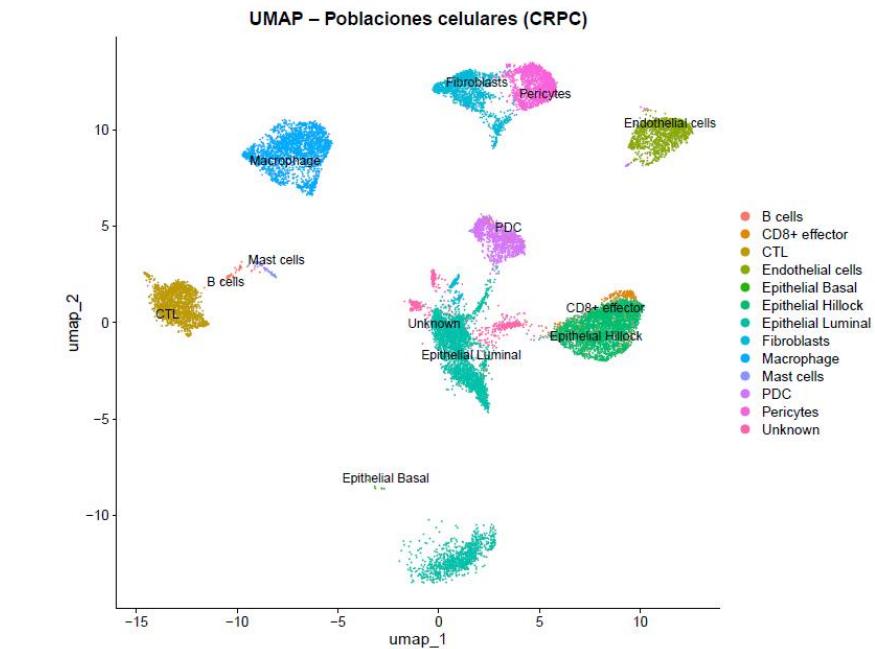


Figura 14: UMAP anotado - Condición CRPC.

Esta anotación preliminar sirvió como base para el análisis crítico de cada clúster y su posterior validación mediante anotación manual.

## 4.5. Composición celular por condición biológica de los clústeres.

Con el objetivo de explorar cómo varía la organización celular a lo largo de la progresión del cáncer de próstata, se analizó la distribución de células en los distintos clústeres según su condición biológica. La representación mediante un gráfico de barras apiladas (Figura 15) muestra la composición relativa de cada clúster en función de la condición de origen (Healthy, PCa y CRPC).

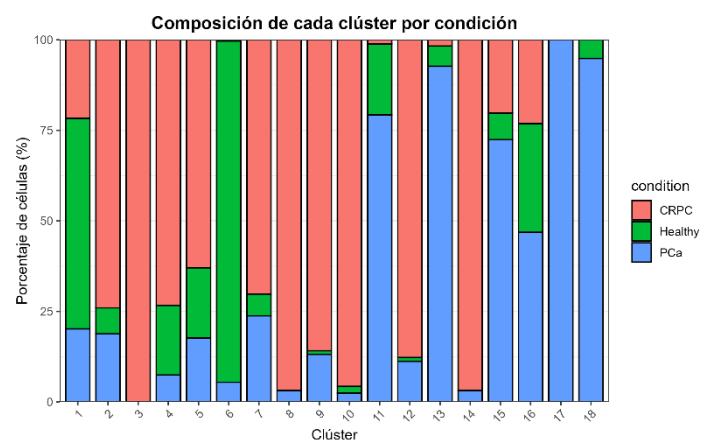


Figura 15: Composición relativa de cada clúster según la condición biológica.

Los clústeres 2 al 5, 6 al 10, 12 y 14 presentan una mayor proporción de células provenientes de muestras CRPC, siendo el clúster 3 exclusivo de esta condición. Los clústeres 11, 13 y del 15 al 18 están compuestos mayoritariamente por células de PCa, siendo exclusivo el clúster 17. En cuanto a las muestras Healthy, su representación es mayoritaria en los clústeres 1 y 6, y se observa en menor proporción en la mayoría de los clústeres, con excepción de los clústeres 3 y 17, donde no se detecta.

Este análisis resume la distribución relativa de células por condición en cada clúster, proporcionando un marco descriptivo para los análisis transcriptómicos posteriores.

## 4.6. Composición de poblaciones celulares anotadas manualmente.

Tras la anotación automática inicial, se realizó la anotación manual de los clústeres, el resultado de la cual se muestra en la Figura 16.

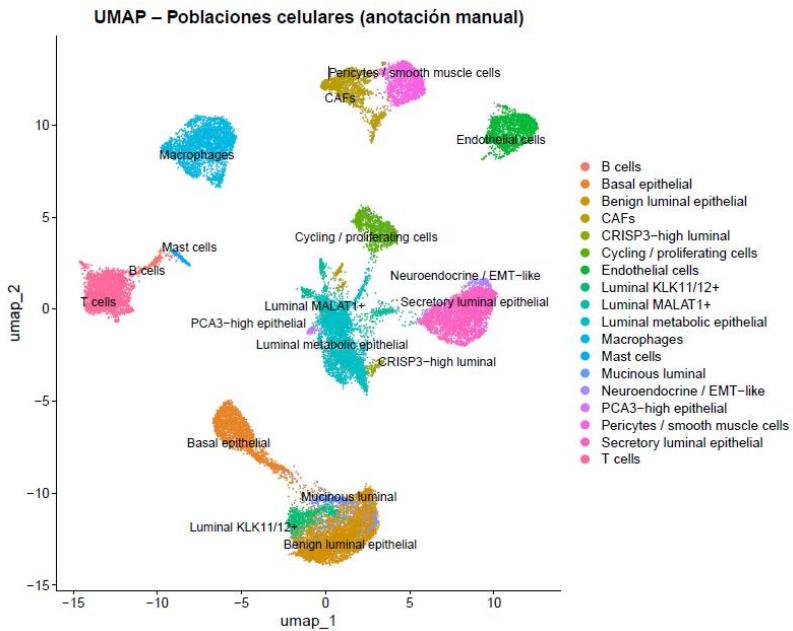


Figura 16: UMAP global con anotación manual de tipos celulares.

La distribución espacial de las poblaciones revela una clara segregación entre compartimentos epiteliales (luminales benignos, tumorales, basales), células del microambiente (inmunes, endoteliales, CAFs, pericitos) y células proliferativas. Destaca la localización central de los subtipos luminas tumorales (KLK11/12+, MALAT1+, mucinosas, CRISP3-high y PCA3-high), así como la agrupación periférica de poblaciones inmunes (macrófagos, linfocitos T y B, mastocitos) y vasculares.

Las visualizaciones estratificadas por condición biológica (Figura 17 – 19) muestran diferencias claras en la representación de tipos celulares entre los estados Healthy, PCa y CRPC. En tejido sano predominan las condiciones Basal, Endothelial, Benign luminal y células inmunes. En PCa se observa una expansión de poblaciones tumorales específicas (KLK11/12+, Mucinosas, CRISP3-high, PCA3-high), mientras que en CRPC estas se ven reemplazadas o complementadas por subtipos más agresivos, como MALAT1+ y Neuroendocrine / EMT-like, junto con un aumento de CAFs y células en proliferación.

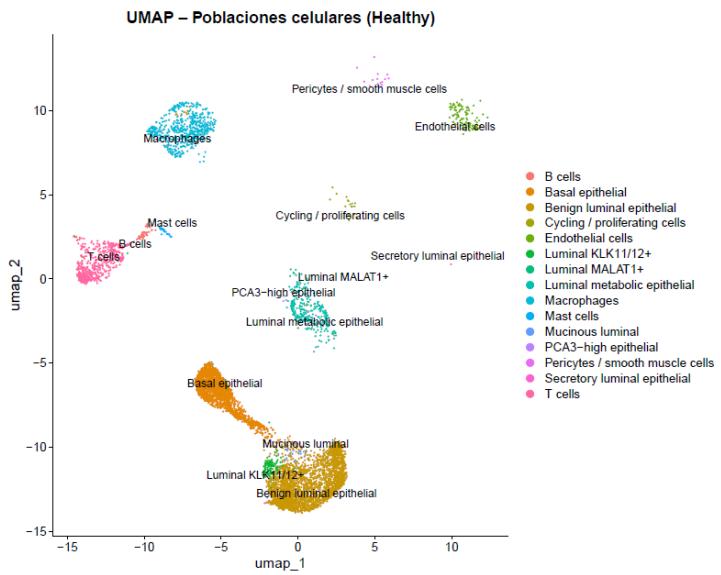


Figura 17: UMAP para Healthy.

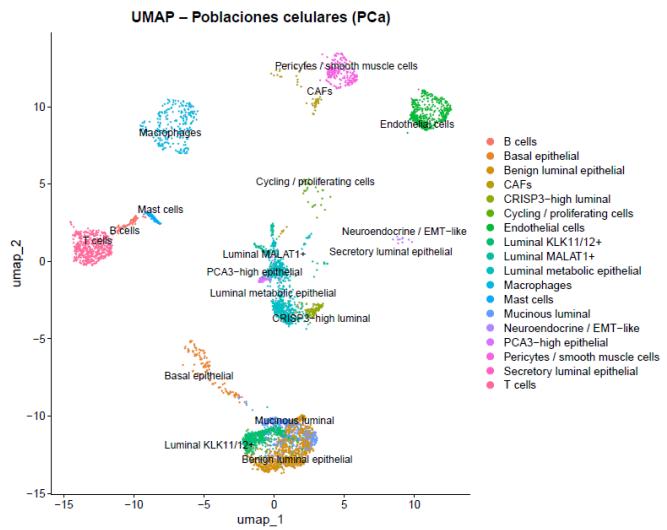


Figura 18: UMAP para PCa.

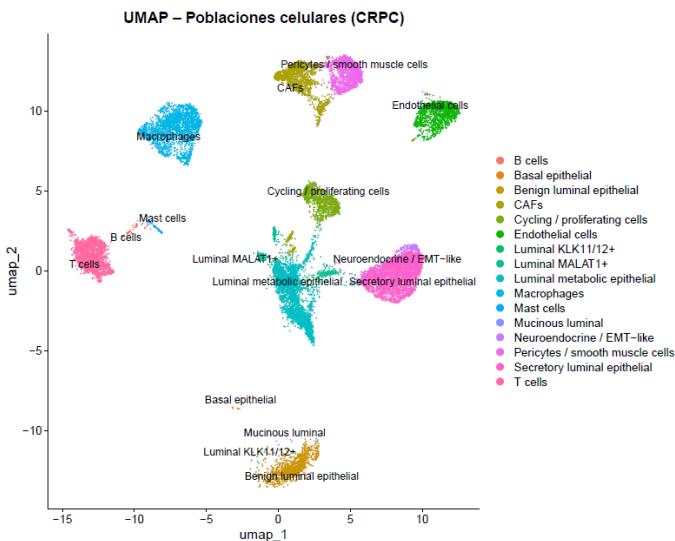


Figura 19: UMAP para CRPC.

A pesar del filtrado aplicado para eliminar células con características transcriptómicas de baja calidad, se identificó una subpoblación etiquetada como “Luminal MALAT1+”, definida principalmente por la alta expresión de los lncRNAs MALAT1 y MIAT.

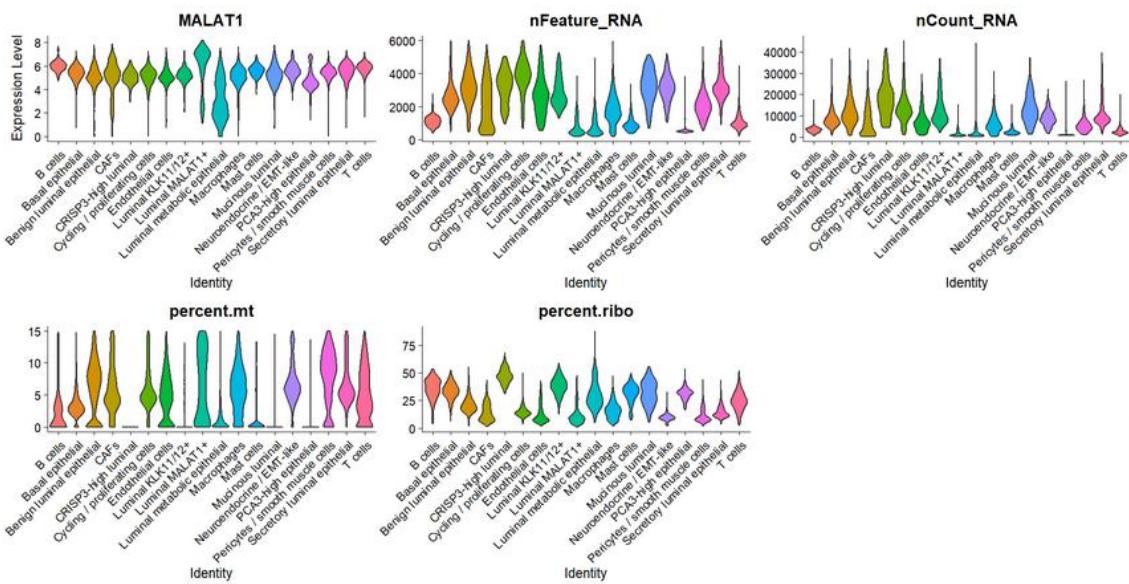


Figura 20: Métricas de calidad por poblaciones celulares.

Tal como se muestra en la Figura 20, esta población presenta niveles elevados de expresión de MALAT1, lo cual justifica su clasificación como clúster independiente. No obstante, los valores de percent.mt y percent.ribo se mantienen dentro de los márgenes aceptables, y tanto el número de genes (nFeature\_RNA) como el número total de cuentas (nCount\_RNA) se encuentran dentro del rango de las demás poblaciones epiteliales, por lo que no fue eliminada durante el filtrado inicial.

Estos datos sugieren que se trata de una subpoblación transcripcionalmente activa, más que de un artefacto técnico. MALAT1 es un lncRNA nuclear involucrado en la regulación de la expresión génica, ampliamente descrito en tejidos epiteliales y contextos tumorales. Su

expresión elevada no implica necesariamente mala calidad celular, sino un estado funcional específico, lo cual ha sido reportado en otros estudios de cáncer de próstata (Stone, 2017; D. Xu et al., 2024; D. Zhang et al., 2021a).

La composición celular por condición se muestra en la Figura 20. Se observa que en las muestras sanas predominan las células epiteliales benignas (Benign luminal epithelial) y Basal epitelial, junto con proporciones relevantes de T cells, Endothelial cells y Macrophages. En PCa se observa un aumento de subtipos tumorales epiteliales, especialmente KLK11/12+, Mucinosas, CRIPS3-high y PCA3-high, con una reducción relativa de las poblaciones benignas. En CRPC aparecen subtipos como Luminal MALAT1+ y Neuroendocrine / EMT-like, así como una expansión de Cycling / proliferating cells y de CAFs.

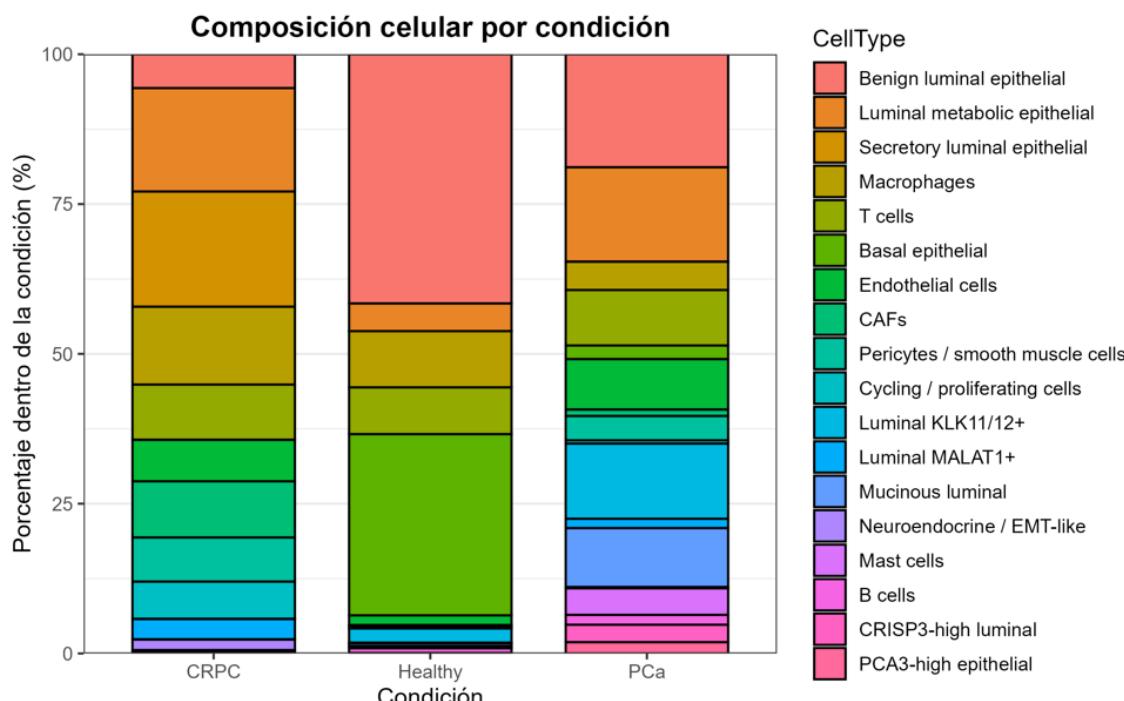


Figura 21: Composición celular por condición.

La Figura 22 muestra la distribución de condiciones dentro de cada tipo celular, permitiendo identificar qué poblaciones están restringidas o enriquecidas en cada estado. Esto revela que poblaciones como Benign luminal epitelial y Basal epitelial son casi exclusivas del tejido sano, mientras que Neuroendocrine / EMT-like y MALAT1+ se asocian de manera casi exclusiva a CRPC. Las poblaciones CRIPS3-high y PCA3-high son específicas de PCa.

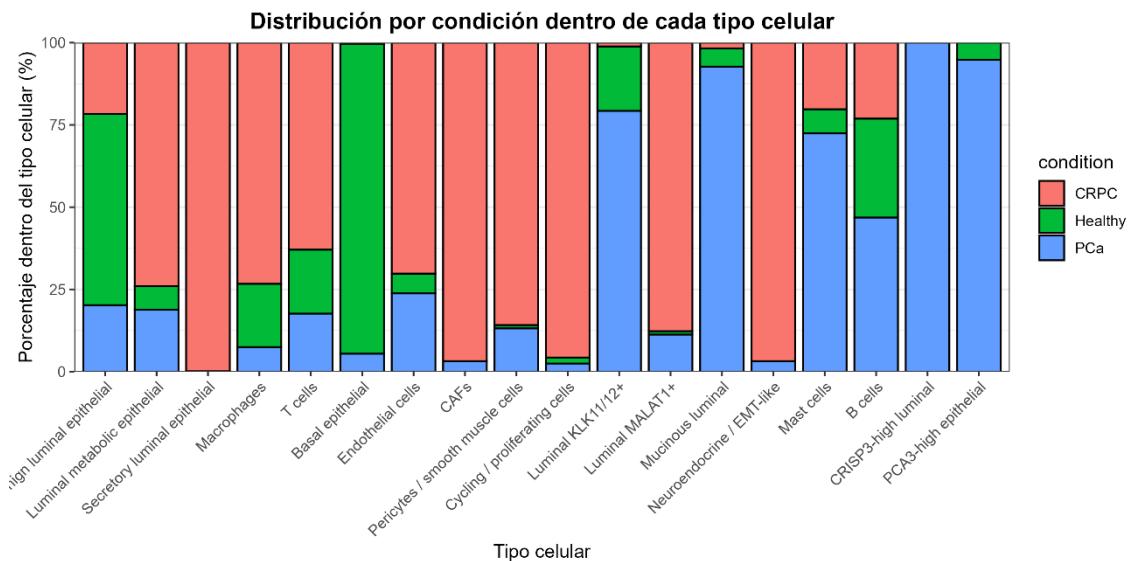


Figura 22: Distribución por condición dentro de cada población.

El criterio seguido para asignar identidades celulares a cada clúster, incluyendo el nombre de la población y sus genes característicos, se encuentra recogido en el Anexo: [9.5. Criterio seguido para caracterización de poblaciones celulares](#).

#### 4.7. Análisis de expresión diferencial por tipo celular.

De entre las poblaciones anotadas manualmente, se seleccionaron Benign luminal epithelial, Luminal Metabolic Epithelial, Macrophages, T cells y Endothelial cells para realizar el análisis de expresión diferencial entre condiciones biológicas. Esta selección se basó en tres criterios:

- (i) Su presencia suficiente en las tres condiciones.
- (ii) Su relevancia biológica en el contexto del cáncer de próstata.
- (iii) Su representación abundante en el conjunto de datos con tal de realizar comparaciones robustas.

En concreto, estas poblaciones mostraron presencia a lo largo de la progresión tumoral (Figura 23).

```

> ct_counts
# A tibble: 18 × 5
  CellType_Manual      CRPC Healthy   PCa ok_a113
  <chr>          <int>  <int>  <int>  <lgl>
1 Benign luminal epithelial    995   2664   925 TRUE
2 Luminal metabolic epithelia  3040    295   772 TRUE
3 Macrophages                2291    601  232 TRUE
4 T cells                     1618    499  453 TRUE
5 Endothelial cells           1216   104  412 TRUE
6 Secretory luminal epithelial 3382     1    1 FALSE
7 Basal epithelial              7   1938   112 FALSE
8 CAFs                         1654     0    54 FALSE
9 Pericytes / smooth muscle cells 1293     15   198 FALSE
10 Cycling / proliferating cells 1095     21    28 FALSE
11 Luminal KLK11/12+             9     9  152 FALSE
12 Luminal MALAT1+               594     7    76 FALSE
13 Mucinous luminal              9     29   482 FALSE
14 Neuroendocrine / EMT-like       306     0    10 FALSE
15 Mast cells                   61     22   218 FALSE
16 B cells                      40     52    81 FALSE
17 CRISP3-high luminal            0     0   143 FALSE
18 PCA3-high epithelial           0     5   92 FALSE
> # Lista recomendada (presentes en las 3 con N suficiente)
> ct_ok <- ct_counts %>% filter(ok_a113) %>% pull(CellType_Manual)
> ct_ok
[1] "Benign luminal epithelial" "Luminal metabolic epithelial" "Macrophages"
[3] "T cells"
[5] "Endothelial cells"

```

Figura 23: Número de células por tipo celular y condición y selección de poblaciones con representación suficiente en las tres condiciones.

El análisis completo de expresión diferencial se realizó mediante la generación de tablas (top\_degs\_tables) para cada población y comparación, que sirvieron de base para la elaboración de los volcano plots y heatmaps. Dado que los principales hallazgos se exponen gráficamente, no se presentan estas tablas de forma individual debido a su extensión.

#### 4.7.1. Células epiteliales luminales benignas.

##### Healthy vs. PCa.

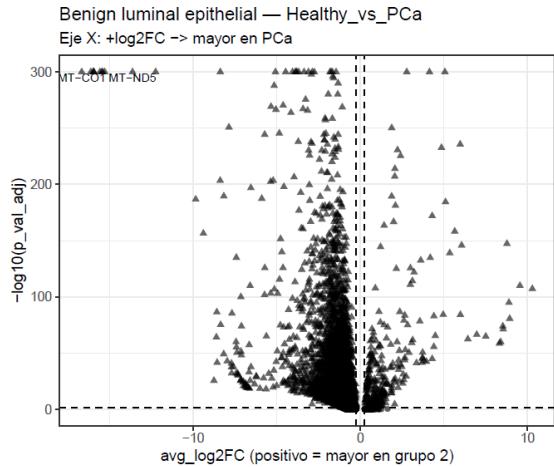


Figura 24: Volcano plot de Benign luminal epithelial para Healthy vs. PCa.

Benign luminal epithelial — Healthy vs PCa

AverageExpression (SCT data) | TOP UP/DOWN

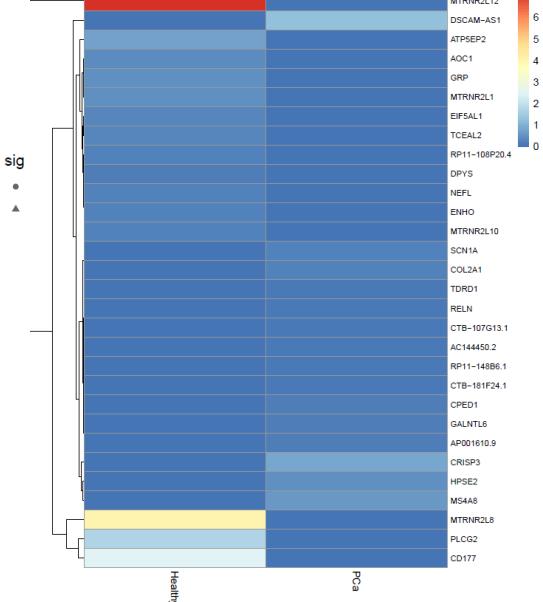


Figura 25: Heatmap de Benign luminal epithelial para Healthy vs. PCa.

El análisis mediante volcano plot (Figura 24) muestra una clara asimetría hacia la izquierda, indicando que un número considerable de genes presentan una mayor expresión en el

grupo Healthy. En concreto, los genes MT-ND5 y MT-CO1 destacan como los más diferencialmente expresados, con los valores de log2FC más negativos, lo que indica una fuerte sobreexpresión en tejido sano. Además, una gran concentración de genes alrededor del eje vertical sugiere que una parte importante del transcriptoma presenta cambios modestos o no significativos entre las dos condiciones.

Esta tendencia se confirma con el heatmap (Figura 25), que muestra los 30 genes más representativos según su expresión promedio. El gen MTRNR2L12 resalta por su alta expresión en el grupo Healthy y casi nula en PCa, posicionándolo como uno de los marcadores más distintivos del tejido sano. De manera similar, genes como MTRNR2L8, PLCG2 y CD177 presentan una mayor expresión en el grupo Healthy.

Por el contrario, genes como DSCAM-AS1, CRISP3, HPSE2 y MS4A8 se encuentran significativamente más expresados en PCa, indicando un posible papel en la progresión tumoral o en la desregulación endotelial asociada al cáncer.

### *Healthy vs CRPC.*

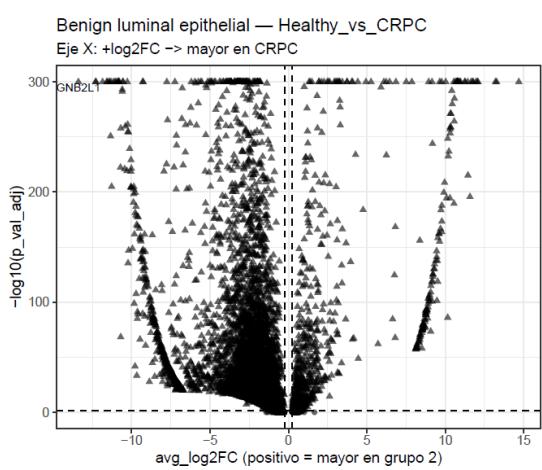


Figura 27: Volcano plot de Benign luminal epithelial para Healthy vs. CRPC.

