



---

**Universidad de Valladolid**

**FACULTAD DE CIENCIAS**

**TRABAJO FIN DE GRADO**

**Grado en Estadística**

**Evaluación de algoritmos de *clustering* en  
Citometría de Flujo de alta dimensionalidad**

**Autora: María García Pizarro  
Tutoras: Itziar Fernández Martínez  
Sara Cuesta Sancho  
2024/2025**



## Resumen

La citometría de flujo es una técnica fundamental para el análisis celular que, gracias al avance tecnológico, permite medir de forma simultánea un mayor número de marcadores. Este aumento en la dimensionalidad ha generado grandes desafíos en el análisis manual de estos datos, promoviendo el desarrollo de algoritmos que automatizan este proceso.

El principal objetivo de este trabajo es comparar metodológica y computacionalmente dos algoritmos de *clustering*, FlowSOM y flowMeans, aplicados a datos de citometría de flujo. Para ello, se emplean tanto conjuntos de datos públicos como datos reales de alta dimensionalidad proporcionados por el Instituto de Biomedicina y Genética Molecular (IBGM). Se evalúan métricas de rendimiento como el F-score, así como la estabilidad de los resultados, el coste computacional y el impacto de aplicar una clasificación jerárquica.

Los resultados muestran que, si bien ambos algoritmos presentan resultados comparables, FlowSOM tiene una clara ventaja computacional y una mayor eficacia en contextos de alta dimensionalidad. Asimismo, la aplicación de FlowSOM de forma jerárquica mejora significativamente los resultados.

**Palabras clave:** FlowSOM, flowMeans, *clustering*, citometría de flujo, alta dimensionalidad, clasificación jerárquica.

## Abstract

Flow cytometry is a key technique for cellular analysis that, thanks to technological advances, allows simultaneous measurement of a large number of markers. This increased dimensionality has posed significant challenges for manual analysis, leading to the development of automated clustering algorithms.

The main objective of this study is to perform a comparative evaluation of the performance of two algorithms, FlowSOM and flowMeans, applied to flow cytometry data. Both public datasets and real high-dimensional data provided by the Institute of Biomedicine and Molecular Genetics (IBGM) are used.

Performance metrics such as the F-score are assessed, along with result stability, computational efficiency, and the impact of applying a hierarchical classification approach. The findings show that, although both algorithms yield comparable results, FlowSOM offers a clear computational advantage and greater effectiveness with high-dimensional data. Moreover, applying FlowSOM hierarchically leads to a notable improvement in performance.

**Key words:** FlowSOM, flowMeans, clustering, flow cytometry, high dimensionality, hierarchical classification.

## Agradecimientos

En primer lugar, me gustaría agradecer a mis tutoras, Itziar Fernández y Sara Cuesta, por su paciencia e implicación en este trabajo, que me ha permitido empezar a conocer de su mano el mundo de la bioestadística. Su orientación, correcciones y sugerencias han sido clave para poder alcanzar los objetivos de este proyecto.

También quiero dar las gracias al personal del laboratorio de *Inmunología de las Mucosas y Alergia* por acogerme y facilitarme los datos con los que he trabajado. En especial, a Carol y a Elisa por su paciencia y apoyo, por el tiempo dedicado a enseñarme nociones de citometría, y por todos los recursos que me han proporcionado para poder llevar a cabo este trabajo.

Por último, a mi familia y amigos, gracias por haberme acompañado durante este proceso.

# Índice

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>1</b>  |
| 1.1. Objetivos . . . . .  | 4         |
| 1.2. Asignaturas relacionadas . . . . .                             | 5         |
| <b>2. Metodología</b>   | <b>6</b>  |
| 2.1. Conjuntos de datos . . . . .                                   | 6         |
| 2.1.1. Consideraciones sobre los datos públicos . . . . .           | 7         |
| 2.1.2. Consideraciones sobre el conjunto del IBGM . . . . .         | 7         |
| 2.2. Preprocesado de los datos . . . . .                            | 8         |
| 2.3. Algoritmos de <i>clustering</i> . . . . .                      | 10        |
| 2.3.1. FlowSOM . . . . .  | 10        |
| 2.3.2. FlowMeans . . . . .  | 12        |
| 2.3.3. Selección del número de <i>clusters</i> . . . . .            | 13        |
| 2.4. Evaluación del rendimiento . . . . .                           | 14        |
| 2.5. Estabilidad de los resultados . . . . .                        | 17        |
| 2.6. Clasificación jerárquica . . . . .                             | 19        |
| 2.7. Implementación computacional . . . . .                         | 20        |
| <b>3. Resultados</b>  | <b>21</b> |
| 3.1. Comparativa de rendimiento: flowMeans frente FlowSOM . . . . . | 21        |
| 3.1.1. Comparativa de F-score . . . . .                             | 21        |