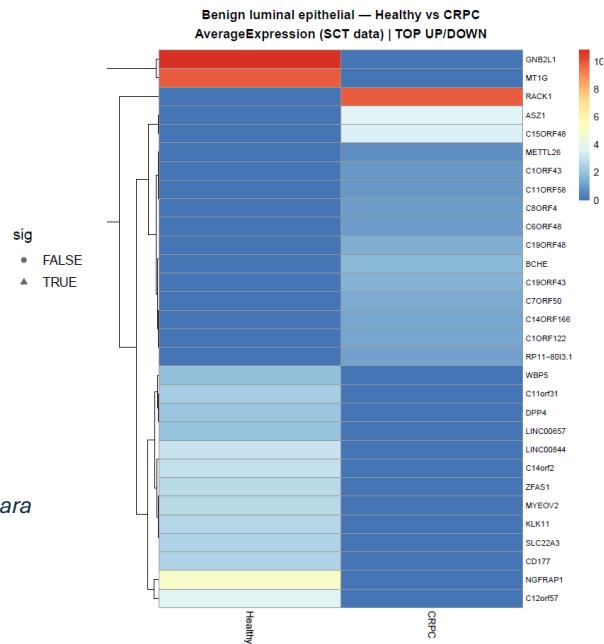


Figura 26: Heatmap de Benign luminal epithelial para Healthy vs. CRPC.

En la Figura 26 (volcano plot), se observa que la mayoría de los genes se concentran cerca del eje vertical, lo que indica cambios no significativos en la expresión entre grupos. No obstante, se evidencia una simetría clara hacia el lado izquierdo, que refleja una sobreexpresión generalizada en el grupo Healthy frente a CRPC. En particular, el gen GNB2L1 destaca como un outlier extremo, con un log2FC altamente negativo y una significancia estadística muy elevada, lo que sugiere una fuerte sobreexpresión en tejido sano.

La Figura 27 (heatmap) confirma esta tendencia, mostrando una expresión notablemente mayor en Healthy para GNB2L1 y también para MT1G, ambos situados en la parte superior del mapa de calor. En el extremo opuesto, se identifican genes como RACK1, ASZ1 y

C15ORF48 con mayor expresión en CRPC, lo que los posiciona como posibles marcadores relevantes en este grupo tumoral.

Además, en el tercio inferior del heatmap se aprecia un grupo de genes con expresión moderadamente superior en Healthy, entre ellos: WBP5, C11orf31, DPP4, LINC00657, C14orf2, ZFAS1, MYEOV2, KLK11, SLC22A3, CD177 y NGFRAP1. Aunque estas diferencias no son tan marcadas como las previamente mencionadas, podrían estar asociadas a procesos celulares diferenciadores del epitelio sano. Por su parte, el resto de genes analizados muestran una expresión ligeramente mayor en CRPC, sin alcanzar niveles de gran significancia.

#### PCA vs. CRPC.

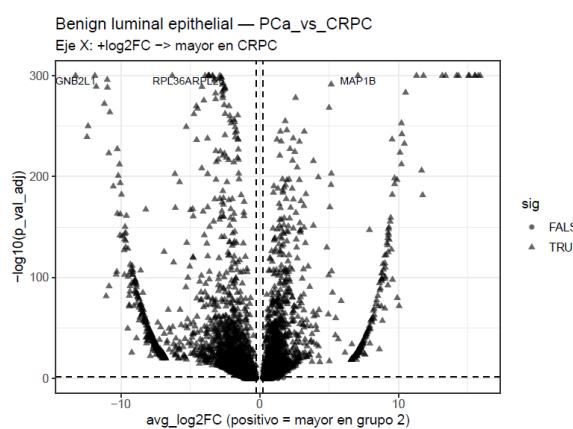


Figura 28: Volcano plot de Benign Luminal epithelial para PCA vs. CRPC.

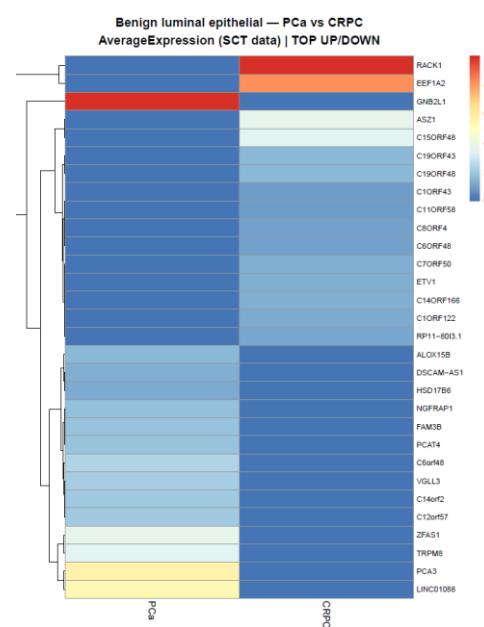


Figura 29: Heatmap de Benign luminal epithelial para PCA vs. CRPC.

El análisis mostrado en el volcano plot (Figura 28) revela una distribución en la que la mayoría de los genes se agrupan en la región central, lo que indica diferencias no significativas en su expresión. Sin embargo, se observa una simetría hacia la izquierda del eje X, reflejando una sobreexpresión generalizada en el grupo PCa. En este contexto, destacan tres genes: GNB2L1 —el más significativo del conjunto—, junto a RPL36A y RPL26, todos ellos ubicados en la región con log2FC negativo. Por otro lado, en el cuadrante superior derecho, se identifica una columna de genes más expresados en CRPC, entre ellos el gen MAP1B, que resalta por su significancia.

Este patrón se complementa con lo observado en la Figura 29 (heatmap), donde se confirma que GNB2L1 presenta una expresión claramente alta en PCa y prácticamente ausente en CRPC, coherente con su ubicación en el volcano plot. En el grupo CRPC, el gen RACK1 muestra una expresión marcadamente elevada, seguido por EEF1A2, ambos con un perfil característico de este grupo. Además, en la parte inferior del mapa, se observan genes predominantes en PCa, como PCA3, LINC01088, ZFAS1 y TRPM8.

Por otro lado, genes como ASZ1 y C15ORF48 presentan una expresión algo mayor en CRPC, aunque con diferencias menos pronunciadas. Es destacable también el patrón general del heatmap: los genes situados en la parte superior muestran en su mayoría una expresión superior en CRPC, mientras que los de la parte inferior se asocian a una mayor expresión en PCa, lo que refuerza la tendencia bidireccional detectada en el volcano plot.

#### 4.7.2. Células endoteliales.

##### *Healthy vs. PCa.*

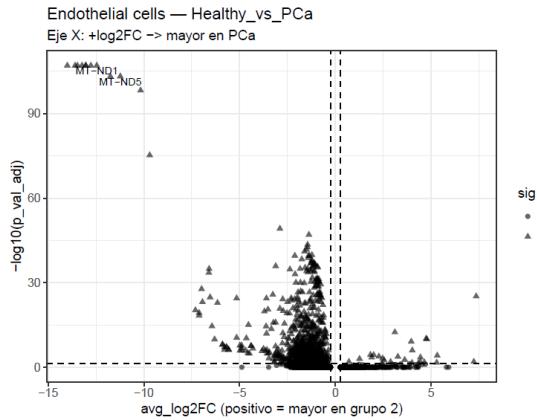


Figura 30: Volcano plot de Endothelial cells para Healthy vs. PCa.

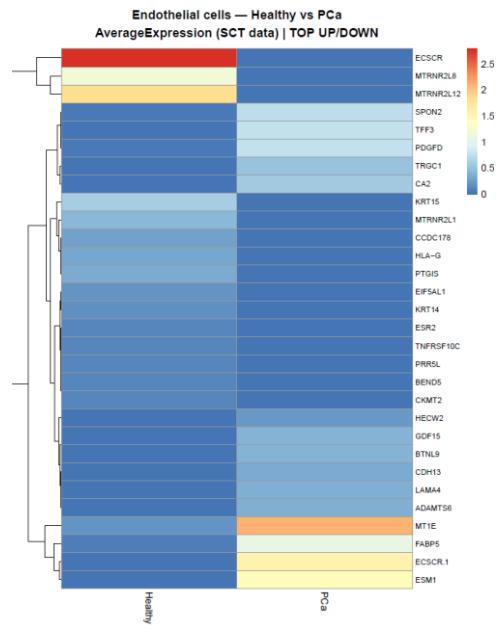


Figura 31: Heatmap de Endothelial cells para Healthy vs. PCa.

El análisis del volcano plot (Figura 30) revela que la mayoría de los genes se agrupan en torno al eje vertical, lo que indica que no presentan cambios significativos en su expresión entre las muestras sanas (Healthy) y tumorales (PCa). No obstante, se aprecia una simetría clara hacia la izquierda del gráfico, lo que refleja una sobreexpresión predominante en el grupo Healthy. En concreto, los genes MT-ND1 y MT-ND5 destacan como los más significativamente expresados en tejido sano, con log2FC negativo elevado y gran significancia estadística. Por el contrario, no se identifican genes notablemente sobreexpresados en el grupo PCa, ya que no se observa densidad destacada en la región derecha del gráfico.

Esta tendencia es coherente con lo mostrado en el heatmap (Figura 31), donde varios genes presentan mayor expresión en el grupo Healthy, especialmente ECSCR, MTRNR2L8 y MTRNR2L12, que se sitúan entre los más representativos. Por otro lado, algunos genes como MT1E, FABP5, ECSCR.1 y ESM1 muestran una expresión algo aumentada en PCa, aunque en menor magnitud y con menor contraste. De forma global, se observa una clara tendencia a una expresión génica más elevada en células endoteliales de muestras sanas, en comparación con el tejido tumoral primario, lo cual podría indicar un perfil funcional.

### Healthy vs. CRPC.

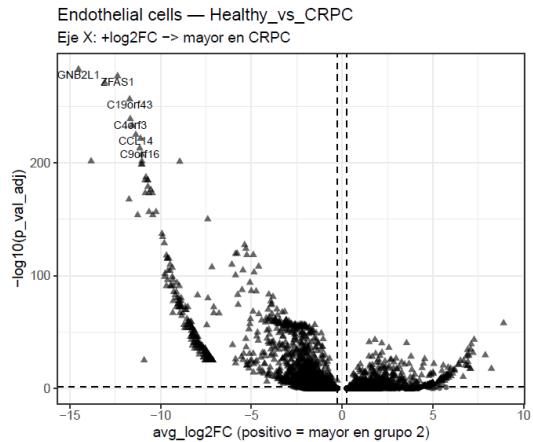


Figura 33: Volcano plot de Endothelial cells para Healthy vs. CRPC.

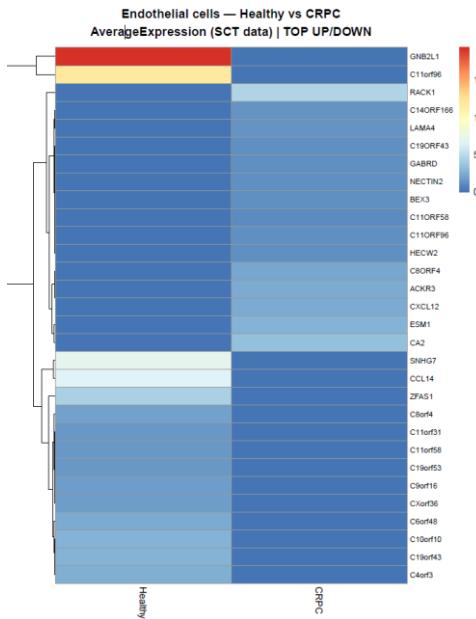


Figura 32: Heatmap de Endothelial cells para Healthy vs. CRPC.

En el volcano plot (Figura 32), se identifican seis genes con mayor expresión en el grupo Healthy, todos localizados en el cuadrante derecho del gráfico: GNB2L1 (el más significativo), ZFAS1, C19orf43, C4orf3, CCL14 y C9orf16. Esta agrupación indica una tendencia generalizada hacia la sobreexpresión en tejido sano, con valores elevados de significancia estadística y log2FC negativo. Por el contrario, no se detectan genes significativamente sobreexpresados en CRPC, ya que el cuadrante izquierdo del gráfico carece de puntos destacados. La zona central presenta una alta densidad de genes con cambios menores o no significativos, reflejando un perfil transcriptómico ampliamente conservado entre ambas condiciones para muchos genes.

El heatmap (Figura 33) apoya esta observación: el gen GNB2L1 muestra una expresión marcadamente elevada en Healthy, seguido —aunque con menor diferencia— por C11orf96. En la parte inferior del mapa, se aprecia que genes como ZFAS1, CCL14, SNHG7, C11orf31, C9orf16 y C19orf53 también presentan niveles de expresión notablemente más bajos en CRPC, lo que sugiere un patrón de represión o pérdida de función en la condición tumoral avanzada.

Aunque la mayoría de los genes más expresados se asocian a Healthy, también se identifican algunos genes con expresión moderadamente aumentada en CRPC, como CA2, ESM1, CXCL12, ACKR3 y C8ORF4, situados en la zona superior del mapa. Este grupo podría representar genes asociados a la desregulación del endotelio o a procesos relacionados con la progresión tumoral y la angiogénesis.

### PCa vs. CRPC.

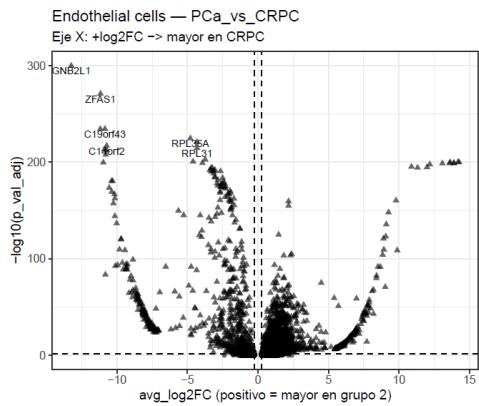


Figura 35: Volcano plot de Endothelial cells para PCa vs. CRPC.

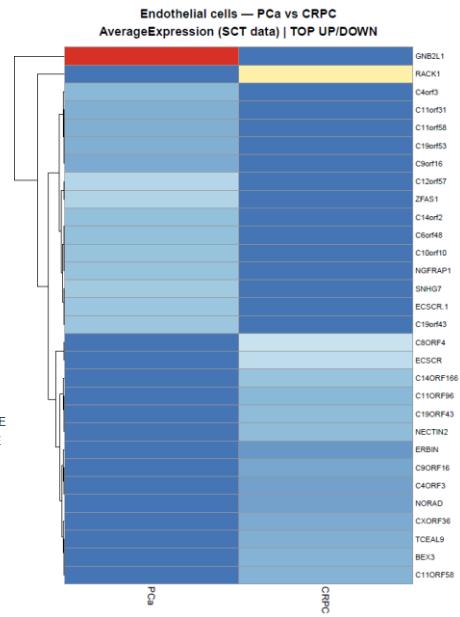


Figura 34: Heatmap de Endothelial cells para PCa vs. CRPC.

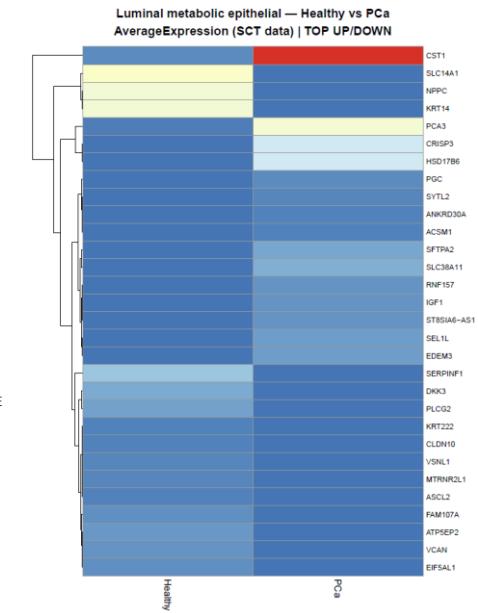
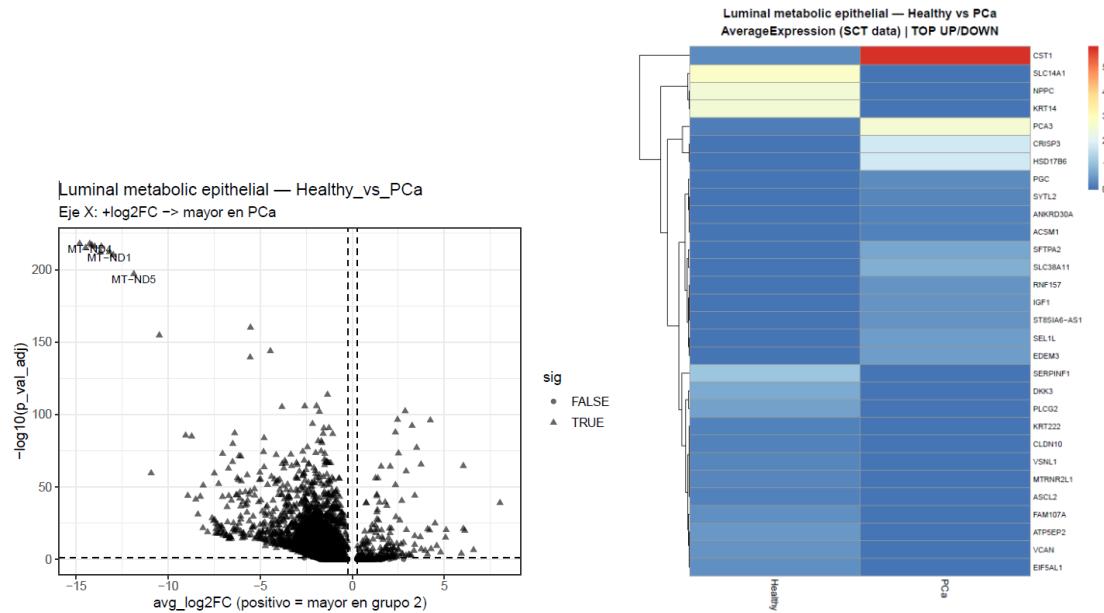
El volcano plot (Figura 34) muestra una clara asimetría hacia la izquierda, indicando una mayor expresión génica en el grupo PCa en comparación con CRPC. Destacan varios genes altamente significativos situados en el cuadrante superior izquierdo, como GNB2L1, ZFAS1, C19orf43, C14orf2, RPL36A y RPL31, todos ellos con log2FC negativo y valores de p ajustado muy bajos. Estos genes constituyen un núcleo de marcadores potenciales sobreexpresados en tejido tumoral primario. En contraste, no se observan genes destacadamente sobreexpresados en CRPC, y la mayoría de los genes se localizan en la zona central, indicando diferencias poco significativas entre ambas condiciones para la mayor parte del transcriptoma evaluado.

Esta tendencia se confirma en el heatmap (Figura 35), donde se observa que el gen GNB2L1 presenta una expresión extremadamente elevada en PCa y prácticamente ausente en CRPC. También destacan otros genes como ZFAS1 y C2orf57, que presentan niveles moderadamente más altos en PCa. En contraposición, RACK1 aparece como uno de los genes más expresados en CRPC, marcando una diferencia notable con respecto a su expresión en PCa.

Finalmente, genes como C8ORF4 y ECSCR muestran una expresión ligeramente mayor en CRPC, aunque las diferencias son menos pronunciadas. Este patrón mixto, aunque dominado por la sobreexpresión en PCa, sugiere que ciertos genes vinculados a la progresión tumoral pueden activarse de forma diferencial durante la transición hacia un estado resistente a la castración.

#### 4.7.3. Epiteliales luminales metabólicas.

*Healthy vs. PCa.*



El volcano plot (Figura 36) muestra una asimetría hacia la derecha, correspondiente a gentes más expresados en el grupo Healthy. En esta zona destacan los genes MT-ND4, MT-ND1 y MT-ND5, situados en el cuadrante superior derecho. La zona izquierda presenta una menor densidad de genes significativamente expresados en PCa. La mayoría de los genes se encuentran en torno al eje central.

En el heatmap (Figura 37), el gen CST1 presenta una expresión elevada en el grupo PCa, siendo el más destacado visualmente. En el grupo Healthy, se observan mayores niveles de expresión de los genes SLC14A1, NPPC y KRT14. En PCa, también se registran niveles más altos en los genes PCA3, CRIP3 y HSD17B6, aunque con menor intensidad relativa.

## Healthy vs. CRPC.

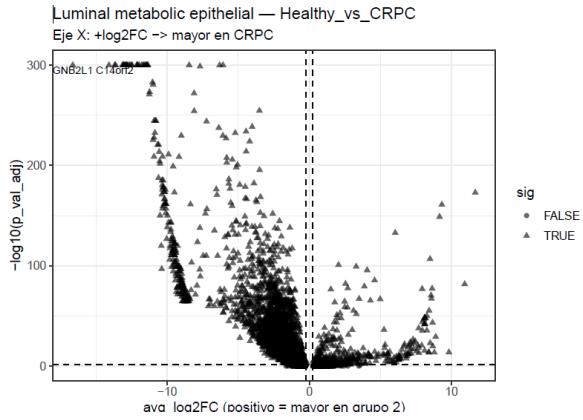


Figura 38: Volcano plot de Luminal metabolic epithelial para Healthy vs. CRPC.

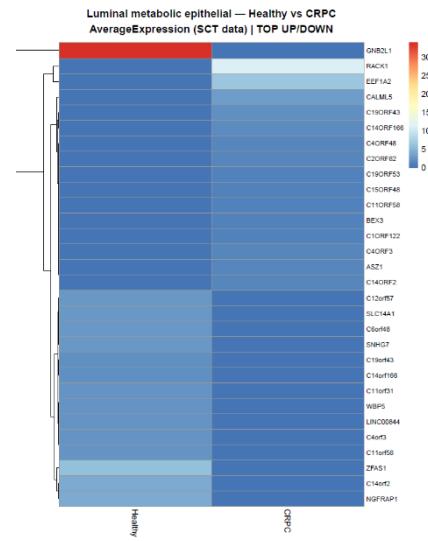


Figura 39: Heatmap de Luminal metabolic epithelial para Healthy vs. CRPC.

El volcano plot (Figura 38) muestra una asimetría hacia la izquierda, correspondiente a genes más expresados en el grupo Healthy. En esta región destacan los genes GNB2L1 y C14orf2, situados en el cuadrante superior izquierdo. No se observan genes claramente destacados en la zona derecha del gráfico, correspondiente a genes más expresados en CRPC. La mayoría de los genes se encuentran en torno al eje central.

En el heatmap (Figura 39), el gen GNB2L1 presenta una expresión elevada en el grupo Healthy, siendo prácticamente inexistente en CRPC. Asimismo, el gen ZFAS1 muestra una mayor expresión en Healthy. En contraste, los genes RACK1 y EEF1A2 presentan niveles de expresión más elevados en el grupo CRPC.

### PCa vs. CRPC.

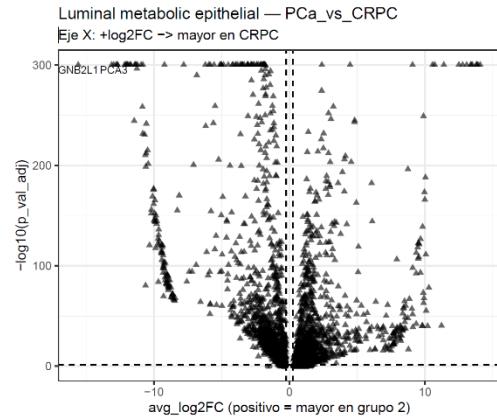


Figura 40: Volcano plot de Luminal metabolic epithelial para PCa vs CRPC.

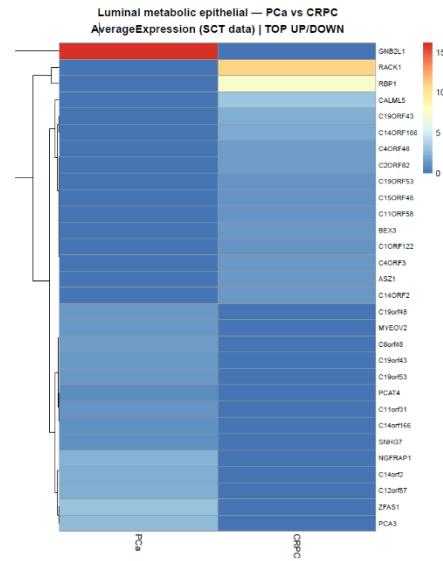


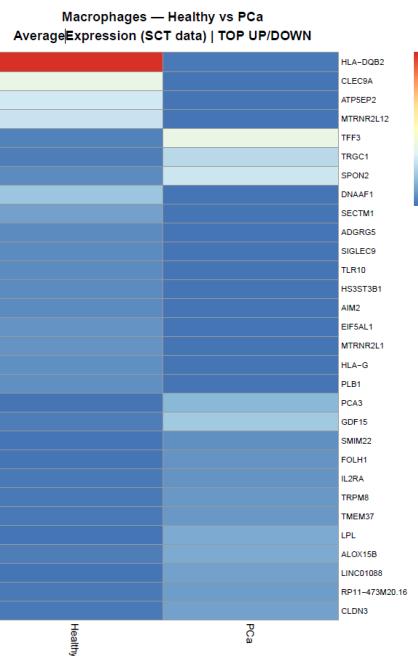
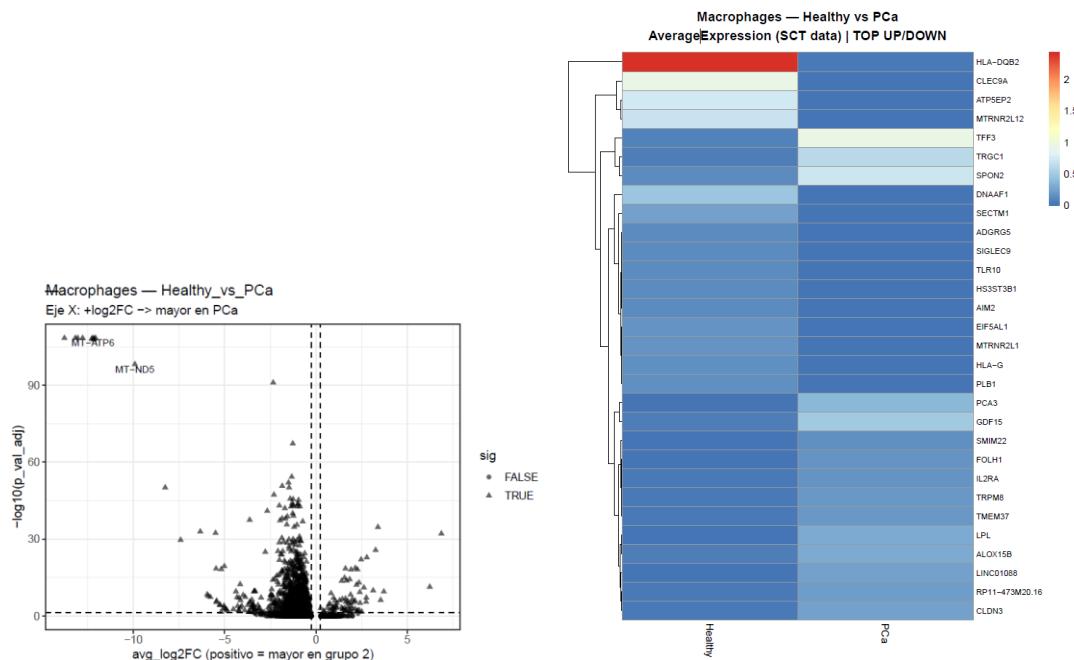
Figura 41: Heatmap de Luminal metabolic epithelial para PCa vs. CRPC.

El volcano plot (Figura 40) muestra una distribución bastante simétrica, con una ligera asimetría hacia la izquierda, correspondiente a genes más expresados en el grupo PCa. En esta región destacan los genes GNB2L1 y PCA3, situados en el cuadrante superior izquierdo. No se observan genes anotados en la zona derecha del gráfico, correspondiente a CRPC. La mayoría de los genes se concentran en torno al eje central, mostrando que no hay diferencias significativas.

En el heatmap (Figura 41), el gen GNB2L1 presenta una expresión muy elevada en el grupo PCa, siendo prácticamente inexistente en CRPC. En este último grupo, destacan los genes RACK1 y RBP1, con niveles de expresión más elevados comparados con PCa.

#### 4.7.4. Macrófagos.

*Healthy vs. PCa.*



El volcano plot (Figura 42) presenta una asimetría hacia la izquierda, correspondiente a genes más expresados en el grupo Healthy. En el cuadrante superior izquierdo destacan visualmente los genes MT-ATP6 y MT-ND5. La mayoría de los genes se concentran en la región central del gráfico, lo que indica ausencia de diferencias significativas en muchos casos.

En el heatmap (Figura 43), el gen HLA-DQB2 muestra una expresión marcadamente más alta en el grupo Healthy, siendo prácticamente inexistente en PCa. Otros genes como CLEC9A, ATP5E92 y MTRNR2L12 también presentan una mayor expresión en el grupo sano, aunque de forma más moderada. En el grupo PCa, se observan niveles algo más elevados de los genes TFF3, TRGC1 y SPON2, aunque con menor intensidad relativa.

### Healthy vs. CRPC.

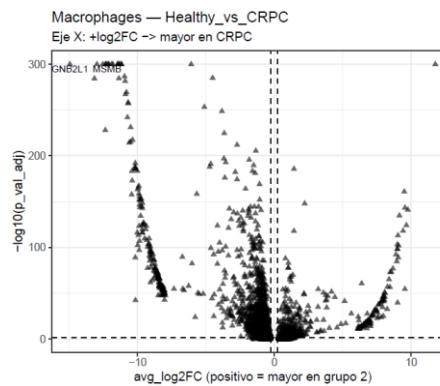


Figura 44: Volcan plot de Macrófagos para Healthy vs. CRPC.

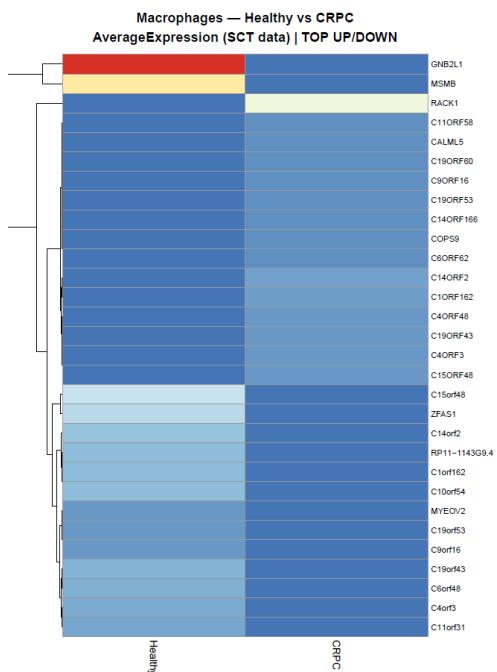


Figura 45: Heatmap de Macrófagos para Healthy vs. CRPC.

El volcano plot (Figura 44) muestra una simetría clara hacia la izquierda, correspondiente a genes con mayor expresión en el grupo Healthy. En el cuadrante superior izquierdo destacan los genes GNB2L1 y MSMB, ambos con alta significación y log2FC negativo. No se identifican genes anotados en el lado derecho del gráfico, correspondiente a CRPC. La mayoría de los genes se sitúan próximos al eje vertical, indicando ausencia de cambios significativos en gran parte del transcriptoma.

En el heatmap (Figura 45), el gen GNB2L1 presenta una expresión muy elevada en Healthy, seguido por MSMB, también con una expresión claramente superior en esta condición. En el grupo CRPC, destaca el gen RACK1 por su mayor expresión relativa. Además, se observa que la parte inferior del heatmap está dominada por genes con mayor expresión en el grupo Healthy.

### PCa vs. CRPC.

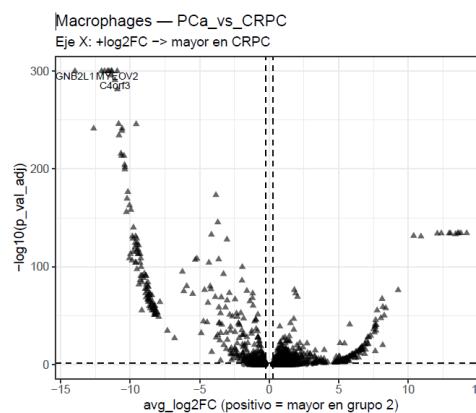


Figura 46: Volcano plot de Macrófagos para PCa vs. CRPC.

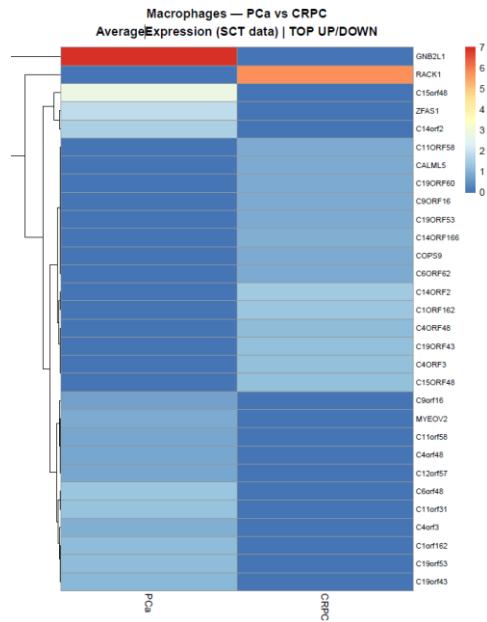


Figura 47: Heatmap de Macrófagos para PCa vs. CRPC.

El volcano plot (Figura 46) presenta una asimetría marcada hacia la izquierda, correspondiente a genes con mayor expresión en el grupo PCa. En el cuadrante superior izquierdo destacan los genes GNB2L1, MYEOV2 y C4orf3, todos ellos con log2FC negativo y alta significación estadística. No se identifican genes anotados en el lado derecho del gráfico, correspondiente al grupo CRPC. La mayoría de los genes se encuentran en la zona central, sin cambios significativos.

En el heatmap (Figura 47), el gen GNB2L1 muestra una expresión muy elevada en el grupo PCa, representado en rojo intenso, mientras que en CRPC su expresión es mínima. También se observan niveles más altos en PCa para los genes C15orf48, ZFAS1 y C14orf2, aunque con menor intensidad relativa. En contraste, en el grupo CRPC, destaca el gen RACK1 con mayor expresión, mientras que en PCa esta es baja o nula.

#### 4.7.5. Células T.

*Healthy vs. PCa.*

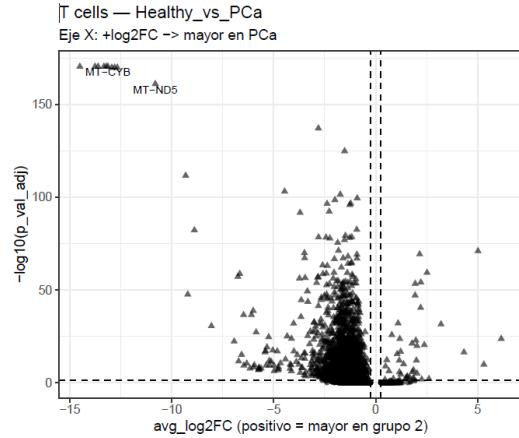


Figura 48: Volcano plot de Células T para Healthy vs. PCa.

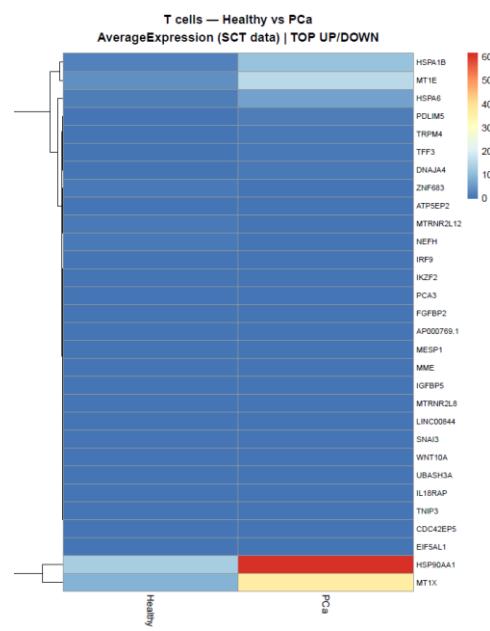


Figura 49: Heatmap de Células T para Healthy vs. PCa.

El volcano plot (Figura 48) muestra una asimetría hacia la izquierda, correspondiente a genes más expresados en el grupo Healthy. En esta zona destacan los genes MT-CYB y MT-ND5, situados en el cuadrante superior izquierdo. La mayoría de los genes se concentran en torno al eje central, indicando cambios no significativos en la expresión génica entre condiciones.

En el heatmap (Figura 49), el gen HSP90AA1 presenta una expresión muy elevada en PCa, siendo el más destacado visualmente. También muestran mayor expresión en este grupo los genes MT1X y MT1E, aunque en menor medida. En el grupo Healthy, los niveles de expresión son muy bajos en general, destacando únicamente HSP90AA1 y MT1X, pero con una intensidad considerablemente menor en comparación con PCa.

### Healthy vs. CRPC.

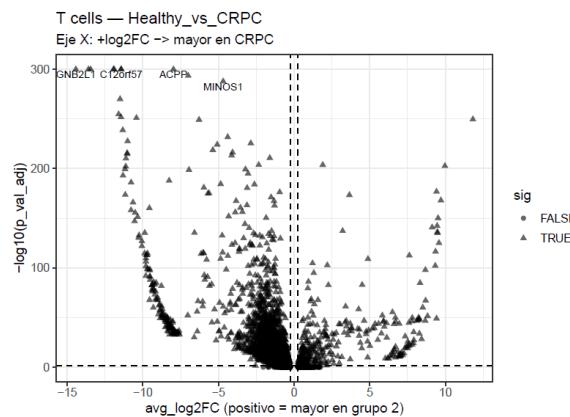


Figura 50: Volcano plot de Células T para Healthy vs. CRPC.

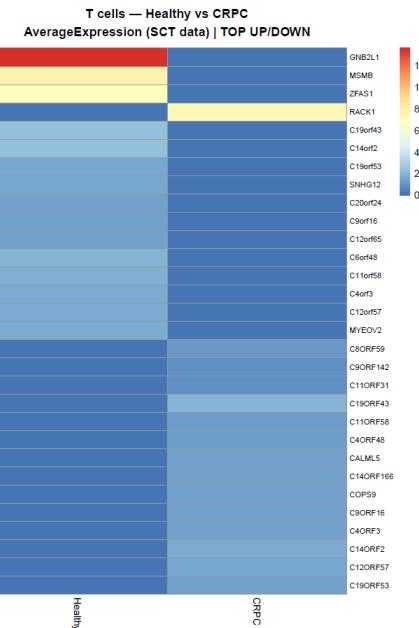


Figura 51: Heatmap de Células T para Healthy vs. CRPC.

El volcano plot (Figura 50) muestra una asimetría marcada hacia la izquierda, indicando una mayor expresión génica en el grupo Healthy. En el cuadrante superior izquierdo destacan los genes GNB2L1, C12orf57, ACPP y MINOS1 como los más diferencialmente expresados. En el extremo derecho no se observan genes anotados, y la mayoría se agrupan en torno al eje central, lo que indica ausencia de diferencias significativas en muchos casos.

En el heatmap (Figura 51), el gen GNB2L1 destaca por su expresión notablemente elevada en el grupo Healthy. Le siguen, aunque con menor intensidad relativa, los genes MSMB y ZFAS1. En el grupo CRPC, el gen más destacable es RACK1, que presenta niveles elevados de expresión comparados con el grupo sano.

### PCa vs. CRPC.

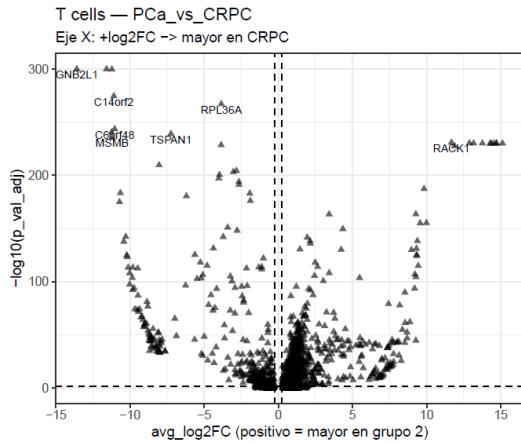


Figura 52: Volcano plot de Células T para PCa vs. CRPC.

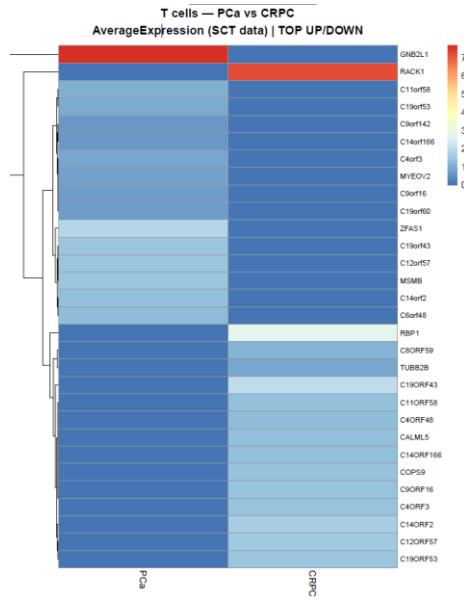


Figura 53: Heatmap de Células T para PCa vs. CRPC.

El volcano plot (Figura 52) muestra una ligera asimetría hacia la izquierda, lo que indica una mayor expresión génica en el grupo PCa. En el cuadrante superior izquierdo destacan los genes GNB2L1, C14orf2, C6orf48, MSMB y TSPAN1 como los más diferencialmente expresados en PCa. En el lado derecho del gráfico, correspondiente al grupo CRPC, RACK1 es el único gen anotado visualmente, presentando una sobreexpresión significativa en este grupo. La mayoría de los genes se concentran en la zona central, indicando ausencia de cambios significativos.

En el heatmap (Figura 53), se observa una expresión elevada del gen GNB2L1 en el grupo PCa, mientras que en CRPC destaca RACK1 como el gen más diferencialmente expresado. La parte superior del heatmap está dominada por genes más expresados en PCa, mientras que la parte inferior agrupa principalmente genes con mayor expresión en CRPC.

## 5. Discusión.

En este trabajo se ha desarrollado un análisis integral del tejido prostático humano en distintos estados de la enfermedad -tejido sano, cáncer de próstata localizado (PCa) y cáncer de próstata resistente a la castración (CRPC)- mediante técnicas de transcriptómica unicelular (scRNA-seq). A partir de la integración de múltiples datasets públicos y el uso de herramientas bioinformáticas avanzadas, se ha logrado caracterizar con alta resolución la diversidad celular del epitelio, del compartimento inmune y del estroma. La discusión a continuación aborda los principales hallazgos del estudio en términos de cambios en la composición celular, aparición de subpoblaciones tumorales específicas, remodelación del microambiente y perfiles de expresión diferencial, todo ello en el contexto del conocimiento actual sobre la progresión del cáncer de próstata.

### 5.1. Composición celular por condición.

El análisis de la composición celular entre condiciones permitió observar una reorganización profunda del tejido prostático durante la progresión de la enfermedad, consistente con la transformación maligna y los mecanismos de resistencia terapéutica. En el tejido sano predominan las poblaciones epiteliales benignas, especialmente las células luminales y basales, acompañadas por componentes vasculares e inmunes. Este perfil celular es coherente con lo reportado por Henry et al. (2018) y Karthaus et al. (2020), quienes caracterizaron el epitelio prostático humano sano y su microambiente mediante scRNA-seq, describiendo una clara compartmentalización epitelial con estabilidad fenotípica.

Durante el paso a cáncer de próstata localizado (PCa); se detectó una disminución de las poblaciones epiteliales benignas y un incremento de subtipos epiteliales tumorales, entre los que destacan los fenotipos KLK11/12+, mucinosos, CRISP3-high y PCA3-high. Este cambio refleja la adquisición de un programa tumoral asociado a secreción aberrante, desdiferenciación y alteración de la homeostasis epitelial, en línea con los perfiles identificados por Berglund et al. (2018) y Chen et al. (2021), quienes describen estas subpoblaciones como representativas de estadios intermedios de progresión y con potencial valor pronóstico.

En la etapa de CRPC, la emergencia de subtipos como Luminal MALAT1+ y Neuroendocrine/EMT-like, junto con una expansión de poblaciones proliferativas (Cycling cells) y fibroblastos asociados a cáncer (CAFs), apunta a un estado avanzado de plasticidad celular, activación de programas de transición epitelio-mesénquima y remodelación estromal. Estos fenómenos han sido vinculados a mecanismos de evasión terapéutica y progresión agresiva en estudios previos como los de Beltran et al. (2016) y He et al. (2021), que identifican poblaciones NEPC-like y EMT-activated en CRPC con perfiles transcriptómicos similares a los hallados en este estudio.

La distribución de condiciones dentro de cada tipo celular refuerza esta transición: las poblaciones benignas aparecen restringidas al tejido sano, mientras que las subpoblaciones malignas emergen de forma progresiva y específica según el estadio, lo que sugiere un proceso secuencial de remodelación tisular en el que se combinan cambios intrínsecos epiteliales y alteraciones en el microambiente tumoral.

## 5.2. Expresión diferencial.

### 5.2.1. Células epiteliales luminales benignas.

El análisis de expresión diferencial en células epiteliales luminales benignas entre condiciones biológicas reveló patrones consistentes con una progresiva pérdida de identidad epitelial sana durante la transformación tumoral. En la comparación entre tejido sano y PCa, se observó una fuerte sobreexpresión de genes mitocondriales como MT-ND5 y MT-CO1 en el grupo Healthy, lo cual sugiere una actividad metabólica oxidativa elevada asociada a epitelio funcionalmente diferenciado. Este perfil es coherente con estudios que describen la pérdida de metabolismo oxidativo y la transición hacia rutas metabólicas alteradas en células tumorales prostáticas (Ahmad et al., 2021; Costello & Franklin, 2006). Asimismo, la expresión elevada de MTRNR2L12 y MTRNR2L8 en tejido sano refuerza su papel como posibles marcadores de integridad epitelial y homeostasis celular.

En contraste, genes como CRISP3, DSCAM-AS1 y HPSE2 se encontraron sobreexpresados en PCa. CRISP3 ha sido previamente asociado a subtipos agresivos de cáncer de próstata y vinculado a peor pronóstico en múltiples cohortes (Berglund et al., 2018; Pinzani et al., 2008). Por su parte, DSCAM-AS1, un lncRNA, ha sido descrito como regulador positivo del crecimiento tumoral en otros tipos de cáncer, aunque su papel en próstata aún está en estudio (Liang et al., 2019). Estos cambios sugieren una reprogramación transcripcional activa incluso en poblaciones inicialmente benignas, posiblemente inducida por señales del microambiente tumoral.

La comparación entre tejido sano y CRPC mostró una tendencia similar: la mayoría de los genes diferencialmente expresados presentaban mayor expresión en el grupo sano, como GNB2L1 y MT1G, lo cual puede reflejar la pérdida progresiva de identidad epitelial en estadios avanzados. GNB2L1 (también conocido como RACK1) ha sido descrito como modulador de señalización celular y su pérdida se ha asociado a transformación epitelial y resistencia terapéutica (Patel et al., 2024). En CRPC, la sobreexpresión de genes como ASZ1 o C15ORF48 podría estar relacionada con estados transicionales o desdiferenciación, aunque su función en este contexto requiere mayor validación.

Finalmente, la comparación directa entre PCa y CRPC sugiere una desregulación más marcada en CRPC, con sobreexpresión de genes como RACK1 y EEF1A2, ambos asociados previamente a proliferación tumoral y resistencia a castración (Lyu et al., 2025). El mantenimiento de la expresión de PCA3 en PCa y su descenso en CRPC es consistente con la pérdida de dependencia androgénica en estadios avanzados (Deng et al., 2017).

En conjunto, estos resultados indican que incluso las poblaciones consideradas “benignas” sufren una transformación transcripcional significativa durante la progresión tumoral, y que su perfil de expresión puede aportar biomarcadores diferenciales según el estadio de la enfermedad.

### 5.2.2. Células endoteliales.

El análisis transcriptómico de las células endoteliales reveló una tendencia global a la pérdida de expresión génica en condiciones tumorales, en particular en CRPC, lo que sugiere una disfunción progresiva del endotelio durante la evolución del cáncer de próstata. En la comparación entre tejido sano y PCa, los genes MT-ND1 y MT-ND5, codificantes de subunidades del complejo I mitocondrial, se encontraron significativamente sobreexpresados en las células endoteliales de muestras sanas, indicando un perfil metabólicamente activo y funcional. Esta observación es coherente con estudios previos que describen una pérdida de función mitocondrial y un fenotipo más glucolítico en el endotelio tumoral, asociado a angiogénesis aberrante y permeabilidad vascular aumentada (De Bock et al., 2013; Jain, 2014).

La expresión elevada de genes como ECSCR y MTRNR2L12 en el endotelio sano podría reflejar funciones vasculares normales y homeostáticas, mientras que la escasa sobreexpresión de genes en PCa indica una transición hacia un estado menos diferenciado o parcialmente disfuncional. En CRPC, esta tendencia se acentúa: los genes GNB2L1, ZFAS1 y CCL14 se expresan significativamente más en tejido sano, mientras que el tumor resistente presenta una pérdida marcada de estos genes, en concordancia con una desregulación más severa del comportamiento endotelial. GNB2L1 (RACK1), ya identificado como diferencial en otras poblaciones, actúa como regulador de señalización celular y su pérdida se ha asociado a disrupción del citoesqueleto y de la adhesión celular (Adams et al., 2011), eventos relevantes en la disfunción endotelial tumoral.

Aunque en menor magnitud, algunos genes como CXCL12, ESM1 y ACKR3 mostraron mayor expresión en CRPC, lo cual es relevante, ya que estos genes han sido previamente implicados en angiogénesis tumoral y reclutamiento celular protumoral (Sun et al., 2010). La sobreexpresión de CXCL12 y su receptor ACKR3 (también conocido como CXCR7) se ha descrito como un eje clave en la neovascularización y en la promoción de metástasis en cánceres sólidos, incluido el de próstata.

La comparación entre PCa y CRPC refuerza esta progresión disfuncional: mientras que PCa conserva una mayor expresión de genes asociados al endotelio activo como GNB2L1 y ZFAS1, CRPC muestra un perfil más apagado, con pocas sobreexpresiones destacadas. Este patrón es consistente con un endotelio tumoral cada vez más alterado, posiblemente asociado a una microvasculatura irregular, menos funcional y más favorable al escape tumoral y a la evasión inmunitaria (Jain, 2005; Nagy et al., 2010).

Estos resultados apuntan a que la transición de un endotelio vascular funcional a uno aberrante y desregulado constituye un componente clave del microambiente tumoral en CRPC. Además, los genes diferencialmente expresados identificados podrían representar posibles biomarcadores de desregulación endotelial o dianas para la modulación de la angiogénesis en contextos terapéuticos (Ferrara & Adamis, 2016; Fukumura & Jain, 2007).

### 5.2.3. Células epiteliales luminales metabólicas.

Las células epiteliales luminales metabólicas mostraron un patrón de expresión diferencial coherente con una pérdida progresiva de identidad funcional durante la evolución tumoral. En la comparación entre tejido sano y PCa, se observó una clara sobreexpresión de genes mitocondriales en el grupo Healthy, entre ellos MT-ND1, MT-ND4 y MT-ND5, lo cual refleja una actividad metabólica oxidativa activa, típica de epitelios diferenciados. Estos hallazgos concuerdan con la literatura previa, que indica una pérdida del metabolismo oxidativo mitocondrial durante la transformación prostática, con un cambio hacia perfiles más glucolíticos en las células tumorales (Costello & Franklin, 2006; Loras et al., 2019).

En el grupo PCa, el gen CST1 apareció como el marcador más destacado, con una sobreexpresión clara. CST1 (Cistatina SN) ha sido reportado como regulador de proteasas y potencial promotor de crecimiento tumoral en diversos tipos de cáncer, aunque su función en próstata aún requiere mayor exploración (Dai et al., 2017; Liu et al., 2019). También se observaron niveles elevados de PCA3 y CRISP3, genes ya consolidados como marcadores asociados a malignidad prostática, especialmente PCA3, ampliamente utilizado como biomarcador diagnóstico no invasivo (Ma et al., 2016).

En la comparación entre tejido sano y CRPC, el gen GNB2L1 volvió a destacar como fuertemente expresado en condiciones no tumorales. Su expresión se redujo de forma marcada en CRPC, como ya se había observado en otras poblaciones celulares, lo que sugiere un papel clave en la homeostasis epitelial perdida durante la progresión a estados resistentes a la castración. La mayor expresión de RACK1 y EEF1A2 en CRPC también apunta a un cambio hacia programas de traducción activa y señalización promaligna, en línea con estudios que vinculan estos genes con resistencia terapéutica y plasticidad tumoral (Patel et al., 2024).

La comparación directa entre PCa y CRPC reveló un patrón más sutil pero coherente: GNB2L1 y PCA3 mostraron mayor expresión en PCa, mientras que RACK1 y RBP1 fueron más abundantes en CRPC. Este último gen, RBP1, codifica una proteína de unión a retinol y ha sido relacionado con procesos de desdiferenciación celular y progresión tumoral en otros contextos, aunque su papel en próstata requiere mayor validación (Fu et al., 2024).

En conjunto, estos resultados sugieren que las células epiteliales luminales metabólicas, pese a su perfil inicialmente funcional, atraviesan un proceso de desregulación progresiva tanto en términos metabólicos como en expresión génica asociada a malignidad, especialmente en CRPC. Este tipo de reprogramación metabólica y transcripcional ha sido descrito como un rasgo característico de la progresión tumoral y de la adquisición de resistencia terapéutica en cáncer de próstata (Costello & Franklin, 2006; Loras et al., 2019). Asimismo, la combinación de genes mitocondriales y lncRNAs como ZFAS1 y PCA3 podría representar una firma de transición entre estados epiteliales diferenciados y fenotipos tumorales avanzados, en línea con estudios previos que asocian estos lncRNAs con progresión y agresividad tumoral (Ghafouri-Fard et al., 2021; Rohan Ramnarine et al., 2019).

#### 5.2.4. Macrófagos.

El análisis de expresión diferencial en macrófagos reveló una clara tendencia hacia la pérdida de funciones inmunes clásicas en las condiciones tumorales, especialmente en CRPC. En la comparación entre tejido sano y PCa, se observó una mayor expresión de genes mitocondriales como MT-ATP6 y MT-ND5 en el grupo Healthy, lo cual sugiere un estado metabólicamente activo asociado a funciones inmunes efectivas. En este mismo grupo, genes como HLA-DQB2 y CLEC9A mostraron niveles elevados, indicando un fenotipo compatible con macrófagos presentadores de antígenos y orientados a la inmunovigilancia. CLEC9A, en particular, es un marcador de macrófagos y células dendríticas especializados en presentación cruzada, y su pérdida se ha asociado con menor capacidad para activar respuestas T antitumorales (Sancho et al., 2008).

En cambio, en PCa se observó una sobreexpresión moderada de genes como TFF3, SPON2 y TRGC1. Algunos de estos genes se han asociado a contextos inflamatorios protumoriales o remodelación tisular, lo que podría reflejar un cambio funcional hacia un fenotipo más regulador o incluso inmunosupresor. Este fenómeno es consistente con el proceso de "polarización" de los macrófagos hacia un fenotipo M2 en tumores sólidos, que favorece la progresión y evasión inmunitaria (Mantovani et al., 2002; Ruffell & Coussens, 2015).