|           |  |           |
|-----------|--|-----------|
| 3.1.2.    | Comparativa del <i>accuracy</i> . . . . .  | 23        |
| 3.1.3.    | Comparativa de tiempos de ejecución . . . . .  | 24        |
| 3.1.4.    | Análisis por subpoblaciones celulares . . . . .  | 24        |
| 3.2.      | Evaluación de la estabilidad frente a la semilla aleatoria . . . . .                         | 26        |
| 3.3.      | Clasificación jerárquica con FlowSOM . . . . .   | 28        |
| 3.3.1.    | Resultados utilizando marcadores relevantes por nivel . . . . .                              | 28        |
| 3.3.2.    | Resultados utilizando todos los marcadores disponibles . . . . .                             | 31        |
| 3.3.3.    | Visualización de resultados: gráficos <i>Star Pie</i> . . . . .                              | 32        |
| <b>4.</b> | <b>Discusión</b>   | <b>38</b> |
| 4.1.      | Comparativa entre FlowSOM y flowMeans . . . . .  | 38        |
| 4.2.      | Impacto del método de asignación . . . . .   | 39        |
| 4.3.      | Variabilidad por la semilla aleatoria en FlowSOM . . . . .                                   | 39        |
| 4.4.      | Clasificación jerárquica con FlowSOM . . . . .   | 40        |
| <b>5.</b> | <b>Conclusiones y líneas futuras</b>   | <b>42</b> |
| 5.1.      | Líneas de trabajo futuras . . . . .  | 42        |
|           | <b>Referencias</b>   | <b>44</b> |
| <b>A.</b> | <b>Clasificación jerárquica de las poblaciones celulares en los datos IBGM</b>               | <b>49</b> |
| <b>B.</b> | <b>Resultados adicionales para la comparativa de rendimiento, flowMeans frente a FlowSOM</b> | <b>52</b> |

|   |           |
|---|-----------|
| <b>C. Resultados adicionales de la evaluación de la estabilidad</b> | <b>54</b> |
| <b>D. Resultados adicionales del análisis jerárquico</b>            | <b>56</b> |
| <b>E. Gráficos <i>Star Pie</i> del análisis jerárquico completo</b> | <b>58</b> |

## Índice de figuras

|    |   |    |
|----|---|----|
| 1. | Ilustración del proceso <i>manual gating</i> . La región seleccionada en la figura superior izquierda sirve de filtro para los datos representados en la figura superior derecha. Este procedimiento se repite de forma iterativa hasta aislar la población de interés, en este caso, los basófilos. El proceso completo se muestra en la Figura A.1 (Apéndices). . . . .   | 2  |
| 2. | Aumento del número de referencias a ' <i>flow cytometry</i> ' en PubMed. . . . .  | 3  |
| 3. | Comparativa con <i>manual gating</i> : red del mapa autoorganizado (SOM) y árbol de expansión mínima (MST). Ilustración tomada de la Figura 3 (A) de Van Gassen et al. [12]. El panel izquierdo muestra la red de nodos del SOM. En el panel derecho, los nodos están conectados mediante un MST, y se visualizan con diagramas circulares que indican la proporción de células de cada tipo asignadas a cada nodo según la clasificación de referencia obtenida por <i>manual gating</i> . . . . . | 12 |
| 4. | Estabilidad de los <i>clusters</i> al inicializarse con 30 semillas diferentes (Fuente: Figura S17, Weber y Robinson [14]). . . . .   | 18 |
| 5. | Esquema de la jerarquía de las poblaciones celulares del conjunto IBGM, elaborado a partir de ilustraciones de BioArt [45]. . . . .   | 19 |
| 6. | Evolución del F-score por poblaciones, ordenadas de mayor a menor abundancia, para cada método de asignación y algoritmo utilizado. . . . .   | 25 |
| 7. | Distribución de las métricas de rendimiento ( <i>precision</i> , <i>recall</i> , F-score y <i>accuracy</i> ) obtenidas tras ejecutar FlowSOM con diez semillas distintas sobre el conjunto IBGM. Se comparan los resultados según el tipo de asignación empleado (máximo o húngaro). . . . .  | 27 |
| 8. | <i>Star Pie</i> del nivel <i>General</i> incluyendo solo los marcadores esenciales para la caracterización celular. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . . . . .  | 34 |
| 9. | <i>Star Pie</i> del nivel <i>General</i> incluyendo todos los marcadores. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . . . . .  | 35 |

|      |  |    |
|------|--|----|
| 10.  | <i>Star Pie</i> del nivel <i>T cells</i> incluyendo todos los marcadores esenciales para la caracterización celular. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . . . . .  | 36 |
| 11.  | <i>Star Pie</i> del nivel <i>T cells</i> incluyendo todos los marcadores. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . . . . .   | 37 |
| A.1. | <i>Manual gating</i> completa del conjunto IBGM. . . . .   | 51 |
| E.1. | <i>Star Pie plot</i> del nivel <i>NKT</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . .             | 58 |
| E.2. | <i>Star Pie plot</i> del nivel <i>Tgd</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . .             | 59 |
| E.3. | <i>Star Pie plot</i> del nivel <i>Real T cells</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . .    | 60 |
| E.4. | <i>Star Pie plot</i> del nivel <i>CD4<sup>+</sup></i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . . | 61 |
| E.5. | <i>Star Pie plot</i> del nivel <i>CD8<sup>+</sup></i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . . | 62 |
| E.6. | <i>Star Pie plot</i> del nivel <i>B cells</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . .         | 63 |