En la comparación entre tejido sano y CRPC, este patrón se acentúa: genes clave como GNB2L1 y MSMB —ambos con funciones vinculadas a señalización celular e integridad epitelial— se expresaron de forma significativamente mayor en Healthy. La pérdida de MSMB ha sido ampliamente descrita como un evento frecuente en cáncer de próstata, y su disminución se asocia a progresión tumoral y menor respuesta inmune local (*Expression of the Prostate-Specific Membrane Antigen - PubMed*, n.d.; Lou et al., 2012). Por el contrario, en CRPC, el gen RACK1 mostró niveles elevados, lo que refuerza su asociación con estados resistentes y promotores de supervivencia tumoral, como ya se observó en otras poblaciones celulares.

Finalmente, la comparación directa entre PCa y CRPC muestra que este último presenta una reducción aún mayor de la expresión de genes asociados a inmunidad efectiva, mientras que en CRPC aumentan marcadores como RACK1, previamente vinculados a supervivencia celular y resistencia tumoral. Este patrón es coherente con la progresiva instauración de un microambiente inmunosupresor en el cáncer de próstata avanzado, caracterizado por la disfunción de macrófagos y la pérdida de capacidades antitumorales efectivas (Lopez-Bujanda & Drake, 2017; Petitprez et al., 2020).

La sobreexpresión de MYEOV2 y ZFAS1 en PCa podría reflejar una etapa intermedia de reprogramación inmune y transcripcional, previa a la consolidación del perfil inmunosupresor observado en CRPC, en línea con modelos actuales que proponen una transición gradual del microambiente tumoral durante la progresión de la enfermedad (Biswas & Mantovani, 2010; Chen et al., 2021; Song et al., 2022b).

En conjunto, estos resultados reflejan un cambio fenotípico progresivo en los macrófagos del tejido prostático, desde un estado funcional inmunocompetente en tejido sano hacia un perfil regulador o inmunosupresor en CRPC, en línea con modelos actuales de evasión inmunitaria tumoral. Este tipo de reprogramación inmune ha sido propuesto como un objetivo terapéutico potencial en cáncer de próstata avanzado, mediante la modulación

del fenotipo de los macrófagos intratumorales (Biswas & Mantovani, 2010; Lopez-Bujanda & Drake, 2017).

### 5.2.5. Células T.

La expresión génica en células T mostró un patrón consistente con una pérdida progresiva de activación inmunitaria en las condiciones tumorales. En la comparación entre tejido sano y PCa, los genes mitocondriales MT-CYB y MT-ND5 se expresaron en mayor medida en el grupo Healthy, lo que podría reflejar un metabolismo activo asociado a linfocitos funcionales. Este perfil mitocondrial ha sido vinculado con estados de activación inmunitaria en linfocitos T efectores (Sena et al., 2013). En cambio, en PCa se observó una fuerte sobreexpresión de HSP90AA1, un chaperón molecular cuya expresión elevada en células T ha sido relacionada con disfunción o agotamiento celular en entornos tumorales (Graner et al., 2009).

La comparación entre tejido sano y CRPC refuerza esta tendencia hacia la desactivación inmunológica: genes como GNB2L1, C12orf57, ACPP y MINOS1 mostraron una mayor expresión en el grupo Healthy, mientras que CRPC carece de genes diferencialmente sobreexpresados. ACPP, también conocido como fosfatasa ácida prostática, tiene funciones relevantes en el control del microambiente prostático y su pérdida puede reflejar desdiferenciación epitelial o disfunción linfocitaria (H. Xu et al., 2019). De nuevo, GNB2L1 y MSMB aparecen como genes marcadores de un entorno inmunológicamente activo, pero su pérdida progresiva en CRPC refuerza la idea de un microambiente más tolerante o evasivo frente al sistema inmune.

En contraste, RACK1 fue uno de los genes más expresados en células T de CRPC. Este gen, previamente vinculado con plasticidad celular y resistencia a estrés, ha sido asociado a vías de supervivencia en entornos inmunosupresores (D. Yang et al., 2024). Su expresión elevada en varias poblaciones celulares, incluyendo linfocitos, sugiere que su papel podría estar vinculado no solo a funciones tumorales directas, sino también a la desregulación del entorno inmunitario circundante.

En conjunto, estos resultados revelan una reducción progresiva en la actividad transcripcional asociada a funciones efectoras en las células T durante la evolución de la enfermedad. Este fenómeno es consistente con la presencia de linfocitos disfuncionales o agotados en el microambiente tumoral de CRPC, lo cual ha sido propuesto como uno de los mecanismos clave de evasión inmunitaria en el cáncer de próstata avanzado (Fan et al., 2024; I. Wang et al., 2022). La identificación de estos cambios en la expresión podría ser relevante para el desarrollo de estrategias inmunoterapéuticas dirigidas, especialmente en el contexto de la resistencia a la castración.

## 6. Conclusiones y trabajos futuros.

Este estudio ha permitido caracterizar con alta resolución la heterogeneidad celular del tejido prostático humano en tres condiciones biológicas clave —tejido sano, cáncer de próstata localizado (PCa) y cáncer de próstata resistente a la castración (CRPC)— mediante el análisis de transcriptómica unicelular. Los resultados obtenidos ofrecen una visión integradora de la progresión tumoral desde una perspectiva celular y molecular, revelando patrones de desregulación compartidos y específicos entre los distintos compartimentos del microambiente prostático.

A nivel de composición celular, se identificó una reorganización progresiva del epitelio, con la desaparición de subpoblaciones benignas y la emergencia de fenotipos tumorales específicos en PCa y CRPC, incluyendo subtipos epiteliales como KLK11/12+, MALAT1+ y Neuroendocrine/EMT-like. Estos cambios reflejan procesos de desdiferenciación y plasticidad celular que contribuyen a la agresividad del tumor y a la resistencia terapéutica.

En el análisis de expresión diferencial, se evidenció una pérdida sistemática de genes mitocondriales y de marcadores de homeostasis en prácticamente todas las poblaciones celulares evaluadas —epiteliales, endoteliales, inmunes— durante la progresión hacia CRPC. Esta tendencia se acompañó de una sobreexpresión progresiva de genes asociados a estrés celular, proliferación, remodelación tumoral y supresión inmunitaria, como RACK1, CST1, HSP90AA1 y SPON2, entre otros.

El compartimento inmune, representado por macrófagos y células T, mostró una reducción clara de funciones efectoras, presentando un fenotipo compatible con inmunosupresión tumoral. Esta transformación fue especialmente evidente en CRPC, donde se observaron perfiles transcripcionales apagados y sobreexpresión de genes implicados en evasión inmunitaria, lo cual refuerza la noción del cáncer de próstata avanzado como un tumor "frío" desde el punto de vista inmunológico.

En conjunto, los hallazgos de este trabajo respaldan un modelo de progresión prostática caracterizado por la pérdida de diferenciación celular, la reprogramación metabólica, la remodelación del estroma y la instauración de un microambiente inmunosupresor. La identificación de genes diferencialmente expresados en subpoblaciones específicas abre nuevas vías para la definición de biomarcadores de estadio o pronóstico, así como para la exploración de dianas terapéuticas orientadas a revertir la resistencia y promover la activación inmunitaria en CRPC.

### 6.1. Limitaciones y perspectivas futuras.

Aunque el presente estudio se basa en un enfoque analítico riguroso, es importante reconocer varias limitaciones que condicionan la interpretación y generalización de los resultados.

En primer lugar, el análisis se basa en datos públicos y previamente procesados, lo que implica una dependencia de la calidad experimental y del diseño original de los estudios incluidos. Esto puede introducir sesgos en la representación de ciertos tipos celulares o

condiciones clínicas, especialmente en poblaciones minoritarias o poco caracterizadas como los subtipos neuroendocrinos.

En segundo lugar, la asignación de identidades celulares se realizó mediante anotación bioinformática basada en perfiles de expresión y en bibliografía previa. Si bien este enfoque permite una caracterización robusta, no puede sustituir la validación experimental *in situ* mediante técnicas como inmunohistoquímica, hibridación *in situ* o proteómica espacial.

Tercero, el análisis se centró en perfiles transcriptómicos estáticos, sin incorporar componentes dinámicos como pseudotiempo o inferencias de trayectoria celular, que podrían aportar información adicional sobre las transiciones fenotípicas observadas, particularmente entre PCa y CRPC.

Por otro lado, aunque se han identificado genes diferencialmente expresados con potencial valor funcional, el estudio no aborda experimentalmente su papel causal en la progresión tumoral ni su viabilidad como dianas terapéuticas, lo cual deberá ser explorado en trabajos posteriores.

Finalmente, cabe mencionar que este análisis se basó en un número reducido de muestras ( $n=7$ ), debido a limitaciones computacionales asociadas al procesamiento de datos de alto volumen. Aunque este enfoque permitió la caracterización de las poblaciones celulares, el tamaño muestral restringido limita la generalización de los hallazgos. Por lo tanto, será necesario validar estos resultados en cohortes más amplias y diversas para confirmar su aplicabilidad clínica.

De cara al futuro, este trabajo ofrece una base sólida para avanzar hacia estudios funcionales y de validación clínica. Entre las líneas más prometedoras se incluyen: la validación de firmas transcriptómicas en cohortes independientes; la evaluación funcional de genes como RACK1, GNB2L1 o HSP90AA1 en modelos experimentales; y la integración de datos epigenéticos, mutacionales o de proteómica espacial para una comprensión más profunda del microambiente tumoral en cáncer de próstata.

## 7. Glosario.

| Término             | Definición  |
|---------------------|---|
| <b>scRNA-seq</b>    | (Single-cell RNA sequencing) Técnica que permite medir la expresión génica de forma individual en cada célula, revelando heterogeneidad celular en tejidos complejos.     |
| <b>CRPC</b>         | (Castration-Resistant Prostate Cancer) Subtipo avanzado de cáncer de próstata que progresiona a pesar del bloqueo hormonal.   |
| <b>CAF</b>          | (Cancer-Associated Fibroblast) Fibroblasto asociado al tumor que participa en la remodelación del estroma y la progresión tumoral.  |
| <b>TME</b>          | (Tumor Microenvironment) Conjunto de células, matriz extracelular y señales que rodean a las células tumorales e influyen en su comportamiento.                           |
| <b>EMT</b>          | (Epithelial-Mesenchymal Transition) Proceso mediante el cual las células epiteliales adquieren características migratorias y mesenquimales, asociado a metástasis.        |
| <b>Pseudotiempo</b> | Análisis computacional que ordena células en función de cambios progresivos en su expresión génica, simulando una trayectoria biológica.                                  |
| <b>Volcano plot</b> | Representación gráfica que combina magnitud de cambio ( $\log_{2}FC$ ) y significancia estadística (- $\log_{10}$ p-valor ajustado) en análisis de expresión diferencial. |
| <b>Heatmap</b>      | Mapa de calor que representa visualmente la expresión de genes en distintas condiciones o tipos celulares.  |
| <b>lncRNA</b>       | (Long non-coding RNA) ARN largo no codificante con funciones regulatorias en múltiples procesos biológicos y tumorales.   |
| <b>RACK1</b>        | Gen asociado a rutas de señalización y respuesta al estrés, frecuentemente implicado en progresión tumoral.   |
| <b>MSMB</b>         | Gen supresor prostático cuya expresión se pierde comúnmente en etapas avanzadas del cáncer de próstata.   |

## 8. Bibliografía.

- Adams, D. R., Ron, D., & Kiely, P. A. (2011). RACK1, A multifaceted scaffolding protein: Structure and function. *Cell Communication and Signaling: CCS*, 9, 22. <https://doi.org/10.1186/1478-811X-9-22>
- Ahmad, F., Cherukuri, M. K., & Choyke, P. L. (2021). Metabolic reprogramming in prostate cancer. *British Journal of Cancer*, 125(9), 1185–1196. <https://doi.org/10.1038/S41416-021-01435-5>
- Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J. C., Geurts, P., Aerts, J., Van Den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11), 1083–1086. <https://doi.org/10.1038/NMETH.4463>
- Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R. P., Wolters, P. J., Abate, A. R., Butte, A. J., & Bhattacharya, M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*, 20(2), 163–172. <https://doi.org/10.1038/S41590-018-0276-Y>
- Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., Ginkhoux, F., & Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 2018 37:1, 37(1), 38–44. <https://doi.org/10.1038/nbt.4314>
- Beltran, H., Prandi, D., Mosquera, J. M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B. V. S. K., Varambally, S., Tomlins, S. A., Nanus, D. M., Tagawa, S. T., Van Allen, E. M., Elemento, O., Sboner, A., Garraway, L. A., Rubin, M. A., & Demichelis, F. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nature Medicine*, 22(3), 298–305. <https://doi.org/10.1038/NM.4045>
- Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstråhle, J., Tarish, F., Tanoglidi, A., Vickovic, S., Larsson, L., Salmén, F., Ogris, C., Wallenborg, K., Lagergren, J., Ståhl, P., Sonnhammer, E., Helleday, T., & Lundeberg, J. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications* 2018 9:1, 9(1), 2419-. <https://doi.org/10.1038/s41467-018-04724-5>
- Biswas, S. K., & Mantovani, A. (2010). Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nature Immunology*, 11(10), 889–896. <https://doi.org/10.1038/NI.1937>
- Bluemn, E. G., Coleman, I. M., Lucas, J. M., Coleman, R. T., Hernandez-Lopez, S., Tharakan, R., Bianchi-Frias, D., Dumpit, R. F., Kaipainen, A., Corella, A. N., Yang, Y. C., Nyquist, M. D., Mostaghel, E., Hsieh, A. C., Zhang, X., Corey, E., Brown, L. G., Nguyen, H. M., Pienta, K., ... Nelson, P. S. (2017). Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell*, 32(4), 474-489.e6. <https://doi.org/10.1016/J.CCELL.2017.09.003>

- Bray Bsc, F., Laversanne, | Mathieu, Hyuna, |, Phd, S., Ferlay, J., Siegel Mph, R. L., Soerjomataram, I., Ahmedin, |, & Dvm, J. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263. <https://doi.org/10.3322/CAAC.21834>
- Browaeys, R., Saelens, W., & Saeys, Y. (2019). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nature Methods* 2019 17:2, 17(2), 159–162. <https://doi.org/10.1038/s41592-019-0667-5>
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 2018 36:5, 36(5), 411–420. <https://doi.org/10.1038/nbt.4096>
- Chan, J. M., Zaidi, S., Love, J. R., Zhao, J. L., Setty, M., Wadosky, K. M., Gopalan, A., Choo, Z. N., Persad, S., Choi, J., LaClair, J., Lawrence, K. E., Chaudhary, O., Xu, T., Masilionis, I., Linkov, I., Wang, S., Lee, C., Barlas, A., ... Sawyers, C. L. (2022). Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science*, 377(6611), 1180–1191. <https://doi.org/10.1126/science.abn0478>
- Chen, S., Zhu, G., Yang, Y., Wang, F., Xiao, Y. T., Zhang, N., Bian, X., Zhu, Y., Yu, Y., Liu, F., Dong, K., Mariscal, J., Liu, Y., Soares, F., Loo Yau, H., Zhang, B., Chen, W., Wang, C., Chen, D., ... Ren, S. (2021). Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nature Cell Biology* 2021 23:1, 23(1), 87–98. <https://doi.org/10.1038/s41556-020-00613-6>
- Chesner, L. N., Polessko, F., Graff, J. N., Hawley, J. E., Smith, A. K., Lundberg, A., Das, R., Shenoy, T., Sjöström, M., Zhao, F., Hu, Y. M., Linder, S., Chen, W. S., Hawkins, R. M., Shrestha, R., Zhu, X., Foye, A., Li, H., Kim, L. M., ... Moran, A. E. (2025). Androgen Receptor Inhibition Increases MHC Class I Expression and Improves Immune Response in Prostate Cancer. *Cancer Discovery*, 15(3), 481–494. <https://doi.org/10.1158/2159-8290.CD-24-0559/750492/AM/ANDROGEN-RECEPTOR-INHIBITION-INCREASES-MHC-CLASS-I>
- Clarke, Z. A., & Bader, G. D. (2024). MALAT1 expression indicates cell quality in single-cell RNA sequencing data. *BioRxiv*, 2024.07.14.603469. <https://doi.org/10.1101/2024.07.14.603469>
- Consortium, H. (2019). The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* 2019 574:7777, 574(7777), 187–192. <https://doi.org/10.1038/s41586-019-1629-x>
- Costello, L. C., & Franklin, R. B. (2006). The clinical relevance of the metabolism of prostate cancer; zinc and tumor suppression: connecting the dots. *Molecular Cancer*, 5, 17. <https://doi.org/10.1186/1476-4598-5-17>
- Crowley, L., Cambuli, F., Aparicio, L., Shibata, M., Robinson, B. D., Xuan, S., Li, W., Hibshoosh, H., Loda, M., Rabidan, R., & Shen, M. M. (2020). A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *ELife*, 9, 1–24. <https://doi.org/10.7554/ELIFE.59465>

- Dai, D. nian, Li, Y., Chen, B., Du, Y., Li, S. bing, Lu, S. xun, Zhao, Z. ping, Zhou, A. jun, Xue, N., Xia, T. liang, Zeng, M. sheng, Zhong, Q., & Wei, W. dong. (2017). Elevated expression of CST1 promotes breast cancer progression and predicts a poor prognosis. *Journal of Molecular Medicine (Berlin, Germany)*, 95(8), 873. <https://doi.org/10.1007/S00109-017-1537-1>
- De Bock, K., Georgiadou, M., Schoors, S., Kuchnio, A., Wong, B. W., Cantelmo, A. R., Quaegebeur, A., Ghesquière, B., Cauwenberghs, S., Eelen, G., Phng, L. K., Betz, I., Tembuyser, B., Brepoels, K., Welti, J., Geudens, I., Segura, I., Cruys, B., Bifari, F., ... Carmeliet, P. (2013). Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell*, 154(3). <https://doi.org/10.1016/J.CELL.2013.06.037>
- Deng, J., Tang, J., Wang, G., & Zhu, Y. S. (2017). Long Non-Coding RNA as Potential Biomarker for Prostate Cancer: Is It Making a Difference? *International Journal of Environmental Research and Public Health* 2017, Vol. 14, Page 270, 14(3), 270. <https://doi.org/10.3390/IJERPH14030270>
- Domínguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., Suchanek, O., Polanski, K., King, H. W., Mamanova, L., Huang, N., Szabo, P. A., Richardson, L., Bolt, L., Fasouli, E. S., Mahbubani, K. T., Prete, M., Tuck, L., ... Teichmann, S. A. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594). [https://doi.org/10.1126/SCIENCE.ABL5197/SUPPL\\_FILE/SCIENCE.ABL5197\\_MDAR\\_REPRODUCIBILITY\\_CHECKLIST.PDF](https://doi.org/10.1126/SCIENCE.ABL5197/SUPPL_FILE/SCIENCE.ABL5197_MDAR_REPRODUCIBILITY_CHECKLIST.PDF)
- Dutt, S. S., & Gao, A. C. (2009). Molecular mechanisms of castration-resistant prostate cancer progression. *Future Oncol*, 5(9), 1403–1413. <https://doi.org/10.2217/fon.09.117>
- Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nature Protocols*, 15(4), 1484–1506. <https://doi.org/10.1038/S41596-020-0292-X>
- Expression of the prostate-specific membrane antigen - PubMed.* (n.d.). Retrieved December 31, 2025, from <https://pubmed.ncbi.nlm.nih.gov/7511053/>
- Fan, Q., Wang, Y., Cheng, J., Pan, B., Zang, X., Liu, R., & Deng, Y. (2024). Single-cell RNA-seq reveals T cell exhaustion and immune response landscape in osteosarcoma. *Frontiers in Immunology*, 15. <https://doi.org/10.3389/FIMMU.2024.1362970>
- Ferrara, N., & Adamis, A. P. (2016). Ten years of anti-vascular endothelial growth factor therapy. *Nature Reviews Drug Discovery*, 15(6), 385–403. <https://doi.org/10.1038/NRD.2015.17>
- Franzén, O., Gan, L. M., & Björkegren, J. L. M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, 2019(1), 46. <https://doi.org/10.1093/DATABASE/BAZ046>

- Fu, L. L., Yan, M., Yu, X., Shao, M., Gosau, M., Friedrich, R. E., Vollkommer, T., Smeets, R., Feng, H. C., & Xu, L. (2024). Retinol-binding protein type 1 expression predicts poor prognosis in head and neck squamous cell carcinoma. *BMC Cancer*, 24(1), 1277. <https://doi.org/10.1186/S12885-024-12565-3>
- Fukumura, D., & Jain, R. K. (2007). Tumor microvasculature and microenvironment: Targets for anti-angiogenesis and normalization. *Microvascular Research*, 74(2–3), 72. <https://doi.org/10.1016/J.MVR.2007.05.003>
- Gao, D., Vela, I., Sboner, A., Iaquinta, P. J., Karthaus, W. R., Gopalan, A., Dowling, C., Wanjala, J. N., Undvall, E. A., Arora, V. K., Wongvipat, J., Kossai, M., Ramazanoglu, S., Barboza, L. P., Di, W., Cao, Z., Zhang, Q. F., Sirota, I., Ran, L., ... Chen, Y. (2014). Organoid cultures derived from patients with advanced prostate cancer. *Cell*, 159(1), 176–187. <https://doi.org/10.1016/J.CELL.2014.08.016>
- Ge, R., Wang, Z., & Cheng, L. (2022). Tumor microenvironment heterogeneity an important mediator of prostate cancer progression and therapeutic resistance. *Npj Precision Oncology* 2022 6:1, 6(1), 31-. <https://doi.org/10.1038/s41698-022-00272-w>
- Ghafouri-Fard, S., Kamali, M. J., Abak, A., Shoorei, H., & Taheri, M. (2021). LncRNA ZFAS1: Role in tumorigenesis and other diseases. *Biomedicine & Pharmacotherapy*, 142, 111999. <https://doi.org/10.1016/J.BIOPHA.2021.111999>
- Graner, M. W., Raynes, D. A., Bigner, D. D., & Guerriero, V. (2009). Heat shock protein 70-binding protein 1 is highly expressed in high-grade gliomas, interacts with multiple heat shock protein 70 family members, and specifically binds brain tumor cell surfaces. *Cancer Science*, 100(10), 1870–1879. <https://doi.org/10.1111/J.1349-7006.2009.01269.X>
- Guo, W., Li, L., He, J., Liu, Z., Han, M., Li, F., Xia, X., Zhang, X., Zhu, Y., Wei, Y., Li, Y., Aji, R., Dai, H., Wei, H., Li, C., Chen, Y., Chen, L., & Gao, D. (2020). Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nature Genetics*, 52(9), 908. <https://doi.org/10.1038/S41588-020-0642-1>
- Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, 20(1), 1–15. <https://doi.org/10.1186/S13059-019-1874-1/FIGURES/6>
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13), 3573–3587.e29. <https://doi.org/10.1016/J.CELL.2021.04.048>
- He, M. X., Cuoco, M. S., Crowdus, J., Bosma-Moody, A., Zhang, Z., Bi, K., Kanodia, A., Su, M. J., Ku, S. Y., Garcia, M. M., Sweet, A. R., Rodman, C., DelloStritto, L., Silver, R., Steinharter, J., Shah, P., Izar, B., Walk, N. C., Burke, K. P., ... Van Allen, E. M. (2021). Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nature Medicine* 2021 27:3, 27(3), 426–433. <https://doi.org/10.1038/s41591-021-01244-6>

- Henry, G. H., Malewska, A., Joseph, D. B., Malladi, V. S., Lee, J., Torrealba, J., Mauck, R. J., Gahan, J. C., Raj, G. V., Roehrborn, C. G., Hon, G. C., MacConmara, M. P., Reese, J. C., Hutchinson, R. C., Vezina, C. M., & Strand, D. W. (2018). A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Reports*, 25(12), 3530-3542.e5. <https://doi.org/10.1016/J.CELREP.2018.11.086>
- Heumos, L., Schaar, A. C., Lance, C., Litinskaya, A., Drost, F., Zappia, L., Lücke, M. D., Strobl, D. C., Henao, J., Curion, F., Aliee, H., Ansari, M., Badia-i-Mompel, P., Büttner, M., Dann, E., Dimitrov, D., Dony, L., Frishberg, A., He, D., ... Theis, F. J. (2023). Best practices for single-cell analysis across modalities. *Nature Reviews Genetics* 2023 24:8, 24(8), 550–572. <https://doi.org/10.1038/s41576-023-00586-w>
- Hie, B. L., Kim, S., Rando, T. A., Bryson, B., & Berger, B. (2024). Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nature Protocols*, 19(8), 2283–2297. <https://doi.org/10.1038/S41596-024-00991-3>
- Hirz, T., Mei, S., Sarkar, H., Kfouri, Y., Wu, S., Verhoeven, B. M., Subtelny, A. O., Zlatev, D. V., Wszolek, M. W., Salari, K., Murray, E., Chen, F., Macosko, E. Z., Wu, C. L., Scadden, D. T., Dahl, D. M., Baryawno, N., Saylor, P. J., Kharchenko, P. V., & Sykes, D. B. (2023). Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nature Communications*, 14(1). <https://doi.org/10.1038/s41467-023-36325-2>
- Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., Bai, J., Chen, J., Jiang, W., Yang, K., Ou, Q., Li, X., Wang, P., & Zhang, Y. (2023). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Research*, 51(D1), D870–D876. <https://doi.org/10.1093/NAR/GKAC947>
- Ianevski, A., Giri, A. K., & Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nature Communications* 2022 13:1, 13(1), 1246-. <https://doi.org/10.1038/s41467-022-28803-w>
- Jain, R. K. (2005). Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science (New York, N.Y.)*, 307(5706), 58–62. <https://doi.org/10.1126/SCIENCE.1104819>
- Jain, R. K. (2014). Antiangiogenesis strategies revisited: from starving tumors to alleviating hypoxia. *Cancer Cell*, 26(5), 605–622. <https://doi.org/10.1016/J.CCCELL.2014.10.006>
- Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., Myung, P., Plikus, M. V., & Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. *Nature Communications* 2021 12:1, 12(1), 1088-. <https://doi.org/10.1038/s41467-021-21246-9>
- Jones, R. C., Karkanias, J., Krasnow, M. A., Pisco, A. O., Quake, S. R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., Harper, W., Hemenez, M., Ponnusamy, R., Salehi, A., Sanagavarapu, B. A., Spallino, E., Aaron, K. A., Concepcion, W., Gardner, J. M., Kelly, B., ... Wyss-Coray, T. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, 376(6594).

[https://doi.org/10.1126/SCIENCE.ABL4896/SUPPL\\_FILE/SCIENCE.ABL4896\\_MDAR\\_REPRODUCIBILITY\\_CHECKLIST.PDF](https://doi.org/10.1126/SCIENCE.ABL4896/SUPPL_FILE/SCIENCE.ABL4896_MDAR_REPRODUCIBILITY_CHECKLIST.PDF)

- Kalluri, R. (2016). The biology and function of fibroblasts in cancer. *Nature Reviews Cancer* 2016 16:9, 16(9), 582–598. <https://doi.org/10.1038/nrc.2016.73>
- Karthaus, W. R., Hofree, M., Choi, D., Linton, E. L., Turkekul, M., Bejnood, A., Carver, B., Gopalan, A., Abida, W., Laudone, V., Biton, M., Chaudhary, O., Xu, T., Masilionis, I., Manova, K., Mazutis, L., Pe'er, D., Regev, A., & Sawyers, C. L. (2020). Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science*, 368(6490), 497–505. [https://doi.org/10.1126/SCIENCE.AAY0267/SUPPL\\_FILE/AAY0267-TABLES\\_S1\\_TO\\_S9.ZIP](https://doi.org/10.1126/SCIENCE.AAY0267/SUPPL_FILE/AAY0267-TABLES_S1_TO_S9.ZIP)
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, & Raychaudhuri, S. (2019a). Fast, sensitive, and accurate integration of single cell data with Harmony. *Nature Methods*, 16(12), 1289. <https://doi.org/10.1038/S41592-019-0619-0>
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, & Raychaudhuri, S. (2019b). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* 2019 16:12, 16(12), 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>
- Kumar, V., Patel, S., Tcyganov, E., & Gabrilovich, D. I. (2016). The Nature of Myeloid-Derived Suppressor Cells in the Tumor Microenvironment. *Trends in Immunology*, 37(3), 208–220. <https://doi.org/10.1016/J.IT.2016.01.004>
- Liang, W. H., Li, N., Yuan, Z. Q., Qian, X. L., & Wang, Z. H. (2019). DSCAM-AS1 promotes tumor growth of breast cancer by reducing miR-204-5p and up-regulating RRM2. *Molecular Carcinogenesis*, 58(4), 461–473. <https://doi.org/10.1002/MC.22941>
- Liu, Y., Ma, H., Wang, Y., Du, X., & Yao, J. (2019). Cystatin SN Affects Cell Proliferation by Regulating the ER $\alpha$ /PI3K/AKT/ER $\alpha$  Loopback Pathway in Breast Cancer. *Oncotargets and Therapy*, 12, 11359. <https://doi.org/10.2147/OTT.S234328>
- Lopez-Bujanda, Z., & Drake, C. G. (2017). Myeloid-derived cells in prostate cancer progression: phenotype and prospective therapies. *Journal of Leukocyte Biology*, 102(2), 393–406. <https://doi.org/10.1189/JLB.5VMR1116-491RR>
- Loras, A., Martínez-Bisbal, M. C., Quintás, G., Gil, S., Martínez-Máñez, R., & Ruiz-Cerdá, J. L. (2019). Urinary Metabolic Signatures Detect Recurrences in Non-Muscle Invasive Bladder Cancer. *Cancers* 2019, Vol. 11, Page 914, 11(7), 914. <https://doi.org/10.3390/CANCERS11070914>
- Lou, H., Li, H., Yeager, M., Im, K., Gold, B., Schneider, T. D., Fraumeni, J. F., Chanock, S. J., Anderson, S. K., & Dean, M. (2012). Promoter variants in the MSMB gene associated with prostate cancer regulate MSMB/NCOA4 fusion transcripts. *Human Genetics*, 131(9), 1453. <https://doi.org/10.1007/S00439-012-1182-2>

- Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology*, 15(6). <https://doi.org/10.15252/MSB.20188746>
- Lyu, C., Vaddi, P. K., Elshafae, S., Pradeep, A., Ma, D., & Chen, S. (2025). Unveiling RACK1: a key regulator of the PI3K/AKT pathway in prostate cancer development. *Oncogene*, 44(5), 322–335. <https://doi.org/10.1038/S41388-024-03224-9>
- Ma, G., Tang, M., Wu, Y., Xu, X., Pan, F., & Xu, R. (2016). LncRNAs and miRNAs: potential biomarkers and therapeutic targets for prostate cancer. *American Journal of Translational Research*, 8(12), 5141. <https://pmc.ncbi.nlm.nih.gov/articles/PMC5209471/>
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5), 1202–1214. <https://doi.org/10.1016/J.CELL.2015.05.002>
- Mantovani, A., Sozzani, S., Locati, M., Allavena, P., & Sica, A. (2002). Macrophage polarization: Tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends in Immunology*, 23(11), 549–555. [https://doi.org/10.1016/S1471-4906\(02\)02302-5](https://doi.org/10.1016/S1471-4906(02)02302-5)
- McCarthy, D. J., Campbell, K. R., Lun, A. T. L., & Wills, Q. F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, 33(8), 1179–1186. <https://doi.org/10.1093/BIOINFORMATICS/BTW777>
- Montserrat-Ayuso, T., & Esteve-Codina, A. (2024). High content of nuclei-free low-quality cells in reference single-cell atlases: a call for more stringent quality control using nuclear fraction. *BMC Genomics*, 25(1), 1124-. <https://doi.org/10.1186/S12864-024-11015-5/TABLES/3>
- Nagy, J. A., Chang, S.-H., Shih, S.-C., Dvorak, A. M., & Dvorak, H. F. (2010). *Heterogeneity of the Tumor Vasculature*. <https://doi.org/10.1055/s-0030-1253454>
- Pan, J., Ma, Z., Liu, B., Qian, H., Shao, X., Liu, J., Wang, Q., & Xue, W. (2023). Identification of cancer-associated fibroblasts subtypes in prostate cancer. *Frontiers in Immunology*, 14. <https://doi.org/10.3389/FIMMU.2023.1133160>
- Patel, S. A., Hassan, M. K., & Dixit, M. (2024). Oncogenic activation of EEF1A2 expression: a journey from a putative to an established oncogene. *Cellular & Molecular Biology Letters*, 29(1), 6. <https://doi.org/10.1186/S11658-023-00519-9>
- Petitprez, F., de Reyniès, A., Keung, E. Z., Chen, T. W. W., Sun, C. M., Calderaro, J., Jeng, Y. M., Hsiao, L. P., Lacroix, L., Bougoüin, A., Moreira, M., Lacroix, G., Natario, I., Adam, J., Lucchesi, C., Laizet, Y., Toulmonde, M., Burgess, M. A., Bolejack, V., ... Fridman, W. H. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature*, 577(7791), 556–560. <https://doi.org/10.1038/S41586-019-1906-8>
- Petitprez, F., Fossati, N., Vano, Y., Freschi, M., Becht, E., Lucianò, R., Calderaro, J., Guédet, T., Lacroix, L., Rancoita, P. M. V., Montorsi, F., Fridman, W. H., Sautès-Fridman, C.,

- Briganti, A., Doglioni, C., & Bellone, M. (2019). PD-L1 Expression and CD8+ T-cell Infiltrate are Associated with Clinical Progression in Patients with Node-positive Prostate Cancer. *European Urology Focus*, 5(2), 192–196. <https://doi.org/10.1016/J.EUF.2017.05.013>
- Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* 2013 9:1, 9(1), 171–181. <https://doi.org/10.1038/nprot.2014.006>
- Pinzani, P., Lind, K., Malentacchi, F., Nesi, G., Salvianti, F., Villari, D., Kubista, M., Pazzagli, M., & Orlando, C. (2008). Prostate-specific antigen mRNA and protein levels in laser microdissected cells of human prostate measured by real-time reverse transcriptase-quantitative polymerase chain reaction and immuno-quantitative polymerase chain reaction. *Human Pathology*, 39(10), 1474–1482. <https://doi.org/10.1016/j.humpath.2008.02.012>
- Pliner, H. A., Shendure, J., & Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. *Nature Methods* 2019 16:10, 16(10), 983–986. <https://doi.org/10.1038/s41592-019-0535-3>
- Rago, V., Perri, A., & Di Agostino, S. (2023). New Therapeutic Perspectives in Prostate Cancer: Patient-Derived Organoids and Patient-Derived Xenograft Models in Precision Medicine. *Biomedicines*, 11(10). <https://doi.org/10.3390/BIOMEDICINES11102743>
- Ramsköld, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebtukova, I., Loring, J. F., Laurent, L. C., Schroth, G. P., & Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* 2012 30:8, 30(8), 777–782. <https://doi.org/10.1038/nbt.2282>
- Rawla, P. (2019). Epidemiology of Prostate Cancer. *World Journal of Oncology*, 10(2), 63. <https://doi.org/10.14740/WJON1191>
- Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., ... Yosef, N. (2017). The Human Cell Atlas. *ELife*, 6. <https://doi.org/10.7554/ELIFE.27041>
- Rohan Ramnarine, V., Kobelev, M., Gibb, E. A., Nouri, M., Lin, D., Wang, Y., Buttyan, R., Davicioni, E., Zoubeidi, A., Collins, C. C., & Zabor, E. (2019). *The evolution of long noncoding RNA acceptance in prostatecancer initiation, progression, and its clinical utility in disease management*. 560–561. <https://doi.org/10.1016/j.eururo.2019.07.040>
- Ruffell, B., & Coussens, L. M. (2015). Macrophages and therapeutic resistance in cancer. *Cancer Cell*, 27(4), 462–472. <https://doi.org/10.1016/J.CCELL.2015.02.015>
- Sancho, D., Mourão-Sá, D., Joffre, O. P., Schulz, O., Rogers, N. C., Pennington, D. J., Carlyle, J. R., & Sousa, C. R. (2008). Tumor therapy in mice via antigen targeting to a novel, DC-restricted C-type lectin. *The Journal of Clinical Investigation*, 118(6), 2098–2110. <https://doi.org/10.1172/JCI34584>

- Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M. J., Blüthgen, N., & Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications* 2017 9:1, 9(1), 20-. <https://doi.org/10.1038/s41467-017-02391-6>
- Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22(6), 839–851. <https://doi.org/10.1261/RNA.053959.115/-/DC1>
- Sena, L. A., Li, S., Jairaman, A., Prakriya, M., Ezponda, T., Hildeman, D. A., Wang, C. R., Schumacker, P. T., Licht, J. D., Perlman, H., Bryce, P. J., & Chandel, N. S. (2013). Mitochondria are required for antigen-specific T cell activation through reactive oxygen species signaling. *Immunity*, 38(2), 225–236. <https://doi.org/10.1016/J.IMMUNI.2012.10.020>
- Song, H., Weinstein, H. N. W., Allegakoen, P., Wadsworth, M. H., Xie, J., Yang, H., Castro, E. A., Lu, K. L., Stohr, B. A., Feng, F. Y., Carroll, P. R., Wang, B., Cooperberg, M. R., Shalek, A. K., & Huang, F. W. (2022a). Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nature Communications*, 13(1). <https://doi.org/10.1038/S41467-021-27322-4>
- Song, H., Weinstein, H. N. W., Allegakoen, P., Wadsworth, M. H., Xie, J., Yang, H., Castro, E. A., Lu, K. L., Stohr, B. A., Feng, F. Y., Carroll, P. R., Wang, B., Cooperberg, M. R., Shalek, A. K., & Huang, F. W. (2022b). Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nature Communications* 2022 13:1, 13(1), 1–20. <https://doi.org/10.1038/s41467-021-27322-4>
- Squair, J. W., Gautier, M., Kathe, C., Anderson, M. A., James, N. D., Hutson, T. H., Hudelle, R., Qaiser, T., Matson, K. J. E., Barraud, Q., Levine, A. J., La Manno, G., Skinnider, M. A., & Courtine, G. (2021). Confronting false discoveries in single-cell differential expression. *Nature Communications* 2021 12:1, 12(1), 5692-. <https://doi.org/10.1038/s41467-021-25960-2>
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R., & Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* 2017 14:9, 14(9), 865–868. <https://doi.org/10.1038/nmeth.4380>
- Stone, L. (2017). Escaping enzalutamide: Malat1 contributes to resistance. *Nature Reviews Urology* 2017 14:8, 14(8), 450–450. <https://doi.org/10.1038/nrurol.2017.91>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalex, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019a). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888-1902.e21. <https://doi.org/10.1016/J.CELL.2019.05.031>

- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019b). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888-1902.e21. <https://doi.org/10.1016/J.CELL.2019.05.031>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019c). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888-1902.e21. <https://doi.org/10.1016/J.CELL.2019.05.031>
- Stuart, T., & Satija, R. (2019a). Integrative single-cell analysis. *Nature Reviews Genetics* 2019 20:5, 20(5), 257–272. <https://doi.org/10.1038/s41576-019-0093-7>
- Stuart, T., & Satija, R. (2019b). Integrative single-cell analysis. *Nature Reviews Genetics* 2019 20:5, 20(5), 257–272. <https://doi.org/10.1038/s41576-019-0093-7>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), 15545–15550. <https://doi.org/10.1073/PNAS.0506580102>
- Sun, X., Cheng, G., Hao, M., Zheng, J., Zhou, X., Zhang, J., Taichman, R. S., Pienta, K. J., & Wang, J. (2010). CXCL12 / CXCR4 / CXCR7 chemokine axis and cancer progression. *Cancer Metastasis Reviews*, 29(4), 709–722. <https://doi.org/10.1007/S10555-010-9256-X>
- 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>
- Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* 2019 9:1, 9(1), 1–12. <https://doi.org/10.1038/s41598-019-41695-z>
- Wang, I., Song, L., Wang, B. Y., Kalebasty, A. R., Uchio, E., & Zi, X. (2022). Prostate cancer immunotherapy: a review of recent advancements with novel treatment methods and efficacy. *American Journal of Clinical and Experimental Urology*, 10(4), 210. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9428569/>
- Wang, Y., & Wang, Y. (2025). Transcription Factor Dynamics in Neuroendocrine Prostate Cancer Development. *Holistic Integrative Oncology*, 4(1), 65. <https://doi.org/10.1007/S44178-025-00197-X>
- Watson, P. A., Arora, V. K., & Sawyers, C. L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nature Reviews Cancer* 2015 15:12, 15(12), 701–711. <https://doi.org/10.1038/nrc4016>
- Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J. R., Bartonicek, N., Wang, T., Larsson, L., Kaczorowski, D., Weisenfeld, N. I., Uytingco, C. R., Chew, J. G., Bent, Z. W., Chan, C. L., Gnanasambandapillai, V., ... Swarbrick, A. (2021). A single-cell and spatially resolved

atlas of human breast cancers. *Nature Genetics* 2021 53:9, 53(9), 1334–1347. <https://doi.org/10.1038/s41588-021-00911-1>

Xu, D., Wang, W., Wang, D., Ding, J., Zhou, Y., & Zhang, W. (2024). Long noncoding RNA MALAT-1: A versatile regulator in cancer progression, metastasis, immunity, and therapeutic resistance. *Non-Coding RNA Research*, 9(2), 388–406. <https://doi.org/10.1016/J.NCRNA.2024.01.015>

Xu, H., Wang, F., Li, H., Ji, J., Cao, Z., Lyu, J., Shi, X., Zhu, Y., Zhang, C., Guo, F., Fang, Z., Yang, B., & Sun, Y. (2019). Prostatic Acid Phosphatase (PAP) Predicts Prostate Cancer Progress in a Population-Based Study: The Renewal of PAP? *Disease Markers*, 2019. <https://doi.org/10.1155/2019/7090545>

Yang, D., Lu, X., Zhang, X., Zhang, X., Zhu, L., & Liu, Q. (2024). RACK1 promotes the occurrence and progression of cervical carcinoma. *Journal of Clinical Laboratory Analysis*, 38(4), e25012. <https://doi.org/10.1002/JCLA.25012>

Zhang, D., Fang, C., Li, H., Lu, C., Huang, J., Pan, J., Yang, Z., Liang, E., Liu, Z., Zhou, X., Xin, Z., Chen, Y., & Cai, Q. (2021). Long ncRNA MALAT1 promotes cell proliferation, migration, and invasion in prostate cancer via sponging miR-145. *Translational Andrology and Urology*, 10(6), 2307–2319. <https://doi.org/10.21037/TAU-20-1526>

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8. <https://doi.org/10.1038/NCOMMS14049>

## 9. Anexos.

### Contenido

|        |   |     |
|--------|---|-----|
| 9.1.   | Acceso al código fuente.....  | 90  |
| 9.2.   | Control de calidad de las muestras (QC). .....                      | 90  |
| 9.2.1. | Tabla resumen global de QC. ....,                                   | 90  |
| 9.2.2. | Gráficos por muestra.....   | 91  |
| 9.3.   | Decisión de resolución de clústeres.....                            | 98  |
| 9.4.   | TOP10 genes por clúster.....  | 101 |
| 9.5.   | Criterio seguido para caracterización de poblaciones celulares..... | 106 |
| 9.6.   | Bibliografía.....   | 116 |

### 9.1. Acceso al código fuente.

El código R utilizado para el análisis completo está disponible en el siguiente repositorio:  
<https://github.com/guachinwey/prostate-scrNAseq-tfm>.

### 9.2. Control de calidad de las muestras (QC).

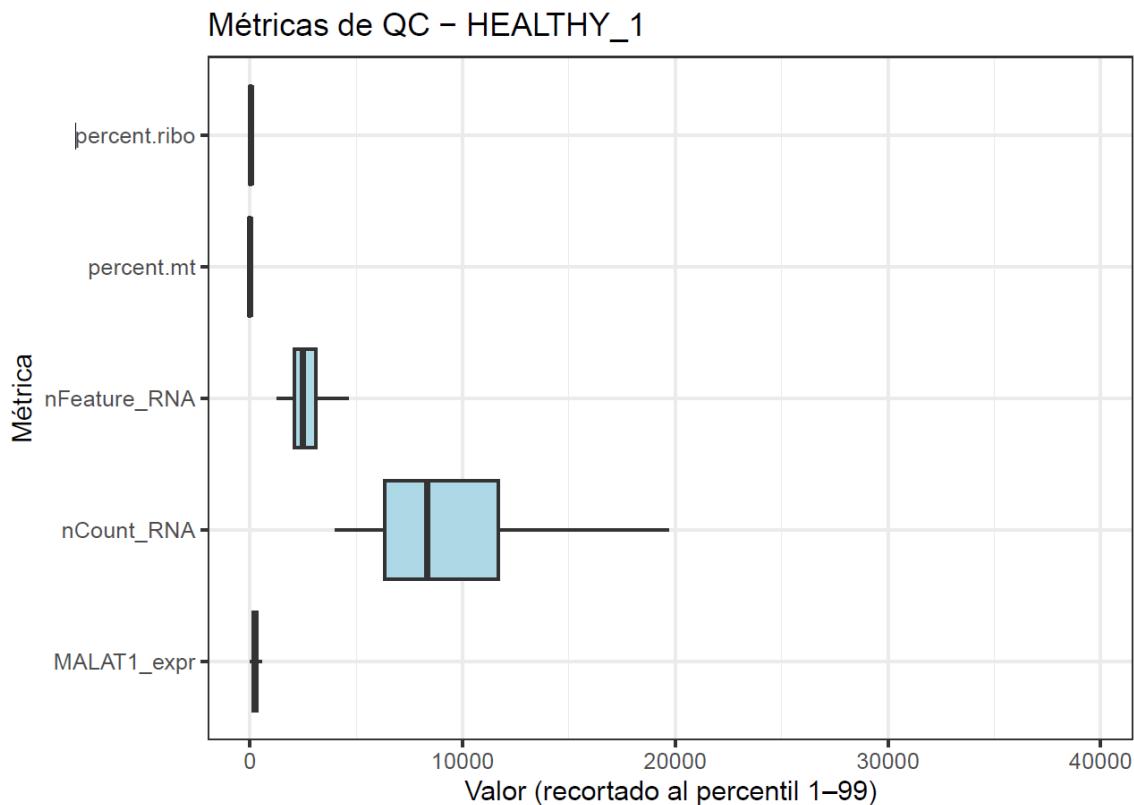
#### 9.2.1. Tabla resumen global de QC.

Tabla 1A: Resultados por muestra tras el control de calidad.

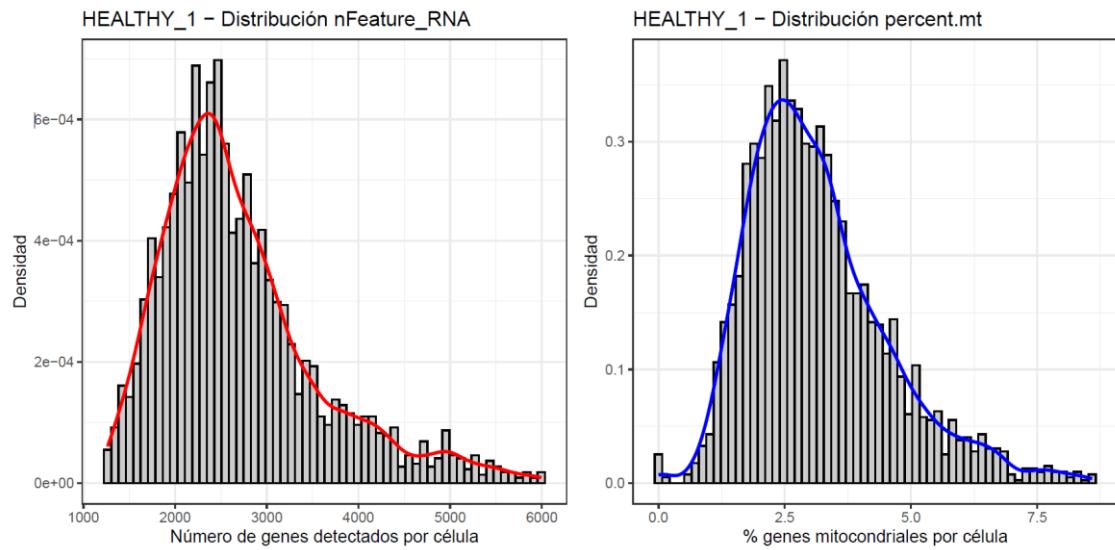
| Sample    | Cells_before | Cells_after | Percent_kept | nFeature_p1 | nFeature_p99 | max_percent_mt_used | MALAT1_p99 | Ribo_p99 |
|-----------|--------------|-------------|--------------|-------------|--------------|---------------------|------------|----------|
| HEALTHY_1 | 2825         | 2716        | 96,14        | 1258,68     | 6699         | 8,59                | 1119,04    | 51,19    |
| HEALTHY_2 | 2303         | 2228        | 96,74        | 1002,02     | 6119,04      | 15                  | 963,64     | 49,65    |
| HEALTHY_3 | 1600         | 1461        | 91,31        | 1008,96     | 6412,73      | 15                  | 1457,26    | 0        |
| PCA       | 5000         | 4904        | 98,08        | 274         | 5053,02      | 0                   | 1728,19    | 60,27    |
| CRPC_1    | 4078         | 3188        | 78,18        | 155         | 8005,92      | 15                  | 2685,3     | 39,98    |
| CRPC_2    | 8084         | 7276        | 90           | 337,83      | 5194,87      | 15                  | 1260,17    | 34,88    |
| CRPC_3    | 7716         | 7146        | 92,61        | 295,15      | 5736,4       | 15                  | 1112,7     | 40,16    |

## 9.2.2. Gráficos por muestra.

Muestra *HEALTHY\_1*.



*Figura 1A: Boxplot de métricas de calidad para Healthy 1.*



*Figura 2A: Histograma de densidad para las métricas de calidad de Healthy 1.*

Muestra HEALTHY\_2.

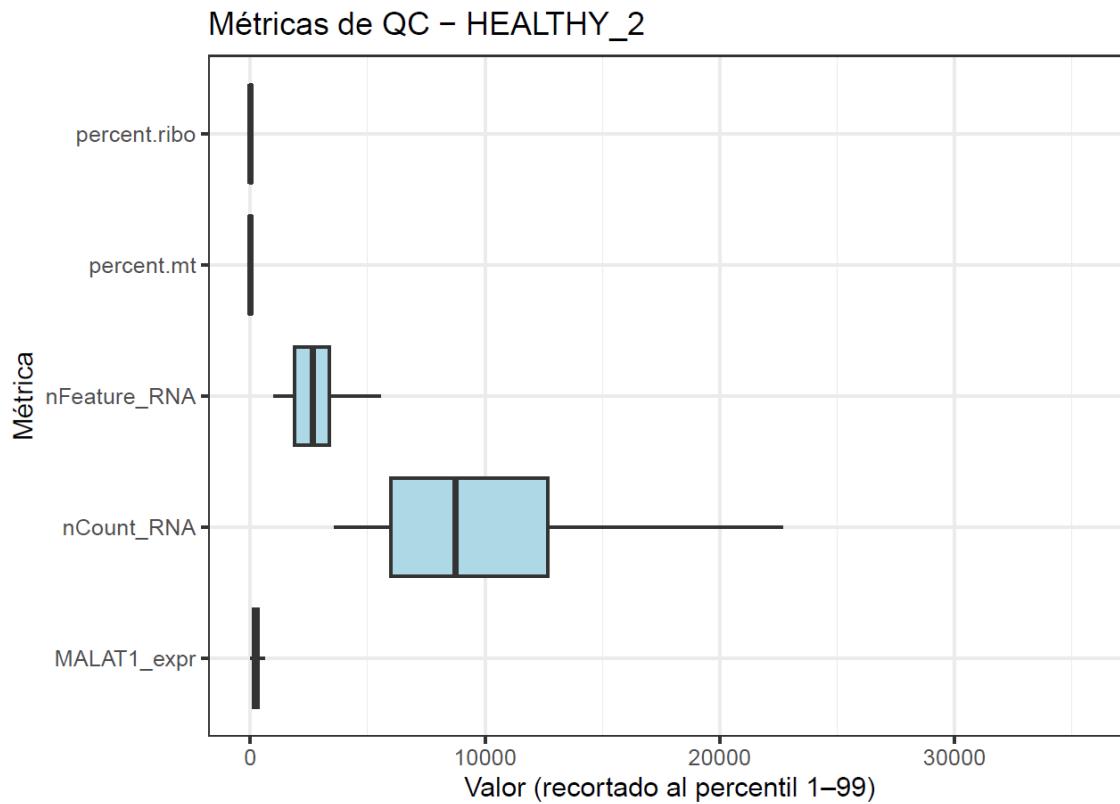


Figura 3A: Boxplot de métricas de calidad para Healthy 2.

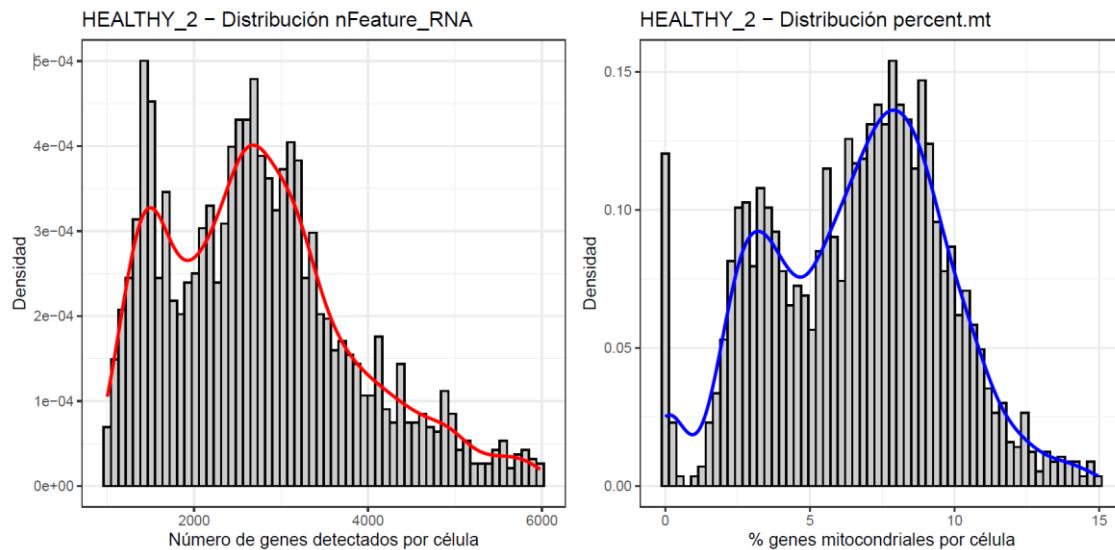


Figura 54A: Histograma de densidad para las métricas de calidad de Healthy 2.

Muestra HEALTHY\_3.