|  |    |
|--|----|
| E.7. <i>Star Pie plot</i> del nivel <i>IgD<sup>-</sup></i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . .      | 63 |
| E.8. <i>Star Pie plot</i> del nivel <i>Plasmablasts</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . .           | 64 |
| E.9. <i>Star Pie plot</i> del nivel <i>Memory B cells</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . .         | 64 |
| E.10. <i>Star Pie plot</i> del nivel <i>No T &amp; No B cells</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . | 65 |
| E.11. <i>Star Pie plot</i> del nivel <i>NK cells</i> , para el modelo que incluye los marcadores esenciales. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . . . . .  | 66 |
| E.12. <i>Star Pie plot</i> del nivel <i>APC</i> . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de <i>meta-clustering</i> . (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles. . . . .                 | 67 |

## Índice de cuadros

|   |    |
|---|----|
| 1. F-score para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . .         | 22 |
| 2. <i>Accuracy</i> para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . . | 23 |

|      |   |    |
|------|---|----|
| 3.   | Tiempo de ejecución para los distintos algoritmos en formato hh:mm:ss. . . . .  | 24 |
| 4.   | Valores promedio de <i>precision</i> , <i>recall</i> , F-score y <i>accuracy</i> obtenidos por FlowSOM en 10 ejecuciones independientes con semillas distintas sobre el conjunto IBGM. Se muestran los valores medios junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> [44]. Se comparan los resultados bajo los métodos de asignación por máximo y por algoritmo del húngaro. . . . . | 28 |
| 5.   | F-score y <i>accuracy</i> en el análisis jerárquico con FlowSOM empleando solo los marcadores relevantes por nivel. Se muestran los valores estimados junto con los intervalos <i>bootstrap</i> de confianza al 95 %. . . . .   | 30 |
| 6.   | F-score y <i>accuracy</i> en el análisis jerárquico con FlowSOM empleando todos los marcadores disponibles. Se muestran los valores estimados junto con los intervalos <i>bootstrap</i> de confianza al 95 %. . . . .   | 32 |
| B.1. | <i>Precision</i> para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . .   | 52 |
| B.2. | <i>Recall</i> para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . .  | 53 |
| C.1. | F-score y <i>accuracy</i> para la ejecución de FlowSOM óptimo, cada fila corresponde a cada una de las diez semillas. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . .   | 54 |
| C.2. | <i>Precision</i> y <i>recall</i> para la ejecución de FlowSOM óptimo, cada fila corresponde a cada una de las diez semillas. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . .  | 55 |
| D.1. | <i>Precision</i> y <i>recall</i> en el análisis jerárquico con FlowSOM empleando solo los marcadores relevantes por nivel. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método <i>cheap bootstrap</i> propuesto por Lam y Liu [44]. . . . .  | 56 |

D.2. *Precision y recall* en el análisis jerárquico con FlowSOM empleando todos los marcadores disponibles. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44]. . . . . 57

## Listado de acrónimos y abreviaturas

|                     |  |
|---------------------|--|
| APC                 | Antigen-presenting cells   |
| arcinh              | Arcoseno hiperbólico   |
| BIC                 | Bayesian information criterion   |
| CD4 <sup>+</sup>    | Real T cells CD4 <sup>+</sup>  |
| CD8 <sup>+</sup>    | Real T cells CD8 <sup>+</sup>  |
| cDC                 | Conventional dendritic cells   |
| DP                  | Double Positive (CD4 <sup>+</sup> CD8 <sup>+</sup> ) cells               |
| FCS                 | Flow Cytometry Standard  |
| FlowCAP             | Flow Cytometry: Critical Assesment of Population Indentification Methods |
| IBGM                | Instituto de Biomedicina y Genética Molecular                            |
| IgD <sup>+</sup>    | IgD <sup>+</sup> B cells   |
| IgD <sup>-</sup>    | IgD <sup>-</sup> B cells   |
| ILC                 | Innate lymphoid cells  |
| ISAC                | International Society for Advancement of Cytometry                       |
| MST                 | Minimal Spanning Tree  |
| NKT                 | Natural Killer T cells   |
| NK cells            | Natural Killer cells   |
| nrT <sub>EMRA</sub> | Nonresident T effector memory reexpressing RA                            |
| PBMC                | Peripheral blood mononuclear cells                                       |
| rT <sub>EMRA</sub>  | Resident T effector memory reexpressing RA                               |
| SOM                 | Self-Organizing Map  |
| Tgd                 | Gamma-delta T cells  |
| Trm                 | Tissue-resident memory cells   |

# 1. Introducción

La citometría de flujo es una tecnología que permite medir características físicas y químicas de células que fluyen a través del citómetro de flujo mientras se encuentran en una solución salina [1, 2]. Haciendo uso de láseres como fuentes de excitación, el citómetro de flujo produce señales de luz que son convertidas en señales electrónicas [2]. Estas señales son procesadas y analizadas en archivos de formato FCS (Flow Cytometry Standard) [3]. Este formato, desarrollado por ISAC (International Society for Advancement of Cytometry) [4], permite almacenar toda la información generada por el citómetro.

En este contexto, el término evento se define como la lista de mediciones del citómetro para una célula concreta, siendo su longitud igual al número de parámetros registrados. Los parámetros corresponden a una característica numérica de la señal producida por uno de los detectores del citómetro [1]. Los marcadores son aquellos parámetros asociados a moléculas específicas que permiten identificar diferentes tipos celulares. A partir de ahora se usará esta terminología (*eventos*, *parámetros* y *marcadores*) para describir los datos contenidos en los archivos FCS.

Uno de los conceptos clave en el análisis de citometría de flujo es el de población celular, entendida como un conjunto de células con características comunes (por ejemplo, linfocitos o monocitos). A su vez, una población de interés hace referencia a aquellas poblaciones especialmente relevantes para un determinado estudio, que pueden incluir poblaciones raras, es decir, aquellas de baja frecuencia pero gran valor biológico.

A lo largo de este documento, el término 'muestra' no se usará en el sentido estadístico habitual, sino para hacer referencia a las distintas muestras biológicas, como puede ser una muestra de sangre, tejido u orina.

Desde su aparición en 1965 [5], la citometría de flujo ha evolucionado considerablemente. El número de parámetros que se pueden medir ha ido aumentando progresivamente, desde los 10 que se podían medir en 1997 [6], pasando por los 12 de 2010 [7], hasta alcanzar los 50 en 2024 [8].

La forma convencional de analizar estos datos, comúnmente denominada *manual gating*, consiste en representar los parámetros dos a dos en gráficos de densidad bidimensionales y seleccionar regiones de interés de forma manual para aislar las poblaciones celulares. Este proceso se repite hasta llegar a las poblaciones de interés. Sin embargo, con el aumento de la dimensionalidad, este enfoque se ha vuelto poco práctico y propenso a omitir poblaciones relevantes [2, 9], una situación especialmente habitual en muestras de pacientes no sanos, donde es común encontrar poblaciones no convencionales de interés.