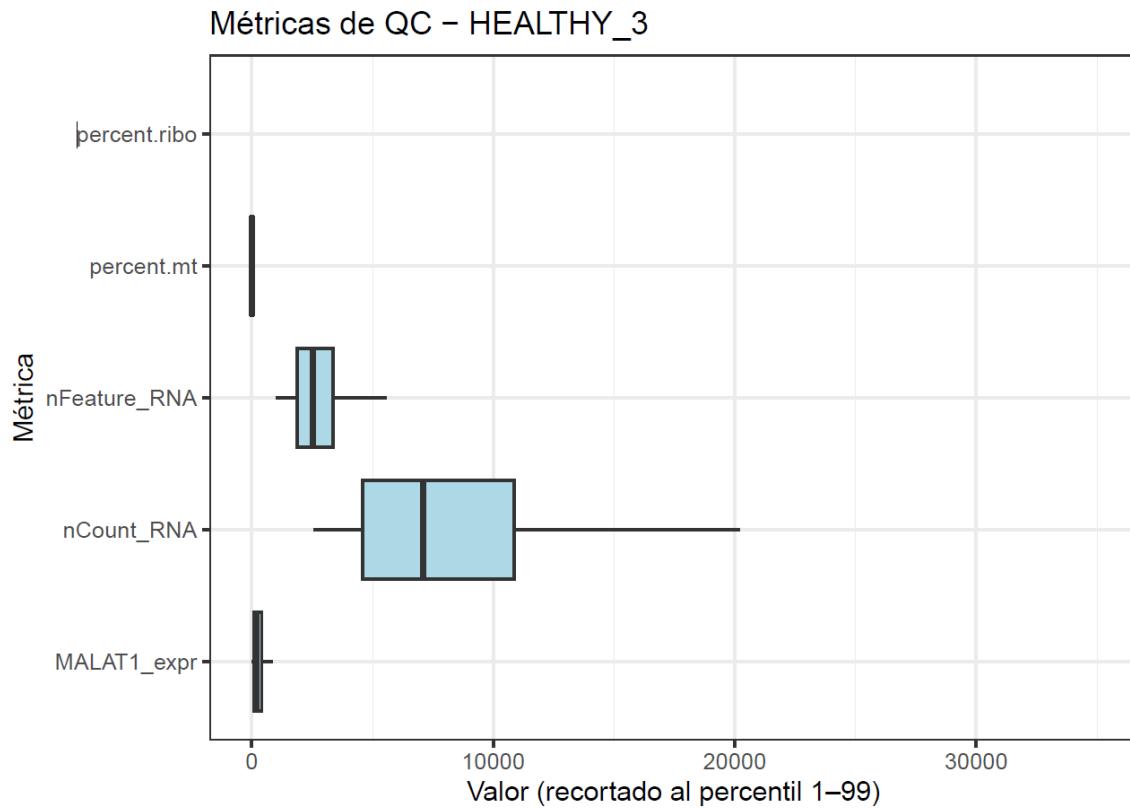


Figura 5A: Boxplot de métricas de calidad para Healthy 3.

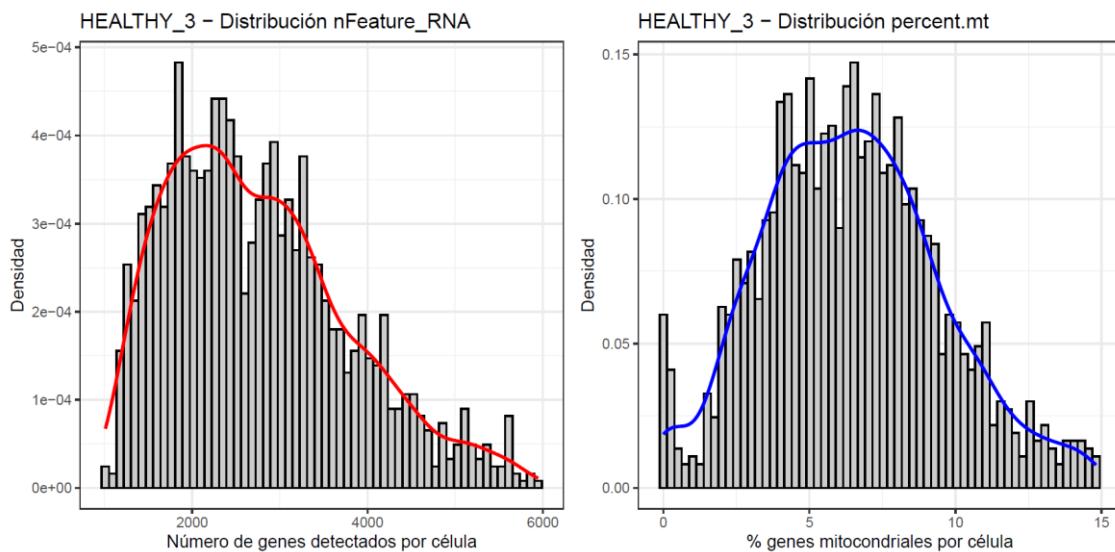


Figura 6A: Histograma de densidad para las métricas de calidad de Healthy 3.

Muestra PCa.

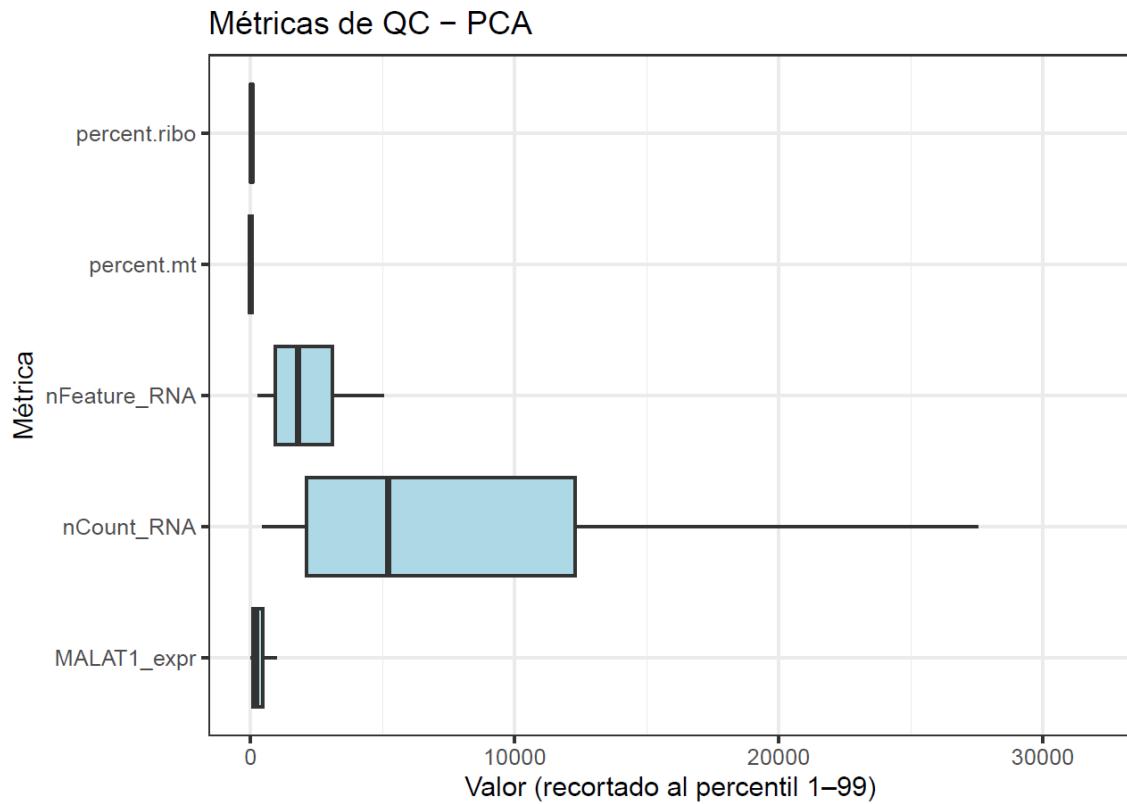


Figura 7A: Boxplot de métricas de calidad para PCa.

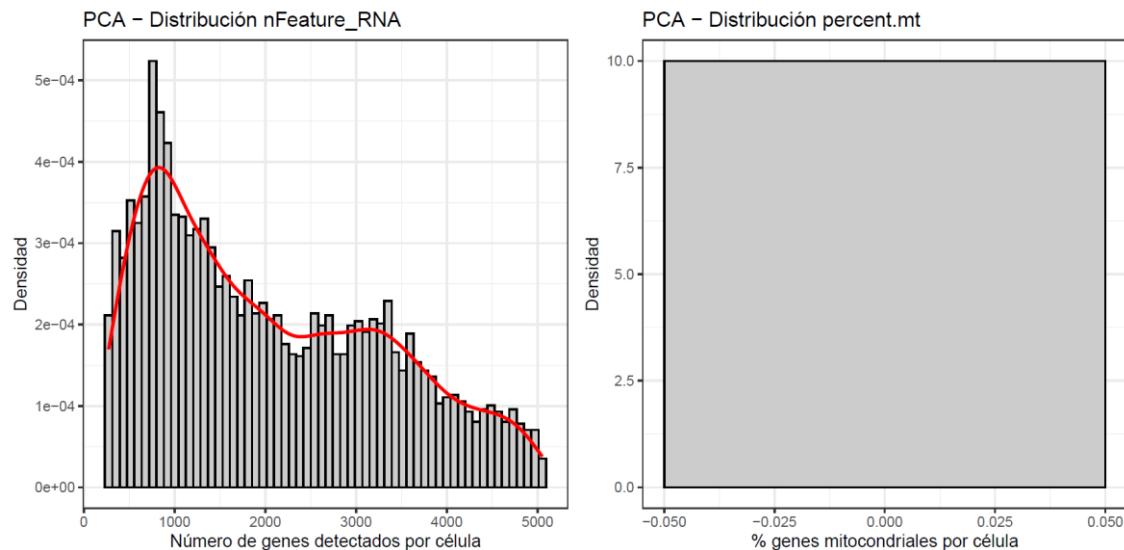


Figura 8A: Histograma de densidad para las métricas de calidad de PCa.

Muestra CRPC\_1.

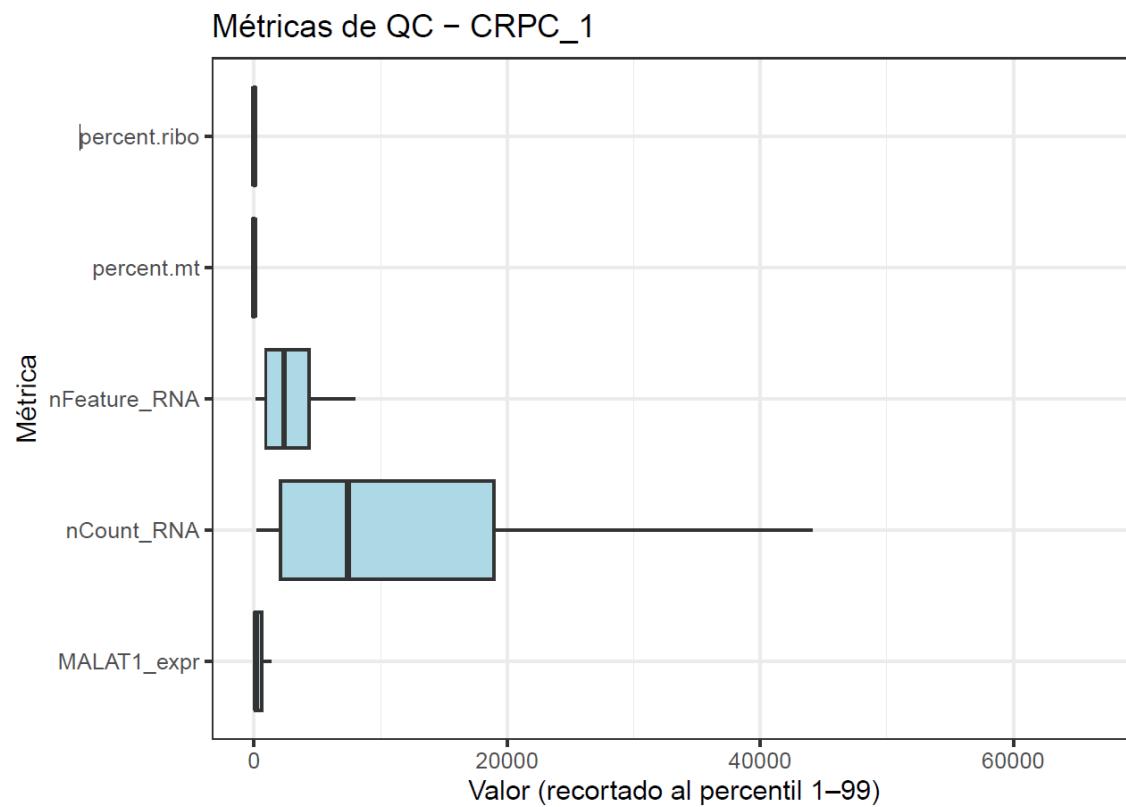


Figura 9A: Boxplot de métricas de calidad para CRPC 1.

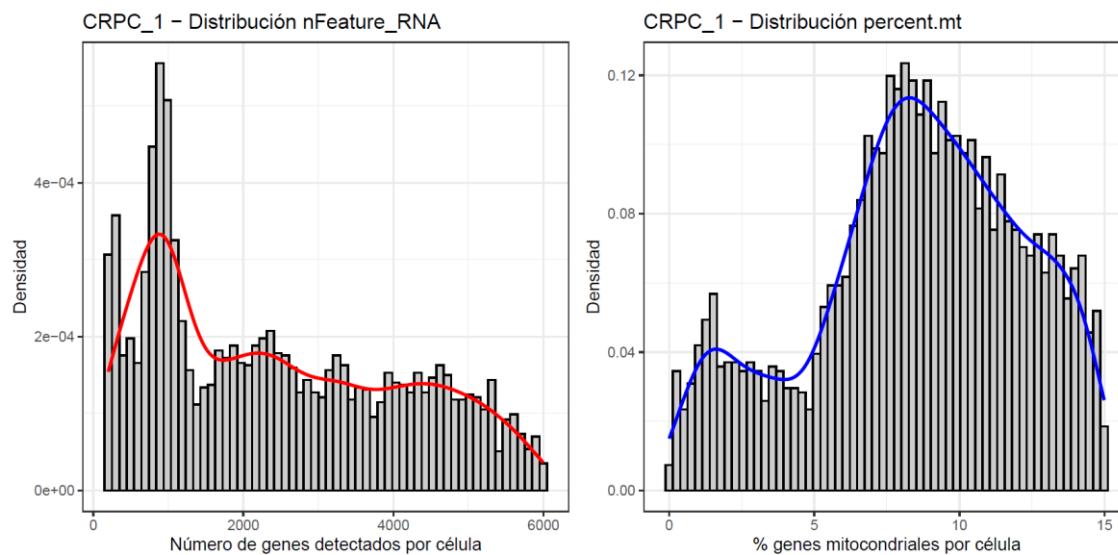


Figura 10A: Histograma de densidad para las métricas de calidad de CRPC 1.

Muestra CRPC\_2.

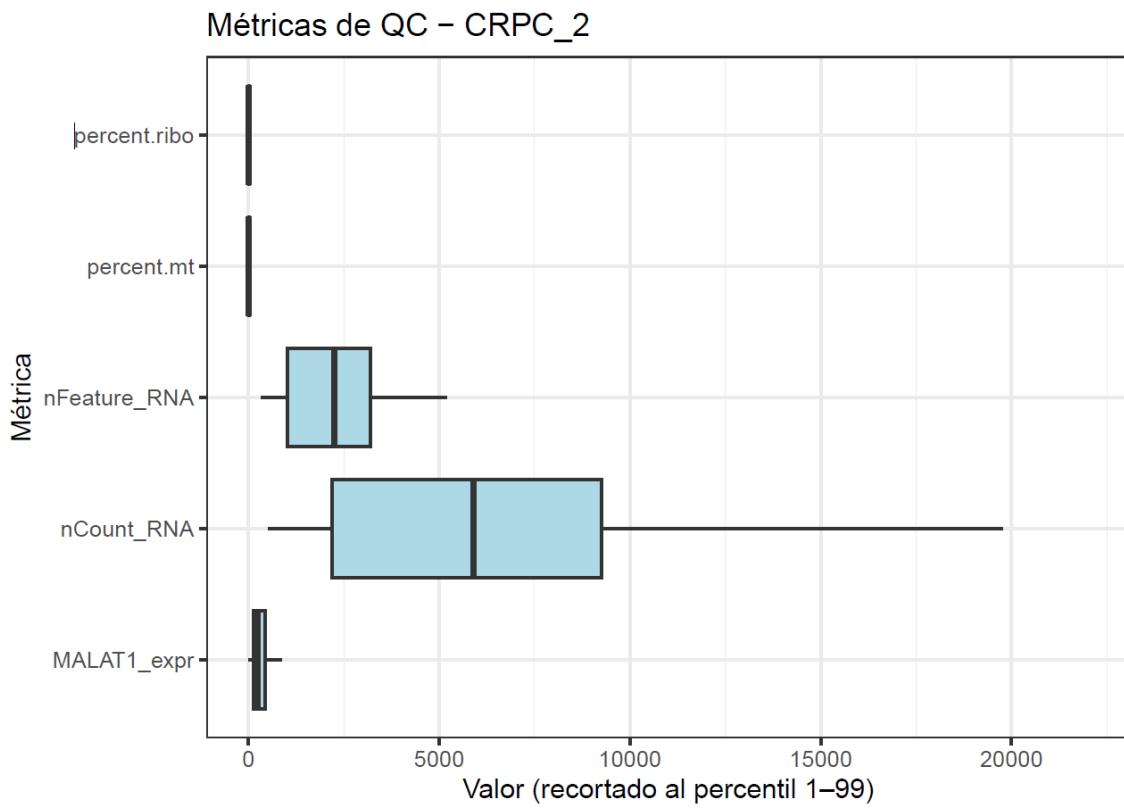


Figura 11A: Boxplot de métricas de calidad para CRPC 2.

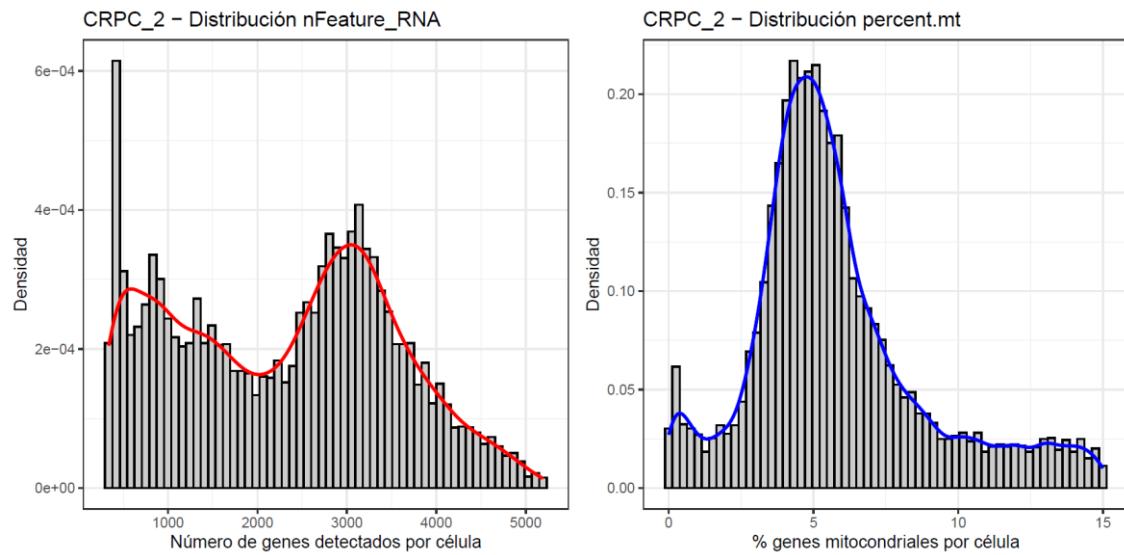


Figura 12A: Histograma de densidad para las métricas de calidad de CRPC 2.

Muestra CRPC\_3.

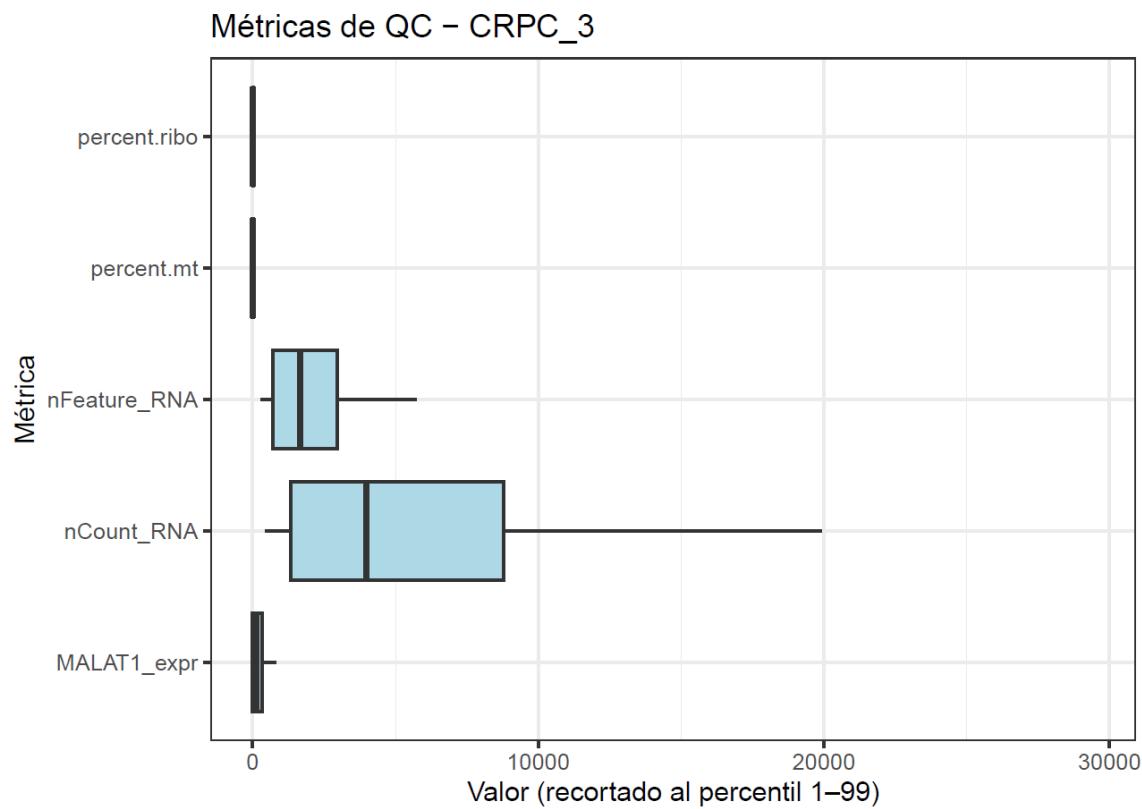


Figura 13A: Boxplot de métricas de calidad para CRPC 3.

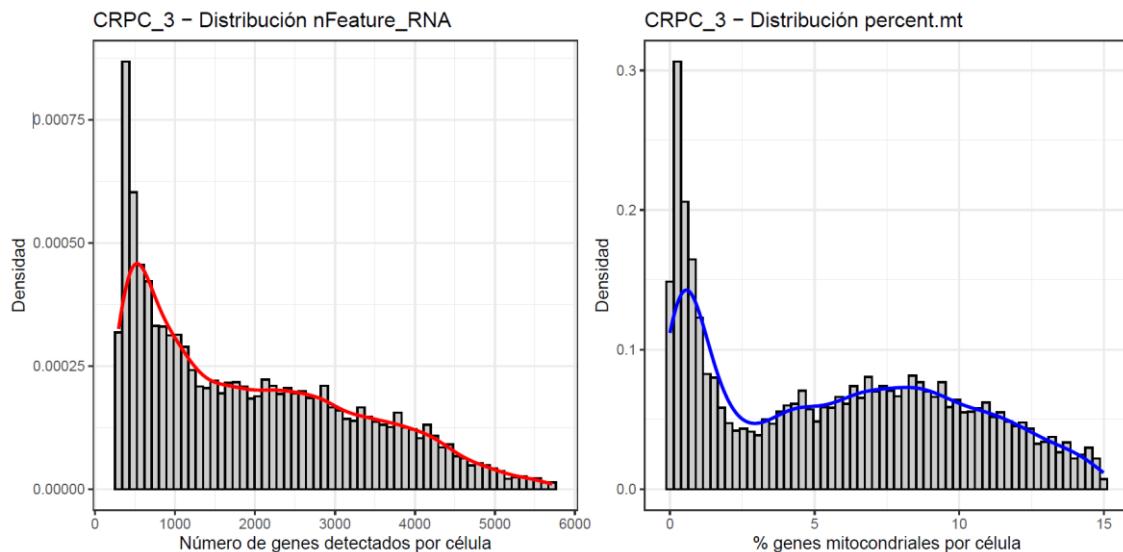


Figura 14A: Histograma de densidad para las métricas de calidad de CRPC 3.

### 9.3. Decisión de resolución de clústeres.

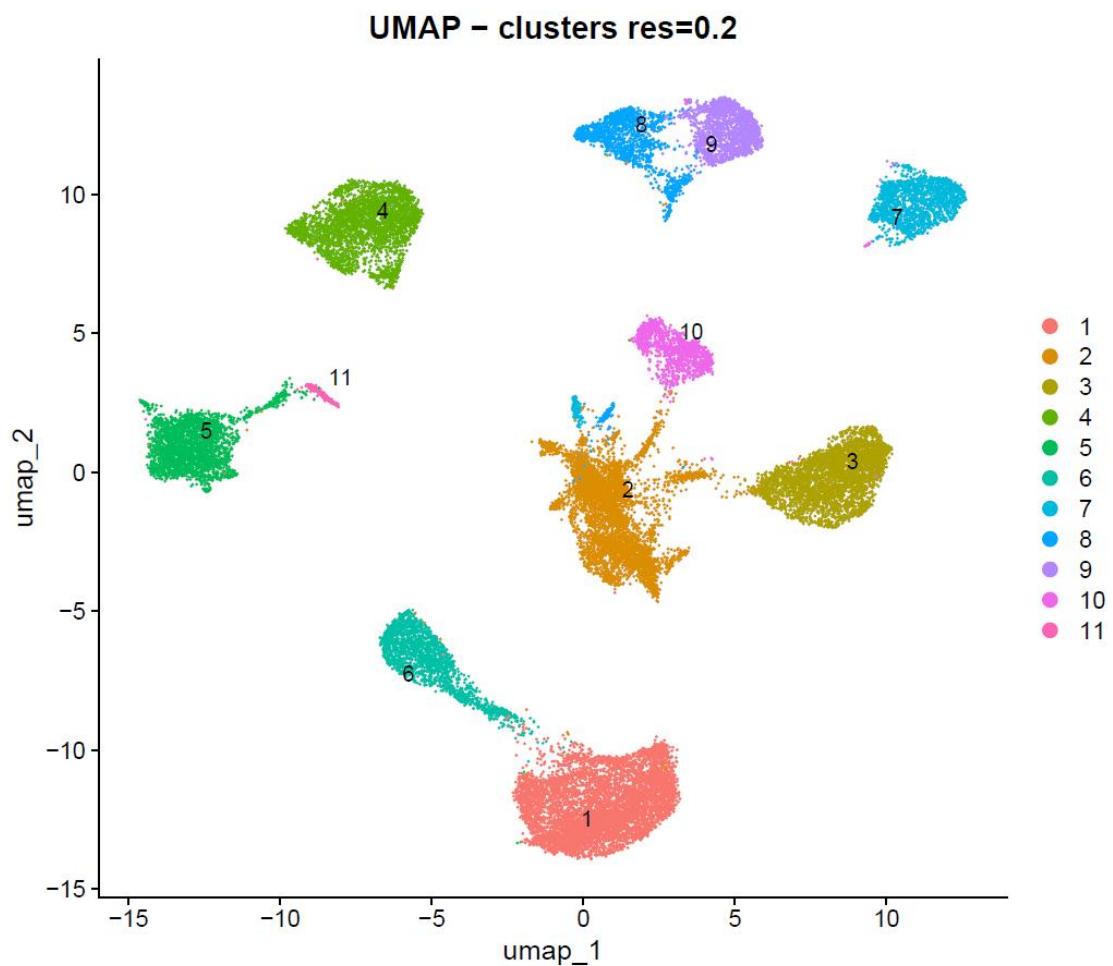


Figura 15A: Proyección UMAP de las células tras clustering con resolución 0.2.