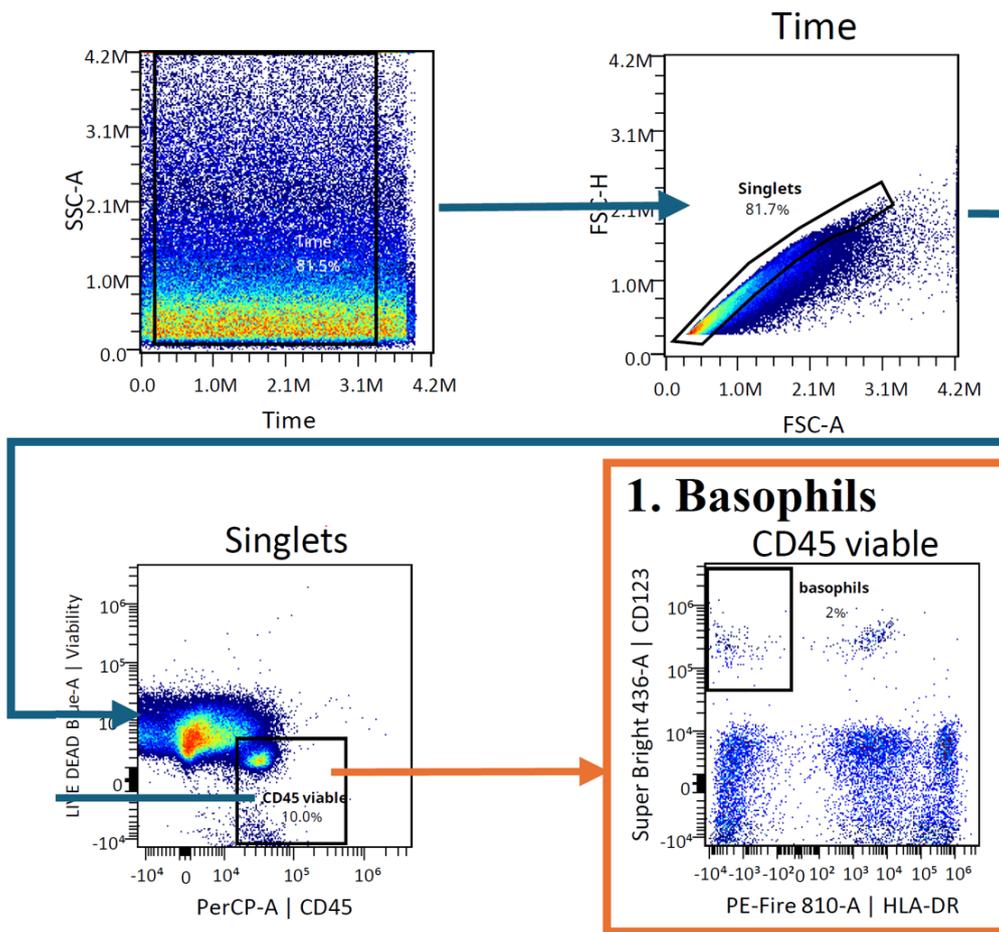


Figura 1: Ilustración del proceso *manual gating*. La región seleccionada en la figura superior izquierda sirve de filtro para los datos representados en la figura superior derecha. Este procedimiento se repite de forma iterativa hasta aislar la población de interés, en este caso, los basófilos. El proceso completo se muestra en la Figura A.1 (Apéndices).

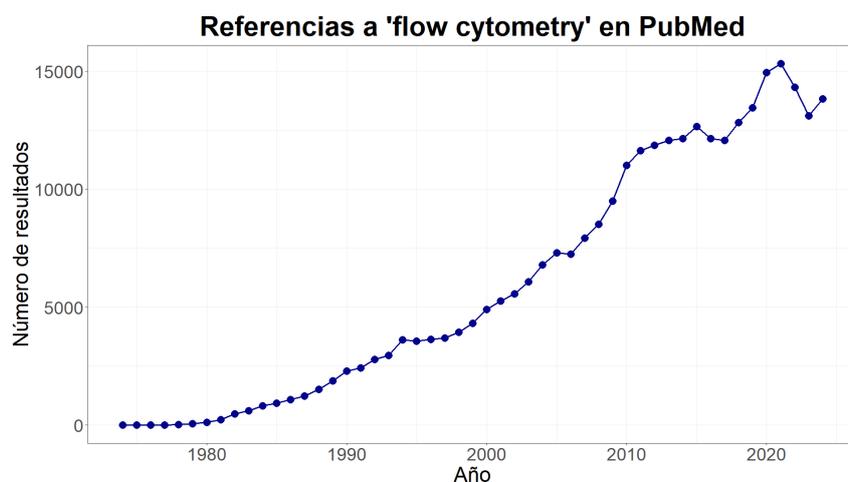


Figura 2: Aumento del número de referencias a 'flow cytometry' en PubMed.

La citometría de flujo presenta múltiples aplicaciones en campos como la inmunología, la biología celular y/o molecular [2]. Como consecuencia, el uso de esta tecnología ha aumentado notablemente en las últimas décadas, como muestra el creciente número de referencias al término 'flow cytometry' en la base de datos biomédica PubMed (Figura 2).

Teniendo en cuenta los aspectos señalados anteriormente, es lógico que hayan surgido nuevas herramientas que automaticen el proceso, como por ejemplo SPADE (*Spanning-tree progression analysis of density-normalized events*) [10], FLOCK (*FLOW clustering without K*) [11] o FlowSOM [12]. Sin embargo, la elección de la herramienta más adecuada no es trivial. En 2013 el FlowCAP-I challenge [13], propuso una primera evaluación sistemática del rendimiento de diferentes algoritmos. Posteriormente, en 2016, Weber y Robinson [14] realizaron un estudio comparativo más exhaustivo, proponiendo mejoras metodológicas. En 2019 se publicaron nuevas recomendaciones metodológicas [15], aunque centradas en citometría de masas, una tecnología relacionada pero basada en isótopos metálicos en lugar de fluorocromos. Esta diferencia técnica hace que sus resultados no sean directamente extrapolables a citometría de flujo.

Desde el punto de vista estadístico, el análisis automático de datos de citometría de flujo mediante técnicas de *clustering* plantea múltiples desafíos. Al tratarse de un problema de *clustering* no supervisado, la falta de etiquetas conocidas para entrenar modelos obliga a desarrollar estrategias sólidas de evaluación externa basadas en anotaciones manuales o conocimiento biológico.

Por otro lado, la elevada dimensionalidad de los datos puede comprometer tanto el rendimiento como la interpretabilidad de los algoritmos clásicos, haciendo necesario el uso de

técnicas de preprocesamiento adecuadas y métodos de *clustering* capaces de adaptarse a la compleja estructura de las poblaciones celulares, incluyendo aquellas con formas irregulares o poco frecuentes.

La aleatoriedad de ciertos algoritmos introduce una fuente adicional de variabilidad que puede condicionar los resultados. Por ello, evaluar su estabilidad frente a distintas inicializaciones y semillas es esencial para garantizar la reproducibilidad y obtener conclusiones más sólidas desde un enfoque estadístico riguroso.

Otro punto clave es la evaluación objetiva del rendimiento de los algoritmos. Para ello, es necesario establecer métricas comparables—como el F-score, la precisión, el *recall* o la *accuracy*—, así como estrategias de asignación consistentes, tales como el algoritmo del húngaro en comparación con criterios de máxima coincidencia.

En este trabajo se propone una evaluación comparativa de dos de los algoritmos más usados en citometría de flujo, FlowSOM y flowMeans, centrada no solo en su rendimiento en distintas métricas, sino también en su eficiencia computacional y su estabilidad ante datos reales de alta dimensionalidad.

## 1.1. Objetivos

El objetivo general de este trabajo es comparar distintos métodos de *clustering* aplicados a datos de citometría de flujo, con el fin de evaluar su rendimiento en la identificación de poblaciones celulares de interés, su eficiencia computacional y su estabilidad en entornos de alta dimensionalidad.

A partir de este objetivo general, se proponen los siguientes objetivos específicos:

- Comparar los resultados obtenidos mediante FlowSOM [12] y flowMeans [16] frente al análisis de referencia basado en *manual gating*.
- Analizar el coste computacional de ambos algoritmos al aplicarlos sobre conjuntos de datos de distinto tamaño y dimensionalidad.
- Evaluar la estabilidad de los resultados de FlowSOM frente a distintas semillas aleatorias.
- Estudiar el comportamiento de FlowSOM cuando se aplica de forma jerárquica sobre subconjuntos definidos por niveles celulares.
- Aplicar los algoritmos a conjuntos de datos reales proporcionados por el IBGM, valorando su aplicabilidad práctica en contextos biomédicos reales.

## 1.2. Asignaturas relacionadas

A lo largo del grado he cursado diferentes asignaturas que me han resultado de gran utilidad a la hora de realizar este trabajo. Entre ellas caben destacar las siguientes:

- **Computación estadística.** En esta asignatura aprendimos a programar en el lenguaje R que ha sido el usado para este trabajo.
- **Técnicas de aprendizaje automático y Análisis multivariante.** A lo largo de estas asignaturas se trabajaron contenidos relacionados con *clustering* y sus distintas métricas de evaluación.
- **Asignaturas de Investigación Operativa.** En asignaturas como Modelos de Investigación Operativa o Programación Entera se han trabajado los problemas de asignación y sus distintas soluciones que ha sido esencial a la hora de asignar los *clusters* identificados con las poblaciones del *manual gating*.

## 2. Metodología

En este apartado se expone la metodología utilizada para llevar a cabo la comparación de algoritmos de *clustering* aplicados a datos de citometría de flujo. Primero se describen los algoritmos de *clustering* seleccionados, flowMeans y FlowSOM. Después, se presentan los distintos conjuntos de datos utilizados y las transformaciones realizadas para su análisis. A continuación, se explican los criterios de evaluación empleados para comparar los resultados, incluyendo tanto las métricas de rendimiento como aspectos relacionados con la estabilidad de los resultados y la elección del número de *clusters*. También se evalúa la aplicación de FlowSOM en un esquema de clasificación jerárquica. Por último, se recogen una serie de consideraciones técnicas sobre el entorno de trabajo y la implementación.

### 2.1. Conjuntos de datos

En este estudio se han utilizado tres conjuntos de datos de citometría de flujo, dos de acceso público, disponibles en el FlowRepository [17], y uno real proporcionado por el grupo de investigación 'Inmunología de las Mucosas y Alergias' del IBGM, caracterizado por tener mayor dimensionalidad. Todos ellos están compuestos por muestras biológicas etiquetadas mediante *manual gating*, lo que permite evaluar el rendimiento de los algoritmos de *clustering* mediante comparación con un estándar de referencia. Aunque este método se considere fiable, incluso en contextos de alta dimensionalidad, conviene recordar que se trata de un procedimiento manual y, por tanto, susceptible a cierta subjetividad [2].

Las características principales de los conjuntos de datos mencionados son las siguientes:

- **Normal Donor**

Conjunto de datos proveniente del FlowCAP-I challenge [13] que tiene como objetivo examinar las diferencias en la respuesta de distintos tipos de células a diversos estímulos en un grupo de pacientes sanos [13]. Contiene 1 778 883 eventos y 14 parámetros, de los cuáles 10 son marcadores. Incluye 7 poblaciones celulares objetivo, con una tasa de anotación del 73.28 %.

- **Symptomatic West Nile virus**

Conjunto de datos proveniente del FlowCAP-I challenge [13]. Este conjunto contiene muestras de células mononucleares de sangre periférica (PBMCs) de pacientes humanos que presentan una infección sintomática del Virus del Nilo Occidental. Las muestras

han sido expuestas a diversos estímulos inmunológicos. El conjunto contiene 1 214 373 eventos y 10 parámetros, de las cuales 6 son marcadores. Tiene 4 poblaciones objetivo y está anotado al 94.26 %.