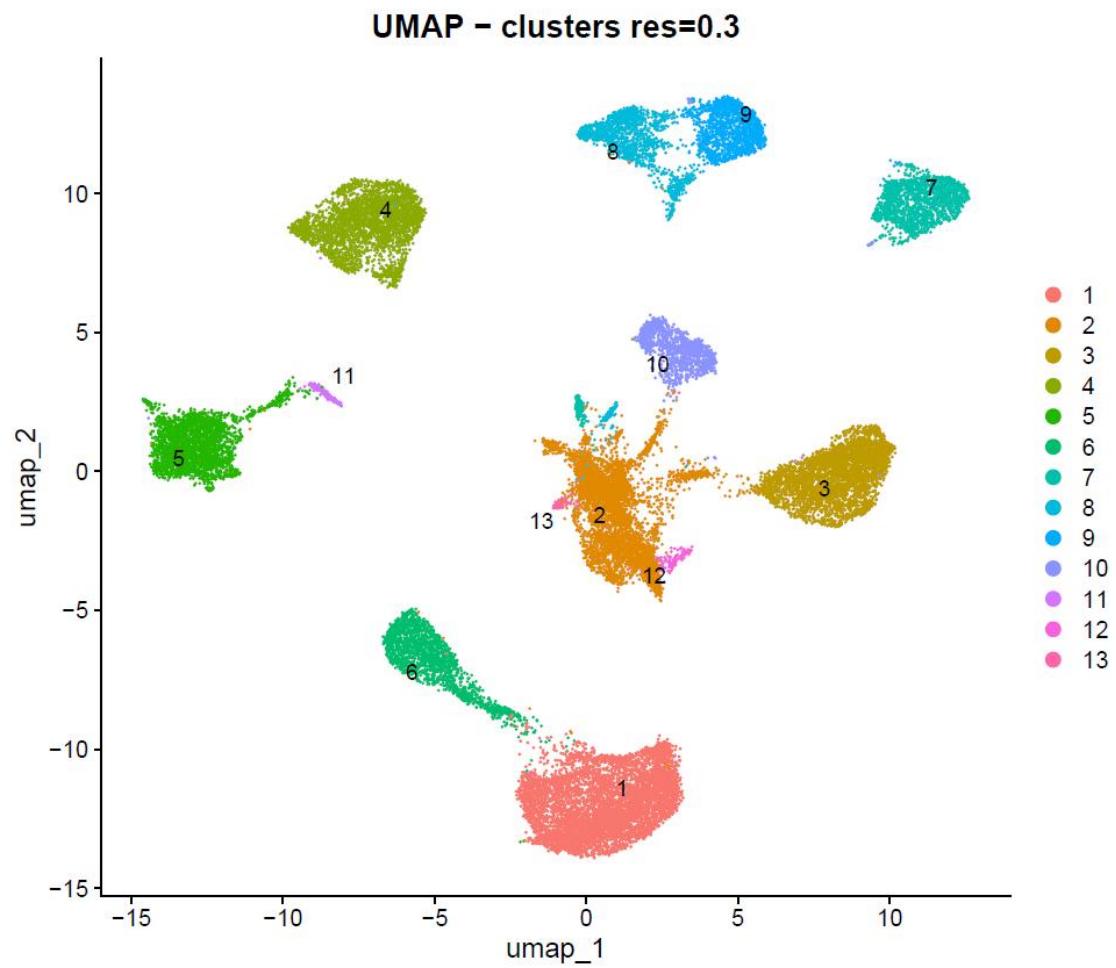


Figura 16A: Proyección UMAP de las células tras clustering con resolución 0.3.

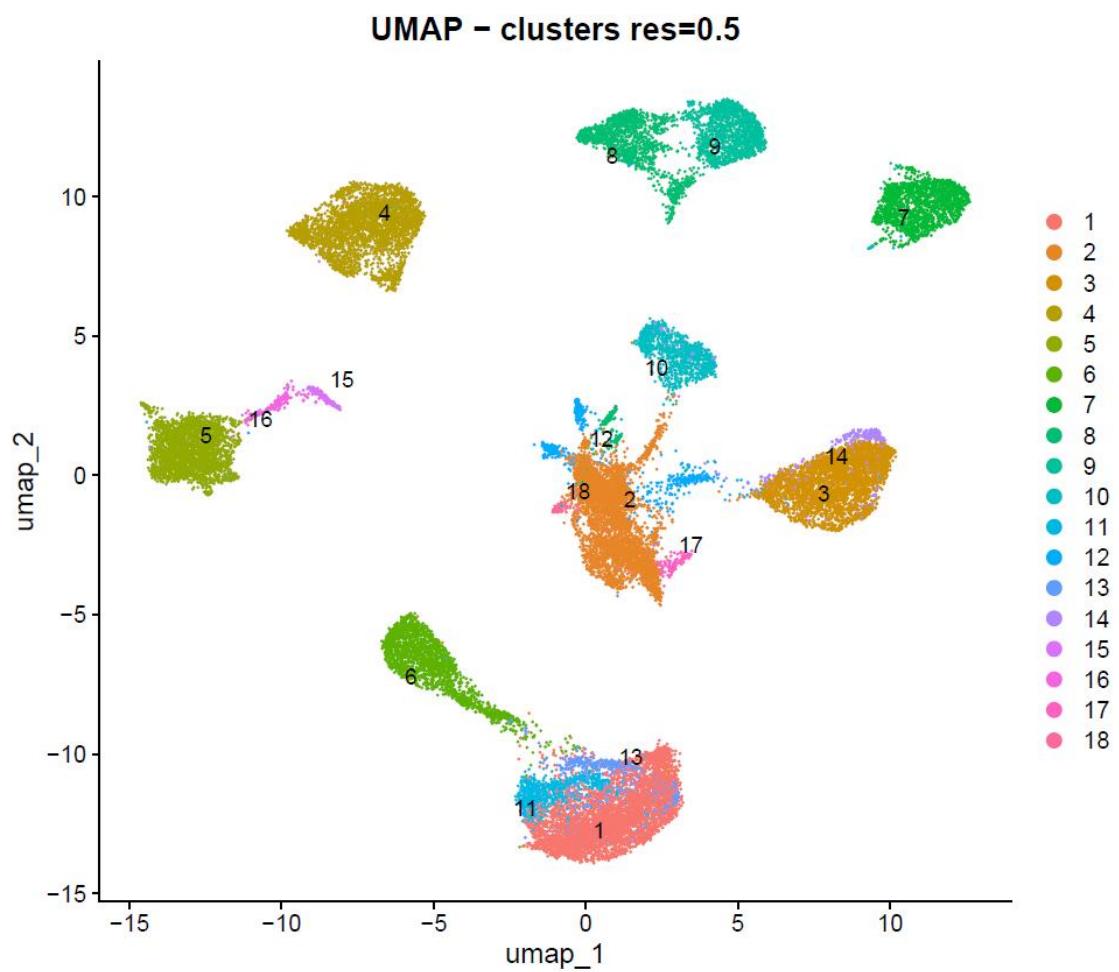


Figura 17A: Proyección UMAP de las células tras clustering con resolución 0.5.

## 9.4. TOP10 genes por clúster.

Tabla 2A: Top 10 genes característicos por clúster.

| p_val    | avg_log2FC | pct.1 | pct.2 | p_val_adj | cluster | gene      |
|----------|------------|-------|-------|-----------|---------|-----------|
| 0        | 7,045621   | 0,298 | 0,004 | 0         | 1       | CD177     |
| 0        | 6,362626   | 0,955 | 0,153 | 0         | 1       | ACPP      |
| 0        | 6,242845   | 0,326 | 0,005 | 0         | 1       | LMAN1L    |
| 0        | 6,022318   | 0,396 | 0,011 | 0         | 1       | SLC22A3   |
| 0        | 5,864158   | 0,304 | 0,007 | 0         | 1       | CPLX3     |
| 0        | 5,646233   | 0,8   | 0,193 | 0         | 1       | MSMB      |
| 0        | 5,598046   | 0,589 | 0,043 | 0         | 1       | NEFH      |
| 0        | 5,44425    | 0,37  | 0,014 | 0         | 1       | PEBP4     |
| 0        | 5,438759   | 0,266 | 0,01  | 0         | 1       | CCK       |
| 0        | 5,425701   | 0,286 | 0,009 | 0         | 1       | PCGEM1    |
| 0        | 3,434213   | 0,502 | 0,106 | 0         | 2       | EEF1A2    |
| 0        | 2,151316   | 0,594 | 0,349 | 0         | 2       | IFT57     |
| 0        | 1,927261   | 0,676 | 0,24  | 0         | 2       | TRGC1     |
| 3,9E-290 | 1,8679     | 0,46  | 0,232 | 8,9E-286  | 2       | HSD17B4   |
| 1,2E-220 | 1,658825   | 0,609 | 0,455 | 2,7E-216  | 2       | ENSA      |
| 0        | 1,601728   | 0,986 | 0,933 | 0         | 2       | RPLP0     |
| 3,7E-136 | 1,601243   | 0,326 | 0,184 | 8,6E-132  | 2       | MAOA      |
| 2,6E-216 | 1,596652   | 0,521 | 0,337 | 6E-212    | 2       | MAP1B     |
| 2,1E-279 | 1,593429   | 0,555 | 0,334 | 4,8E-275  | 2       | IMPDH2    |
| 8,5E-242 | 1,573658   | 0,606 | 0,412 | 2E-237    | 2       | BASP1     |
| 0        | 4,752744   | 0,391 | 0,015 | 0         | 3       | SYNPR     |
| 0        | 4,641729   | 0,351 | 0,016 | 0         | 3       | KRT20     |
| 0        | 4,608528   | 0,259 | 0,01  | 0         | 3       | GSX1      |
| 0        | 4,550468   | 0,376 | 0,017 | 0         | 3       | RAET1L    |
| 0        | 4,407454   | 0,296 | 0,013 | 0         | 3       | LINC01198 |
| 0        | 4,281669   | 0,588 | 0,042 | 0         | 3       | CEACAM6   |
| 0        | 4,269602   | 0,552 | 0,035 | 0         | 3       | SCN3B     |
| 0        | 4,168101   | 0,254 | 0,013 | 0         | 3       | LIPC      |
| 0        | 4,140889   | 0,527 | 0,034 | 0         | 3       | FGF3      |
| 0        | 4,127666   | 0,367 | 0,022 | 0         | 3       | RFX4      |
| 0        | 7,739704   | 0,732 | 0,023 | 0         | 4       | C1QC      |
| 0        | 7,664857   | 0,257 | 0,002 | 0         | 4       | IL1A      |
| 0        | 7,611794   | 0,78  | 0,01  | 0         | 4       | MS4A7     |
| 0        | 7,562631   | 0,769 | 0,034 | 0         | 4       | C1QA      |
| 0        | 7,546844   | 0,749 | 0,036 | 0         | 4       | C1QB      |
| 0        | 7,545471   | 0,567 | 0,006 | 0         | 4       | IGSF6     |
| 0        | 7,521093   | 0,566 | 0,006 | 0         | 4       | OLR1      |
| 0        | 7,44887    | 0,514 | 0,003 | 0         | 4       | CD86      |
| 0        | 7,428095   | 0,258 | 0,001 | 0         | 4       | LILRB2    |

|   |          |       |       |   |   |            |
|---|----------|-------|-------|---|---|------------|
| 0 | 7,419969 | 0,427 | 0,003 | 0 | 4 | LILRB4     |
| 0 | 7,384588 | 0,6   | 0,005 | 0 | 5 | CD3G       |
| 0 | 7,352668 | 0,516 | 0,011 | 0 | 5 | TRBC1      |
| 0 | 7,288551 | 0,779 | 0,012 | 0 | 5 | CD3D       |
| 0 | 7,264195 | 0,701 | 0,009 | 0 | 5 | CD3E       |
| 0 | 7,245815 | 0,286 | 0,005 | 0 | 5 | AC092580.4 |
| 0 | 7,238669 | 0,429 | 0,009 | 0 | 5 | KLRB1      |
| 0 | 7,095633 | 0,788 | 0,011 | 0 | 5 | CD2        |
| 0 | 6,995003 | 0,53  | 0,009 | 0 | 5 | GZMA       |
| 0 | 6,98169  | 0,269 | 0,005 | 0 | 5 | GZMK       |
| 0 | 6,960193 | 0,71  | 0,016 | 0 | 5 | TRAC       |
| 0 | 8,144456 | 0,642 | 0,02  | 0 | 6 | KRT17      |
| 0 | 7,391976 | 0,465 | 0,011 | 0 | 6 | KRT14      |
| 0 | 7,08981  | 0,367 | 0,003 | 0 | 6 | DLK2       |
| 0 | 6,800572 | 0,795 | 0,037 | 0 | 6 | KRT15      |
| 0 | 6,638596 | 0,719 | 0,01  | 0 | 6 | KRT5       |
| 0 | 6,602055 | 0,733 | 0,016 | 0 | 6 | SLC14A1    |
| 0 | 6,379378 | 0,27  | 0,003 | 0 | 6 | C16orf74   |
| 0 | 6,350809 | 0,297 | 0,004 | 0 | 6 | TP63       |
| 0 | 6,331444 | 0,488 | 0,007 | 0 | 6 | TRIM29     |
| 0 | 6,216219 | 0,619 | 0,013 | 0 | 6 | SCUBE2     |
| 0 | 8,160253 | 0,358 | 0,001 | 0 | 7 | NOVA2      |
| 0 | 8,035732 | 0,314 | 0,001 | 0 | 7 | ARHGEF15   |
| 0 | 7,934931 | 0,802 | 0,008 | 0 | 7 | VWF        |
| 0 | 7,889998 | 0,745 | 0,006 | 0 | 7 | PTPRB      |
| 0 | 7,838496 | 0,687 | 0,005 | 0 | 7 | PCDH17     |
| 0 | 7,813401 | 0,589 | 0,002 | 0 | 7 | MYCT1      |
| 0 | 7,672846 | 0,32  | 0,001 | 0 | 7 | FLT4       |
| 0 | 7,651177 | 0,573 | 0,003 | 0 | 7 | ROBO4      |
| 0 | 7,63506  | 0,528 | 0,003 | 0 | 7 | KCNN3      |
| 0 | 7,581975 | 0,744 | 0,005 | 0 | 7 | CD34       |
| 0 | 7,558389 | 0,4   | 0,02  | 0 | 8 | SFRP2      |
| 0 | 7,433757 | 0,422 | 0,015 | 0 | 8 | SFRP4      |
| 0 | 7,303665 | 0,269 | 0,002 | 0 | 8 | FNDC1      |
| 0 | 7,195044 | 0,27  | 0,004 | 0 | 8 | MFAP5      |
| 0 | 6,930683 | 0,704 | 0,043 | 0 | 8 | LUM        |
| 0 | 6,899966 | 0,265 | 0,003 | 0 | 8 | COL11A1    |
| 0 | 6,692808 | 0,253 | 0,004 | 0 | 8 | OMD        |
| 0 | 6,688295 | 0,892 | 0,14  | 0 | 8 | COL1A1     |
| 0 | 6,660595 | 0,534 | 0,011 | 0 | 8 | ISLR       |
| 0 | 6,64987  | 0,337 | 0,054 | 0 | 8 | APOD       |
| 0 | 7,516262 | 0,496 | 0,005 | 0 | 9 | RERGL      |
| 0 | 7,153348 | 0,641 | 0,005 | 0 | 9 | COX4I2     |
| 0 | 7,091326 | 0,395 | 0,003 | 0 | 9 | FAM162B    |
| 0 | 6,993421 | 0,253 | 0,003 | 0 | 9 | CASQ2      |

|          |          |       |       |          |    |             |
|----------|----------|-------|-------|----------|----|-------------|
| 0        | 6,846593 | 0,703 | 0,01  | 0        | 9  | HIGD1B      |
| 0        | 6,457504 | 0,961 | 0,086 | 0        | 9  | RGS5        |
| 0        | 6,451411 | 0,485 | 0,006 | 0        | 9  | CSPG4       |
| 0        | 6,376037 | 0,528 | 0,008 | 0        | 9  | RASL12      |
| 0        | 6,324795 | 0,322 | 0,012 | 0        | 9  | MYH11       |
| 0        | 6,307584 | 0,582 | 0,009 | 0        | 9  | CDH6        |
| 0        | 5,949937 | 0,403 | 0,01  | 0        | 10 | CENPA       |
| 0        | 5,942011 | 0,656 | 0,022 | 0        | 10 | UBE2C       |
| 0        | 5,901352 | 0,465 | 0,008 | 0        | 10 | SGO1        |
| 0        | 5,887775 | 0,746 | 0,032 | 0        | 10 | TOP2A       |
| 0        | 5,818599 | 0,445 | 0,01  | 0        | 10 | DLGAP5      |
| 0        | 5,796507 | 0,613 | 0,012 | 0        | 10 | PBK         |
| 0        | 5,694247 | 0,516 | 0,012 | 0        | 10 | NUF2        |
| 0        | 5,668843 | 0,428 | 0,009 | 0        | 10 | ESCO2       |
| 0        | 5,658584 | 0,606 | 0,015 | 0        | 10 | RRM2        |
| 0        | 5,61793  | 0,253 | 0,007 | 0        | 10 | PLK1        |
| 0        | 7,323534 | 0,321 | 0,002 | 0        | 11 | ABCC11      |
| 0        | 7,003833 | 0,432 | 0,004 | 0        | 11 | SERPINB11   |
| 0        | 5,767225 | 0,276 | 0,006 | 0        | 11 | CYP4F2      |
| 0        | 5,594827 | 0,695 | 0,029 | 0        | 11 | OGN         |
| 0        | 5,513896 | 0,598 | 0,033 | 0        | 11 | KLK12       |
| 0        | 4,727973 | 0,526 | 0,027 | 0        | 11 | KCNN2       |
| 0        | 4,480971 | 0,295 | 0,014 | 0        | 11 | SAMD5       |
| 0        | 4,257674 | 0,845 | 0,126 | 0        | 11 | KLK11       |
| 0        | 4,225664 | 0,439 | 0,023 | 0        | 11 | ST8SIA6-AS1 |
| 0        | 4,1477   | 0,455 | 0,029 | 0        | 11 | PCAT18      |
| 3,5E-100 | 2,15911  | 0,402 | 0,125 | 8,07E-96 | 12 | MIAT        |
| 2,4E-112 | 1,788371 | 1     | 0,992 | 5,5E-108 | 12 | MALAT1      |
| 2,62E-14 | 1,323503 | 0,315 | 0,226 | 6,06E-10 | 12 | TMEM2       |
| 1,13E-39 | 1,229624 | 0,49  | 0,271 | 2,62E-35 | 12 | DST         |
| 1,14E-39 | 1,141825 | 0,374 | 0,175 | 2,63E-35 | 12 | MGAT5       |
| 5,42E-49 | 1,123732 | 0,453 | 0,209 | 1,26E-44 | 12 | RASSF6      |
| 5,01E-27 | 1,104558 | 0,38  | 0,215 | 1,16E-22 | 12 | RAB11FIP1   |
| 2E-51    | 1,095344 | 0,408 | 0,172 | 4,64E-47 | 12 | KREMEN1     |
| 0,002167 | 1,033737 | 0,261 | 0,247 | 1        | 12 | BOD1L1      |
| 1,66E-31 | 1,010409 | 0,601 | 0,378 | 3,84E-27 | 12 | ELF3        |
| 0        | 7,383064 | 0,717 | 0,008 | 0        | 13 | C1orf64     |
| 0        | 6,596727 | 0,25  | 0,006 | 0        | 13 | GPC5-AS1    |
| 0        | 6,54642  | 0,6   | 0,014 | 0        | 13 | PHGR1       |
| 0        | 6,394768 | 0,39  | 0,005 | 0        | 13 | TMPRSS11E   |
| 0        | 6,269175 | 0,421 | 0,009 | 0        | 13 | PGC         |
| 0        | 5,657121 | 0,619 | 0,031 | 0        | 13 | RNF157      |
| 0        | 5,651108 | 0,548 | 0,019 | 0        | 13 | LINC00668   |
| 0        | 5,39737  | 0,452 | 0,027 | 0        | 13 | SFTPA2      |
| 1,2E-306 | 5,093225 | 0,687 | 0,141 | 2,8E-302 | 13 | CST1        |

|          |          |       |       |          |    |               |
|----------|----------|-------|-------|----------|----|---------------|
| 0        | 5,059165 | 0,26  | 0,005 | 0        | 13 | CST4          |
| 0        | 10,07164 | 0,595 | 0,001 | 0        | 14 | FGF4          |
| 0        | 9,590296 | 0,361 | 0,003 | 0        | 14 | DKK4          |
| 0        | 8,500098 | 0,547 | 0,002 | 0        | 14 | LIX1          |
| 0        | 8,182428 | 0,462 | 0,004 | 0        | 14 | ITLN2         |
| 0        | 7,623481 | 0,332 | 0,005 | 0        | 14 | TMEM35A       |
| 0        | 7,439725 | 0,31  | 0,002 | 0        | 14 | RNF182        |
| 0        | 7,067402 | 0,405 | 0,004 | 0        | 14 | SSTR1         |
| 0        | 6,97024  | 0,399 | 0,004 | 0        | 14 | PNMT          |
| 0        | 6,903594 | 0,772 | 0,025 | 0        | 14 | FGF19         |
| 0        | 6,901715 | 0,272 | 0,003 | 0        | 14 | KCNIP4-IT1    |
| 0        | 11,82913 | 0,993 | 0,005 | 0        | 15 | TPSAB1        |
| 0        | 11,3     | 0,997 | 0,012 | 0        | 15 | TPSB2         |
| 0        | 11,23919 | 0,807 | 0,001 | 0        | 15 | MS4A2         |
| 0        | 10,9281  | 0,97  | 0,003 | 0        | 15 | CPA3          |
|          |          |       |       |          |    | RP11-354E11.2 |
| 0        | 10,38522 | 0,329 | 0     | 0        | 15 |               |
| 0        | 7,898133 | 0,645 | 0,006 | 0        | 15 | RGS13         |
| 0        | 7,856924 | 0,289 | 0,002 | 0        | 15 | C1orf186      |
| 0        | 7,621004 | 0,455 | 0,003 | 0        | 15 | IL1RL1        |
| 0        | 7,11645  | 0,638 | 0,013 | 0        | 15 | HPGDS         |
| 0        | 6,639728 | 0,684 | 0,018 | 0        | 15 | KIT           |
| 0        | 11,54637 | 0,277 | 0     | 0        | 16 | FCRLA         |
| 0        | 9,682432 | 0,699 | 0,002 | 0        | 16 | MS4A1         |
| 0        | 8,586805 | 0,468 | 0,003 | 0        | 16 | VPREB3        |
| 0        | 8,330642 | 0,277 | 0,001 | 0        | 16 | CD19          |
| 0        | 8,290568 | 0,295 | 0,002 | 0        | 16 | JCHAIN        |
| 0        | 7,354419 | 0,306 | 0,001 | 0        | 16 | BLK           |
| 0        | 7,298698 | 0,757 | 0,019 | 0        | 16 | CD79A         |
| 2,1E-295 | 7,137493 | 0,624 | 0,045 | 5E-291   | 16 | IGKC          |
| 0        | 7,109409 | 0,364 | 0,007 | 0        | 16 | IGHM          |
| 3,8E-186 | 6,173458 | 0,277 | 0,013 | 8,8E-182 | 16 | IGLC2         |
| 4,4E-185 | 4,462883 | 0,315 | 0,014 | 1E-180   | 17 | SERPINB11     |
| 0        | 4,405971 | 0,769 | 0,043 | 0        | 17 | OGN           |
| 3,2E-250 | 4,329016 | 0,643 | 0,045 | 7,4E-246 | 17 | KLK12         |
| 8,3E-295 | 4,129731 | 0,594 | 0,032 | 1,9E-290 | 17 | ST8SIA6-AS1   |
| 3,59E-46 | 3,759855 | 0,252 | 0,034 | 8,33E-42 | 17 | CRISP3        |
| 1,2E-118 | 3,54196  | 0,42  | 0,039 | 2,7E-114 | 17 | PCAT18        |
| 4,95E-98 | 3,518371 | 0,35  | 0,032 | 1,15E-93 | 17 | C19orf81      |
| 1,8E-126 | 3,506612 | 0,434 | 0,039 | 4,2E-122 | 17 | KCNN2         |
| 2,4E-149 | 3,421823 | 0,531 | 0,049 | 5,6E-145 | 17 | THSD7A        |
| 2,24E-81 | 3,306932 | 0,308 | 0,03  | 5,18E-77 | 17 | ACSM1         |
| 1,5E-247 | 4,674554 | 0,856 | 0,059 | 3,5E-243 | 18 | LINC01088     |
| 8,78E-74 | 3,671964 | 0,66  | 0,116 | 2,03E-69 | 18 | MIPEP         |
| 5,52E-31 | 3,566998 | 0,258 | 0,038 | 1,28E-26 | 18 | FGF12         |

|          |          |       |       |          |    |           |
|----------|----------|-------|-------|----------|----|-----------|
| 4,33E-65 | 3,470248 | 0,567 | 0,092 | 1E-60    | 18 | TRPM8     |
| 1,65E-74 | 3,408504 | 0,268 | 0,018 | 3,82E-70 | 18 | DSCAM-AS1 |
| 2,5E-185 | 3,407164 | 0,907 | 0,083 | 5,7E-181 | 18 | PCA3      |
| 2,86E-24 | 3,122956 | 0,392 | 0,104 | 6,62E-20 | 18 | SYCE1L    |
| 1,69E-27 | 2,840289 | 0,495 | 0,15  | 3,91E-23 | 18 | ABCC4     |
| 2,4E-64  | 2,802595 | 0,99  | 0,383 | 5,57E-60 | 18 | TFF3      |
| 1,55E-70 | 2,695764 | 1     | 0,299 | 3,6E-66  | 18 | TRGC1     |

## 9.5. Criterio seguido para caracterización de poblaciones celulares.

Tabla 3A: Criterios seguidos por clúster para caracterizar las poblaciones celulares.