#### ■ **Datos IBGM**

Conjunto real proporcionado por el IBGM, compuesto por datos de citometría de células de la pared intestinal (íleon o colon) de pacientes humanos con enfermedades inflamatorias intestinales (enfermedad de Crohn y colitis ulcerosa, en fases activa y quiescente), así como de sujetos sanos. Tras un proceso de limpieza manual, el conjunto final consta de 1 567 074 eventos y 49 parámetros, de las cuales 40 son marcadores. Está anotado al 100 % e incluye 51 poblaciones celulares distintas.

### **2.1.1. Consideraciones sobre los datos públicos**

Los conjuntos de datos procedentes del FlowCAP-I challenge presentan diversas repeticiones del experimento, entendiendo como diferentes condiciones de estimulación aplicadas a las mismas poblaciones celulares. Sin embargo, las poblaciones de interés no han sido etiquetadas de forma consistente entre repeticiones, lo que complica una evaluación conjunta. A pesar de los esfuerzos realizados para estandarizar el etiquetado, incluyendo la revisión por parte de inmunólogos, no se dispone de información suficiente para garantizar una asignación precisa y homogénea de las clases en todos los casos.

Por este motivo, se ha decidido aplicar los algoritmos de *clustering* de forma independiente para cada repetición experimental. Asimismo, algunas repeticiones del conjunto Normal Donor han debido ser descartadas, debido a errores de singularidad durante la ejecución de los algoritmos, impidiendo generar una partición válida.

### **2.1.2. Consideraciones sobre el conjunto del IBGM**

Las muestras del IBGM han sido procesadas mediante una tecnología avanzada de la citometría de flujo convencional, conocida como citometría de flujo espectral. Mientras que la citometría de flujo convencional detecta la emisión de los fluorocromos en ventanas espectrales discretas, la citometría de flujo espectral registra el espectro completo de emisión de los fluorocromos, lo que permite medir un número notablemente mayor de parámetros de forma simultánea [18]. Desde el punto de vista del análisis de datos, la principal diferencia respecto a la citometría de flujo convencional radica en el tratamiento de la compensación espectral durante el preprocesamiento. Sin embargo, una vez completado este paso, no existen diferencias

metodológicas entre trabajar con datos de citometría de flujo espectral o convencional en lo que respecta a los análisis posteriores realizados en este trabajo (*clustering*, evaluación de métricas y visualización).

## 2.2. Preprocesado de los datos

Antes de aplicar los algoritmos de *clustering*, los datos se someten a un proceso de preprocesado que incluye pasos de compensación, transformación, limpieza y escalado. Estos pasos son fundamentales en citometría de flujo, ya que los datos crudos pueden contener ruido técnico, solapamientos espectrales y escalas no comparables entre parámetros.

### 1. Compensación

En citometría de flujo, los fluorocromos utilizados para marcar diferentes tipos celulares pueden presentar solapamiento espectral, es decir, emitir fluorescencia en canales adyacentes. Este fenómeno genera señales no deseadas en varios detectores simultáneamente, provocando correlaciones artificiales entre parámetros que no reflejan diferencias biológicas reales. Para corregir este efecto, se aplica una matriz de compensación que cuantifica la contribución cruzada de cada fluorocromo en los distintos canales. Esta matriz, calculada a partir de controles de fluorescencia con marcajes únicos, permite ajustar los valores de forma que cada canal represente exclusivamente la señal del marcador correspondiente. En citometría de flujo espectral este proceso se realiza mediante un desmezclado espectral (*spectral unmixing*), que emplea el espectro completo de emisión registrado por un conjunto de detectores. A partir de los espectros de referencia de cada fluorocromo, se resuelve un sistema de ecuaciones que estima la contribución real de cada marcador a la señal total registrada [18]. En este estudio, se aplicó este proceso de corrección únicamente al conjunto de datos del IBGM, ya que los conjuntos públicos ya incluían la compensación en su versión preprocesada.

### 2. Transformación

La transformación de los datos es un paso fundamental para poder analizarlos, ya que permite reducir la variabilidad de la intensidad de fluorescencia entre las distintas poblaciones celulares [9, 19, 20, 21]. Las intensidades crudas suelen mostrar distribuciones muy asimétricas y heterocedásticas, por lo que su transformación mejora tanto la interpretación visual como la estabilidad de los algoritmos aplicados. Aunque la elección de la transformación es dependiente del conjunto de datos [19], en este trabajo se ha utilizado la transformación arcoseno hiperbólico (*arcsinh*), ampliamente utilizada en el campo [2] y cuya expresión es:

$$f(y) = \ln(y + \sqrt{y^2 + 1})$$

$$f\left(\frac{y}{c}\right) = \ln\left(\frac{y}{c} + \sqrt{\left(\frac{y}{c}\right)^2 + 1}\right); \quad 10 < c < 10,000 \quad (1)$$

La transformación arcsinh requiere definir un cofactor  $c$  para cada marcador. Esta elección suele realizarse de forma empírica, ajustando distintos valores hasta obtener una representación clara en forma de dos picos distinguibles en los histogramas de intensidad [2]. Debido a que este procedimiento puede ser laborioso, se adopta un valor estándar de  $c = 150$  [2, 22, 23].