| N.º Clúster. | Genes característicos (TOP10).   | Herramientas.   | Población asignada.           | Justificación.  |
|--------------|--|---|-------------------------------|---|
| 1            | CD177, ACPP, LMAN1L, SLC22A3, CPLX3, MSMB, NEFH, PEB4, CCK, PCGEM1.    | CellMarker 2.0 (Hu et al., 2023b), literatura.                            | Benign luminal epithelial.    | ACPP es un marcador prostático clásico asociado a células epiteliales luminales diferenciadas (Z. Tang et al., 2017; <i>The Human Protein Atlas</i> , n.d.). La herramienta CellMarker identifica ACPP con alta especificidad para células luminales epiteliales. Aunque algunos genes no arrojaron resultados directos, el perfil global y la distribución de condiciones (presencia relevante de Healthy y PCa) apoyan una identidad epitelial benigna.   |
| 2            | EEF1A2, IFT57, TRGC1, HSD17B4, ENSA RPLP0, MAOA, MAP1B, IMPDH2, BASP1. | Human Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. | Luminal metabolic epithelial. | Los genes presentes no fueron reconocidos como marcadores por CellMarker 2.0, por lo que se consultó su expresión tisular y celular en Human Protein Atlas. EEF1A2 y HSD17B4 se expresan en células epiteliales prostáticas con funciones metabólicas ( <i>The Human Protein Atlas</i> , n.d.). Además, su sobreexpresión se ha asociado con cáncer de próstata avanzado (Patel et al., 2024; Worst et al., 2017). La predominancia de células CRPC sugiere una reprogramación luminal epitelial hacia un estado metabólicamente activo (Zeng & Jiang, 2025). |
| 3            | SYNPR, KRT20, GSX1, RAET1L, LINC01198, CEACAM6,                        | Human Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. | Secretory luminal epithelial. | La presencia de genes como KRT20 y CEACAM6 en este clúster sugiere un perfil relacionado con epitelios glandulares, particularmente aquellos presentes en tejidos   |

|   |   |  |              |   |
|---|---|--|--------------|---|
|   | SCN3B, LIPC, FGF3, RFX4.  |  |              | <p>como el tracto gastrointestinal, donde estos genes se han descrito en contextos de diferenciación y actividad secretora (<i>The Human Protein Atlas</i>, n.d.). Aunque no son marcadores específicos de epitelio prostático, su expresión podría reflejar fenómenos de plasticidad o transdiferenciación epitelial.</p> <p>Por otro lado, la vía de señalización FGF/FGFR ha sido implicada en la regulación de la diferenciación y progresión tumoral en cáncer de próstata, lo que apoya un posible rol de genes como FGF3 en este contexto (Corn et al., 2013). Además, reprogramaciones del metabolismo lipídico son características bien documentadas del cáncer prostático avanzado (Z. Zhang et al., 2023).</p> <p>La baja representación de células sanas y la elevada proporción de células CRPC dentro del clúster refuerzan la hipótesis de una alteración del fenotipo secretor en un contexto tumoral avanzado.</p> |
| 4 | C1QC, IL1A, MS4A7, C1QA, C1QB, IGSF6, OLR1, CD86, LILRB2, LILRB4. | CellMarker 2.0 (Hu et al., 2023b), literatura. | Macrophages. | <p>CellMarker 2.0 identificó con alta especificidad a C1QA y C1QB como marcadores de macrófagos. Además, otros genes presentes, como MS4A7A, CD86 y LILRB4, han sido descritos como altamente expresados en macrófagos o asociados a su función inmunitaria (Nirmal et al., 2018).</p> <p>La composición del clúster muestra predominancia de células CRPC, seguida de un porcentaje mucho menor de Healthy, lo que sugiere que esta población incluye tanto macrófagos residentes como</p>   |

|   |   |  |                    |   |
|---|---|--|--------------------|---|
|   |   |  |                    | TAMs (tumor-associated macrophages). Estos últimos son conocidos por infiltrarse en tumores avanzados, incluyendo el cáncer de próstata resistente a castración, donde promueven evasión inmune y progresión tumoral (Mantovani et al., 2017).  |
| 5 | CD3G, TRBC1, CD3C, AC092580.4, KLRB1, CD2, GZMA, GZMK, TRAC.              | CellMarker 2.0 (Hu et al., 2023b), literatura. | T cells.           | CellMarker 2.0 identificó de manera robusta los genes CD3G, CD3D, CD3E y CD2 como marcadores característicos de células T, lo que queda reflejado en una puntuación de coincidencia elevada para los TOP10 genes (0.67). Además, genes como TRB1C, TRAC, GZMA y GZMK son conocidos como componentes clave del receptor de células T y de la maquinaria citotóxica, respaldando aún más esta asignación (Chifman et al., 2016).  |
| 6 | KRT17, KRT14, DLK2, KRT15, KRT5, SLC14A1, C16orf74, TP63, TRIM29, SCUBE2. | CellMarker 2.0 (Hu et al., 2023b), literatura. | Basal epithelial.  | El clúster muestra una expresión destacada de genes clásicos de células basales como KRT5, KRT115, KRT15 y TP63, ampliamente descritos como marcadores canónicos de células epiteliales basales prostáticas (Henry et al., 2018; Johnson et al., n.d.). La herramienta CellMarker 2.0 identifica específicamente estos genes en poblaciones de células basales. La distribución del clúster indica una predominancia de células Healthy, con un porcentaje menor de PCa y ausencia de CRPC, lo que refuerza la hipótesis de una población epitelial basal no transformada y disminuida en estadios avanzados de cáncer de próstata. |
| 7 | NOVA2, ARHGEF15, VWF, PTPRB, PCDH17,                                      | CellMarker 2.0 (Hu et al., 2023b), Human       | Endothelial cells. | El gen VWF fue identificado por CellMarker 2.0 como específico de endotelial cells, ya que es un reconocido marcador endotelial.  |

|   |   |   |                                  |  |
|---|---|---|----------------------------------|--|
|   | MYCT1, FLT4, ROBO4, KCNN3, CD34.                                  | Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. |                                  | Además, genes como PTPRB y FLT4 están implicados en funciones vasculares y linfangiogénesis, siendo típicamente expresados por células endoteliales ( <i>The Human Protein Atlas</i> , n.d.). La distribución de condiciones muestra una dominancia de células CRPC, con un porcentaje significativo de PCa y menor contribución en Healthy, lo cual podría reflejar un aumento de la vascularización tumoral en estadios avanzados de cáncer de próstata (Ji et al., 2023; Trojan et al., 2004).  |
| 8 | SFRP2, SFRP4, FNDC1, MFAP5, LUM, COL11A1, OMD, COL1A1, ISLR, APOD | CellMarker 2.0 (Hu et al., 2023b), literatura.                      | CAFs.                            | Los genes COL1A1, COL11A1, LUM y MFAP5 están asociados clásicamente con fibroblastos activados y matriz extracelular. La herramienta CellMarker 2.0 identificó un marcador (APOD) con afinidad específica a fibroblastos. Además, múltiples estudios han documentado la sobreexpresión de estos genes con fibroblastos asociados a cáncer (CAFs) en cáncer de próstata y otros tumores sólidos(Wu et al., 2021; J. Zhang et al., 2023). La casi exclusividad del clúster en la condición CRPC sugiere un papel importante de estas células en la progresión tumoral y en la reprogramación del microambiente prostático resistente a la castración, consistente con funciones descritas para CAFs, como la promoción de la invasión tumoral, remodelación tisular y resistencia a tratamientos hormonales (Öhlund et al., 2014; Pan et al., 2023). |
| 9 | RERGL, COX4I2, FAM162B,   | CellMarker 2.0 (Hu et al., 2023b),                                  | Perycites / smooth muscle cells. | Del conjunto de genes presentes, solo MYH11 y RGS fueron identificados como marcadores   |

|    |   |   |                                      |  |
|----|---|---|--------------------------------------|--|
|    | CASQ2,<br>HIGD1B,<br>RGS5, CSPG4,<br>RASL12,<br>MYH11, CDH6   | Human<br>Protein Atlas<br>( <i>The Human<br/>Protein Atlas</i> ,<br>n.d.),<br>literatura.   |                                      | de fibroblastos en CellMarker 2.0. Para los demás genes, se consultó The Human Protein Atlas, donde se observó expresión específica en pericitos y células de músculo liso vascular, lo que sugiere que esta población representa células en la línea mural vascular. La predominancia de células provenientes de la condición CRPC podría indicar un aumento de células de soporte vasculatura asociadas al microambiente tumoral avanzado, consistentes con funciones de remodelación celular, contracción y posible interacción con células tumorales (Ji et al., 2023; Trojan et al., 2004). |
| 10 | CENPA,<br>UBE2C, SGO1,<br>TOP2A,<br>DLGAP5, PBK,<br>NUF2, ESCO2,<br>RRM2, PLK1                          | CellMarker<br>2.0 (Hu et al.,<br>2023b),<br>Human<br>Protein Atlas<br>( <i>The Human<br/>Protein Atlas</i> ,<br>n.d.),<br>literatura. | Cycling /<br>proliferating<br>cells. | Este cluster present alta expresión en genes fuertemente asociados con el ciclo celular y la proliferación activa, como TOP2A, CENPA, UBE2C, PLK1 y RRM2. Ninguno de estos genes fue identificado como marcador específico en CellMarker 2.0. Sin embargo, su expresión tisular y celular en The Human Protein Atlas los vincula con células en división activa, especialmente en contextos tumorales. La alta proporción de células CRPC sugiere un estado proliferativo agresivo, consistente con el fenotipo observado en etapas avanzadas del cáncer de próstata resistente a castración.    |
| 11 | ABCC11,<br>SERPINB11,<br>CYP4F2, OGN,<br>KLK12,<br>KCNN2,<br>SAMD5,<br>KLK11,<br>ST8SIA6-AS1,<br>PCAT18 | Human<br>Protein Atlas<br>( <i>The Human<br/>Protein Atlas</i> ,<br>n.d.),<br>literatura.   | Luminal<br>KLK11/12+.                | Este clúster fue clasificado como epitelial luminal con expresión de KLK11/12 debido a la presencia de genes del grupo calicreína, típicamente asociados al epitelio glandular prostático. En particular, KLK11 presenta expresión a nivel transcriptómico en células glandulares y  |

|    |  |             |                     |   |
|----|--|-------------|---------------------|---|
|    |  |             |                     | <p>luminales de la próstata, con perfiles elevados en adenocarcinoma prostático comparado con otros tejidos, lo que respalda su carácter epitelial luminal en contexto tumoral (<i>The Human Protein Atlas</i>, n.d.). Aunque KLK12 no muestra una categoría de expresión clara en próstata en Human Protein Atlas, miembros de la familia de calicreínas como grupo están implicados en funciones de secreción epitelial y en la biología del cáncer de próstata, incluida potencial influencia en progresión tumoral y regulación del microambiente (Boyukozer et al., 2020).</p> <p>La predominancia de este clúster en PCa, con menor presencia en Healthy y muy marginal en CRPC, apoya que representa un subgrupo de células epiteliales luminas diferenciadas con actividad específica en calicreínas.</p> |
| 12 | MIAT, MALAT1, TMEM2, DST, MGAT5, RASSF6, RAB11FIP1, KREMEN1, BOD1L1, ELF3. | Literatura. | Luminal<br>MALAT1+. | <p>Este cluster se caracterizó por una elevada expresión de MALAT y otros genes epiteliales como ELF3, DST y TMEM2, lo que sugiere una identidad luminal epitelial diferenciada, posiblemente asociada a estados avanzados del cáncer prostático. MALAT1 (Metastasis associated lung adenocarcoma transcript 1) es un lncRNA ampliamente descrito como promotor de la progresión tumoral, metástasis y resistencia terapéutica en varios tipos de cáncer, incluido el CRPC (D. Zhang et al., 2021b). La sobreexpresión de MALAT1m ha sido vinculada con proliferación, migración y evasión inmune en células tumorales prostáticas, y su inhibición reduce el</p>   |

|    |  |   |                            |   |
|----|--|---|----------------------------|---|
|    |  |   |                            | crecimiento tumoral en modelos preclínicos (D. Xu et al., 2024). La presencia predominante de este clúster en muestras de CRPC (>95%) y su escasa presencia en PCa y Healthy refuerzan su asociación con estados avanzados del cáncer.  |
| 13 | C1orf64, GPC5-AS1, PHGR1, TMPRSS11E, PGC, RNF157, LINC00668, SFTPA2, CST1, CST4. | Human Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. | Mucinous luminal.          | Aunque CellMarker 2.0 no clasificó directamente estos genes como marcadores específicos de células prostáticas, el análisis de expresión en Human Protein Atlas indica que varios de ellos (PHGR1, TMPRSS11E, CST1) están enriquecidos en epitelio glandular de órganos secretores, lo que es compatible con una identidad luminal mucinosa. Particularmente, CST1 y CST4 codifican cistatinas con funciones en la regulación de proteasas en secreciones glandulares, y han sido reportados en contextos de adenocarcinomas con diferenciación mucinosa, como los de pulmón o tracto gastrointestinal (Liu & Yao, 2019; Uhlén et al., 2015; J. Yang et al., 2022). La distribución de este clúster, altamente enriquecido en PCa, sugiere una subpoblación epitelial diferenciada con actividad secretora. |
| 14 | FGF4, DKK4, LIX1, ITLN2, TMEM35A, RNF182, SSTR1, PNMT, FGF19, KCNIP4-IT1.        | Human Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. | Neuroendocrine / EMT-like. | CellMarker 2.0 no da una clasificación directa de estos genes. No obstante, varios de ellos, como SSTR1, DKK4, PNMT o FGF4, presentan patrones de expresión compatibles con tejidos neuroendocrinos o de transición epitelio-mesénquima (EMT) según <i>The Human Protein Atlas</i> . Este perfil, junto con la alta representación en CRPC y la   |

|    |  |   |             |   |
|----|--|---|-------------|---|
|    |  |   |             | literatura que describe la emergencia de fenotipos neuroendocrinos y EMT-like en estadios avanzados y resistentes del cáncer de próstata, sugiere la presencia de esta subpoblación, posiblemente asociada a mecanismos de progresión tumoral y resistencia terapéutica (Beltran et al., 2019; Yamada & Beltran, 2021).   |
| 15 | TPSAB1,<br>TPSB2,<br>MS4A2, CPA3,<br>RP11-<br>354E11.2,<br>RGS13,<br>C1orf186,<br>IL1RL1,<br>HPGDS, KIT. | Human<br>Protein Atlas<br>( <i>The Human<br/>Protein Atlas</i> ,<br>n.d.),<br>literatura.   | Mast cells. | No se logra una clasificación directa a partir de CellMarker 2.0, pero múltiples genes como TPSAB1, TPSB2, MS4A2, CPA3, RGS13, HPGDS y KIT son marcadores de mastocitos en diversos tejidos ( <i>The Human Protein Atlas</i> , n.d.). El clúster se encuentra principalmente en PCa y en menor proporción en CRPC. Este patrón sugiere que los mastocitos pueden estar implicados en etapas iniciales del microambiente tumoral, pero ser desplazados o suprimidos en fases avanzadas de la enfermedad, lo que es consistente con observaciones previas en cáncer de próstata (Fleischmann et al., 2009; Johansson et al., 2010). |
| 16 | FCRLA,<br>MS4A1,<br>VPREB3,<br>CD19,<br>JCHAIN, BLK,<br>CD79A, IGKC,<br>IGHM, IGLC2.                     | CellMarker<br>2.0 (Hu et al.,<br>2023b),<br>Human<br>Protein Atlas<br>( <i>The Human<br/>Protein Atlas</i> ,<br>n.d.),<br>literatura. | B cells.    | Este clúster expresa múltiples genes característicos de células B, como CD19, MS4A1, CD79A, IGKC, IGHM y JCHAIN, lo cual respalda su identidad como linfocitos B maduros o en estado activo. CellMarker 2.0 lo clasifica principalmente como “B cells” y “GR+ regulatory B cell”, y The Human Protein Atlas confirma la expresión de estos marcadores en tejidos linfoides (Monaco et al., 2019; Szabo et al., 2019). El clúster está presente en las tres condiciones analizadas, con una mayor proporción en PCa,   |

|    |   |   |                      |  |
|----|---|---|----------------------|--|
|    |   |   |                      | seguido de Healthy y CRPC en proporciones similares. Esta distribución podría reflejar un rol inmunomodulador de las células B en estadios tempranos del cáncer de próstata (Sharonov et al., 2020).   |
| 17 | SERPINB11, OGN, KLK12, ST8SIA6-AS1, CRISP3, PCAT18, C19orf81, KCNN2, THSD7A, ACSM1. | Human Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. | CRISP3-high luminal. | CellMarker 2.0 no proporciona una clasificación directa, pero el patrón de expresión observado en The Human Protein Atlas sugiere una identidad luminal prostática especializada. CRISP3 (Cysteine-rich secretory protein 3), en particular, ha sido descrito como marcador de células epiteliales prostáticas con expresión aumentada en adenocarcinomas de próstata, y se asocia con una diferenciación secretora dentro del epitelio glandular (Fagerberg et al., 2014; Pathak et al., 2016). La presencia exclusiva de este clúster en muestras PCa, y su ausencia tanto en CRPC como en tejido sano, sugiere que podría representar una subpoblación tumoral diferenciada propia de estadios iniciales o menos agresivos del cáncer prostático. |
| 18 | LINC01088, MIPEP, FGF12, TRPM8, DSCAM-AS1, PCA3, SYCE1L, ABCC4, TFF3, TRGC1.        | Human Protein Atlas ( <i>The Human Protein Atlas</i> , n.d.), literatura. | PCA3-high luminal.   | Este clúster presenta alta expresión de genes vinculados al epitelio prostático y al fenotipo luminal. No se clasifican directamente por CellMarker 2.0, pero The Human Protein Atlas y estudios previos respaldan su identidad epitelial glandular diferenciada. PCA3 (Prostate Cancer ANtigen 3) es un long non-coding RNA altamente específico del tejido prostático, sobreexpresado en cáncer de próstata y ampliamente utilizado como marcador diagnóstico ( <i>DD3: A New Prostate-Specific Gene, Highly Overexpressed in Prostate Cancer - PubMed</i> , n.d.).  |

|  |  |  |  |   |
|--|--|--|--|---|
|  |  |  |  | Asimismo, TFF3 y TRPM8 también han sido relacionados con funciones epiteliales y eventos tempranos en la carcinogénesis prostática (Vestergaard et al., 2010). Su presencia casi exclusiva en PCa, con mínima representación en tejido sano y ausencia en CRPC, sugiere una subpoblación tumoral epitelial asociada a estadios tempranos y diferenciados del cáncer prostático. |
|--|--|--|--|---|

## 9.6. Bibliografía.

- Beltran, H., Hruszkewycz, A., Scher, H. I., Hildesheim, J., Isaacs, J., Yu, E. Y., Kelly, K., Lin, D., Dicker, A., Arnold, J., Hecht, T., Wicha, M., Sears, R., Rowley, D., White, R., Gulley, J. L., Lee, J., Meco, M. D., Small, E. J., ... Nelson, P. S. (2019). The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, 25(23), 6916–6924.  
<https://doi.org/10.1158/1078-0432.CCR-19-1423>
- Boyukozer, F. B., Tanoglu, E. G., Ozen, M., Ittmann, M., & Aslan, E. S. (2020). Kallikrein gene family as biomarkers for recurrent prostate cancer. *Croatian Medical Journal*, 61(5), 450.  
<https://doi.org/10.3325/CMJ.2020.61.450>
- Chifman, J., Pullikuth, A., Chou, J. W., Bedognetti, D., & Miller, L. D. (2016). Conservation of immune gene signatures in solid tumors and prognostic implications. *BMC Cancer*, 16(1).  
<https://doi.org/10.1186/S12885-016-2948-Z>
- Corn, P. G., Wang, F., McKeehan, W. L., & Navone, N. (2013). Targeting fibroblast growth factor pathways in prostate cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, 19(21), 5856–5866.  
<https://doi.org/10.1158/1078-0432.CCR-13-1550>
- DD3: a new prostate-specific gene, highly overexpressed in prostate cancer - PubMed. (n.d.). Retrieved January 5, 2026, from <https://pubmed.ncbi.nlm.nih.gov/10606244/>
- Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C. A. K., Skogs, M., Ottosson Takanen, J., Berling, H., Tegel, H., Mulder, J., ... Uhlen, M. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics : MCP*, 13(2), 397–406. <https://doi.org/10.1074/MCP.M113.035600>
- Fleischmann, A., Schlomm, T., Köllermann, J., Sekulic, N., Huland, H., Mirlacher, M., Sauter, G., Simon, R., & Erbersdobler, A. (2009). Immunological microenvironment in prostate cancer: high mast cell densities are associated with favorable tumor characteristics and good prognosis. *The Prostate*, 69(9), 976–981.  
<https://doi.org/10.1002/PROS.20948>
- Henry, G. H., Malewska, A., Joseph, D. B., Malladi, V. S., Lee, J., Torrealba, J., Mauck, R. J., Gahan, J. C., Raj, G. V., Roehrborn, C. G., Hon, G. C., MacConmara, M. P., Reese, J. C., Hutchinson, R. C., Vezina, C. M., & Strand, D. W. (2018). A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Reports*, 25(12), 3530-3542.e5.  
<https://doi.org/10.1016/J.CELREP.2018.11.086>
- Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., Bai, J., Chen, J., Jiang, W., Yang, K., Ou, Q., Li, X., Wang, P., & Zhang, Y. (2023). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Research*, 51(D1), D870–D876. <https://doi.org/10.1093/NAR/GKAC947>

- Ji, S., Wu, W., & Jiang, Q. (2023). Crosstalk between Endothelial Cells and Tumor Cells: A New Era in Prostate Cancer Progression. *International Journal of Molecular Sciences*, 24(23). <https://doi.org/10.3390/IJMS242316893>
- Johansson, A., Rudolfsson, S., Hammarsten, P., Halin, S., Pietras, K., Jones, J., Stattin, P., Egevad, L., Granfors, T., Wikström, P., & Bergh, A. (2010). Mast cells are novel independent prognostic markers in prostate cancer and represent a target for therapy. *The American Journal of Pathology*, 177(2), 1031–1041. <https://doi.org/10.2353/AJPATH.2010.100070>
- Johnson, G. L., Masias, E. J., & Correspondence, J. A. L. (n.d.). Cellular Heterogeneity and Lineage Restriction during Mouse Digit Tip Regeneration at Single-Cell Resolution. <https://doi.org/10.1016/j.devcel.2020.01.026>
- Liu, Y., & Yao, J. (2019). Research progress of cystatin SN in cancer. *OncoTargets and Therapy*, 12, 3411. <https://doi.org/10.2147/OTT.S194332>
- Mantovani, A., Marchesi, F., Malesci, A., Laghi, L., & Allavena, P. (2017). Tumour-associated macrophages as treatment targets in oncology. *Nature Reviews. Clinical Oncology*, 14(7), 399–416. <https://doi.org/10.1038/NRCLINONC.2016.217>
- Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y. Y., Carré, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., Zippelius, A., Pedro de Magalhães, J., & Larbi, A. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports*, 26(6), 1627–1640.e7. <https://doi.org/10.1016/J.CELREP.2019.01.041>
- Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., & Freeman, T. C. (2018). Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumors. *Cancer Immunology Research*. <https://doi.org/10.1158/2326-6066.CIR-18-0342>
- Öhlund, D., Elyada, E., & Tuveson, D. (2014). Fibroblast heterogeneity in the cancer wound. *The Journal of Experimental Medicine*, 211(8), 1503–1523. <https://doi.org/10.1084/JEM.20140692>
- Pan, J., Ma, Z., Liu, B., Qian, H., Shao, X., Liu, J., Wang, Q., & Xue, W. (2023). Identification of cancer-associated fibroblasts subtypes in prostate cancer. *Frontiers in Immunology*, 14. <https://doi.org/10.3389/FIMMU.2023.1133160>
- Patel, S. A., Hassan, M. K., & Dixit, M. (2024). Oncogenic activation of EEF1A2 expression: a journey from a putative to an established oncogene. *Cellular & Molecular Biology Letters*, 29(1), 6. <https://doi.org/10.1186/S11658-023-00519-9>
- Pathak, B. R., Breed, A. A., Apte, S., Acharya, K., & Mahale, S. D. (2016). Cysteine-rich secretory protein 3 plays a role in prostate cancer cell invasion and affects expression of PSA and ANXA1. *Molecular and Cellular Biochemistry*, 411(1–2), 11–21. <https://doi.org/10.1007/S11010-015-2564-2>
- Sharonov, G. V., Serebrovskaya, E. O., Yuzhakova, D. V., Britanova, O. V., & Chudakov, D. M. (2020). B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nature Reviews. Immunology*, 20(5), 294–307. <https://doi.org/10.1038/S41577-019-0257-X>

- Szabo, P. A., Levitin, H. M., Miron, M., Snyder, M. E., Senda, T., Yuan, J., Cheng, Y. L., Bush, E. C., Dogra, P., Thapa, P., Farber, D. L., & Sims, P. A. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications*, 10(1). <https://doi.org/10.1038/S41467-019-12464-3>
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, 45(W1), W98–W102. <https://doi.org/10.1093/NAR/GKX247>
- The Human Protein Atlas. (n.d.). Retrieved January 1, 2026, from [https://www.proteinatlas.org/?utm\\_source=chatgpt.com](https://www.proteinatlas.org/?utm_source=chatgpt.com)
- Trojan, L., Thomas, D., Friedrich, D., Grobholz, R., Knoll, T., Alken, P., & Michel, M. S. (2004). Expression of Different Vascular Endothelial Markers in Prostate Cancer and BPH Tissue: An Immunohistochemical and Clinical Evaluation. *Anticancer Research*, 24(3A), 1651–1656. [https://doi.org/10.1016/s1569-9056\(04\)90362-4](https://doi.org/10.1016/s1569-9056(04)90362-4)
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., ... Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220). [https://doi.org/10.1126/SCIENCE.1260419/SUPPL\\_FILE/1260419\\_UHLEN.SM.PDF](https://doi.org/10.1126/SCIENCE.1260419/SUPPL_FILE/1260419_UHLEN.SM.PDF)
- Vestergaard, E. M., Nexø, E., Tørring, N., Borre, M., Ørntoft, T. F., & Sørensen, K. D. (2010). Promoter hypomethylation and upregulation of trefoil factors in prostate cancer. *International Journal of Cancer*, 127(8), 1857–1865. <https://doi.org/10.1002/IJC.25209>
- Worst, T. S., Waldbillig, F., Abdelhadi, A., Weis, C. A., Gottschalt, M., Steidler, A., Von Hardenberg, J., Michel, M. S., & Erben, P. (2017). The EEF1A2 gene expression as risk predictor in localized prostate cancer. *BMC Urology*, 17(1). <https://doi.org/10.1186/S12894-017-0278-3>
- Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J. R., Bartonicek, N., Wang, T., Larsson, L., Kaczorowski, D., Weisenfeld, N. I., Uytingco, C. R., Chew, J. G., Bent, Z. W., Chan, C. L., Gnanasambandapillai, V., ... Swarbrick, A. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics* 2021 53:9, 53(9), 1334–1347. <https://doi.org/10.1038/s41588-021-00911-1>
- Xu, D., Wang, W., Wang, D., Ding, J., Zhou, Y., & Zhang, W. (2024). Long noncoding RNA MALAT-1: A versatile regulator in cancer progression, metastasis, immunity, and therapeutic resistance. *Non-Coding RNA Research*, 9(2), 388–406. <https://doi.org/10.1016/J.NCRNA.2024.01.015>
- Yamada, Y., & Beltran, H. (2021). Clinical and Biological Features of Neuroendocrine Prostate Cancer. *Current Oncology Reports*, 23(2). <https://doi.org/10.1007/S11912-020-01003-9>
- Yang, J., Luo, G., Li, C., Zhao, Z., Ju, S., Li, Q., Chen, Z., Ding, C., Tong, X., & Zhao, J. (2022). Cystatin SN promotes epithelial-mesenchymal transition and serves as a prognostic

biomarker in lung adenocarcinoma. *BMC Cancer*, 22(1). <https://doi.org/10.1186/S12885-022-09685-Z>

Zeng, Y., & Jiang, Z. (2025). Analysis of fatty acid metabolism in prostate cancer and discovery of a new programmed cell death-associated luminal cell subpopulation. *Discover Oncology*, 16(1). <https://doi.org/10.1007/S12672-025-01982-W>

Zhang, D., Fang, C., Li, H., Lu, C., Huang, J., Pan, J., Yang, Z., Liang, E., Liu, Z., Zhou, X., Xin, Z., Chen, Y., & Cai, Q. (2021). Long ncRNA MALAT1 promotes cell proliferation, migration, and invasion in prostate cancer via sponging miR-145. *Translational Andrology and Urology*, 10(6), 2307. <https://doi.org/10.21037/TAU-20-1526>

Zhang, J., Lu, S., Lu, T., Han, D., Zhang, K., Gan, L., Wu, X., Li, Y., Zhao, X., Li, Z., Shen, Y., Hu, S., Yang, F., Wen, W., & Qin, W. (2023). Single-cell analysis reveals the COL11A1+ fibroblasts are cancer-specific fibroblasts that promote tumor progression. *Frontiers in Pharmacology*, 14, 1121586. <https://doi.org/10.3389/FPHAR.2023.1121586/FULL>