En los conjuntos públicos Normal Donor y West Nile Virus, los datos ya venían transformados con arcsinh utilizando ese cofactor estándar, lo que permite replicar los análisis de estudios previos como el de Weber y Robinson [14]. En el caso del conjunto del IBGM, la transformación también se realizó con arcsinh pero aplicando cofactores específicos para cada marcador, determinados por investigadores del grupo 'Inmunología de las mucosas y Alergias', en función de la estructura de los datos.

### 3. Limpieza

La limpieza se aplicó exclusivamente al conjunto de datos del IBGM. En cambio, los conjuntos públicos del FlowCAP-I challenge se utilizaron directamente en su versión preprocesada, que ya incluye los filtros aplicados por los autores del estudio original.

En el caso del IBGM, la limpieza se realizó en dos fases. En primer lugar, se hizo una limpieza manual, que consistió en la eliminación de artefactos técnicos y eventos no celulares. A continuación, se llevó a cabo una limpieza automática, con el fin de eliminar anomalías en la adquisición de la señal y en los eventos atípicos. Este procedimiento se implementó con el algoritmo PeacoQC [24].

### 4. Escalado

El algoritmo FlowSOM requiere que los datos estén previamente escalados para garantizar que todos los marcadores contribuyan de forma equilibrada al cálculo de distancias. Para ello, los datos se escalaron por columnas, transformando cada marcador para que tenga media cero y varianza uno. Este paso evita que las variables con mayor dispersión dominen la estructura de los *clusters* y asegura una comparación justa entre parámetros.

## 2.3. Algoritmos de *clustering*

A efectos comparativos, se seleccionaron los algoritmos FlowSOM [12] y flowMeans [16], ambos ampliamente utilizados en el análisis de datos de citometría de flujo. FlowSOM destaca por su popularidad en el ámbito biomédico y su capacidad para manejar grandes volúmenes de datos. FlowMeans, por su parte, se basa en el algoritmo clásico de k-medias, conocido por su sencillez, facilidad de implementación y eficiencia [25]. Ambos algoritmos están disponibles como paquetes de código abierto en la plataforma Bioconductor [26], lo que garantiza la reproducibilidad de los resultados.

### 2.3.1. FlowSOM

FlowSOM es un algoritmo de *clustering* basado en mapas auto-organizados (*Self-Organizing Maps*, SOM), técnica de redes neuronales introducida por Kohonen [27]. Su aplicación a citometría de flujo hace posible la representación de grandes volúmenes de datos multivariantes en una estructura espacial representativa, sobre la que posteriormente se aplica *clustering* jerárquico por consenso para generar los denominados *meta-clusters* [12, 28]. El proceso consta de varias fases:

#### 1. Lectura y preprocesamiento de los datos

Los archivos de datos se combinan en una matriz  $n \times d$ , donde cada fila representa un evento y cada columna un marcador. Una vez hecho esto, se realiza una compensación y transformación de los datos según lo detallado en la sección 2.2. Por último, los datos se escalan por columna para darle el mismo peso a todas las columnas. Este último paso es prescindible si los investigadores quieren dar más peso a alguna de las columnas.

#### 2. Construcción del Mapa Auto-Organizado

Los SOM son un tipo de red neuronal artificial introducida por Kohonen [27]. La implementación empleada en FlowSOM se basa en la propuesta de Wehrens et al. [28], adaptada específicamente para datos de citometría de flujo. Un SOM está formado por una red de nodos (también llamados neuronas), donde cada uno de ellos tiene asociado un vector de pesos que representa una posición en el espacio de datos. El objetivo del entrenamiento es organizar estos nodos de forma que los más próximos entre sí en la estructura topológica del mapa también sea similares en el espacio de los datos. Esto permite generar una representación bidimensional interpretable de relaciones complejas de alta dimensionalidad.

A la hora de entrenar el modelo, se inicializan los nodos con puntos aleatorios del conjunto de datos. Usando la distancia de Chebyshev [29] como función de vecindad, se busca el nodo más próximo para cada evento y se actualiza la lista de vecinos de dicho nodo. Tanto el tamaño del vecindario,  $\epsilon$ , como el factor de aprendizaje,  $\alpha$ , decrecen progresivamente a lo largo de las iteraciones, permitiendo un ajuste más fino en las etapas finales del entrenamiento.

### 3. Construcción de un Árbol de Expansión Mínima

Una vez entrenado el SOM, FlowSOM utiliza un árbol de expansión mínima (*Minimal Spanning Tree*, MST) para representar gráficamente su estructura topológica. Un MST es un grafo no dirigido que conecta todos los nodos de una red minimizando la suma total de las distancias entre ellos, sin formar ciclos [30].

En este contexto, cada nodo del SOM se representa como un vértice del grafo y se conecta al nodo más similar según una medida de distancia. Esta visualización permite interpretar las relaciones entre *clusters* de manera intuitiva, facilitando la identificación de transiciones suaves entre poblaciones celulares y la estructura general de los datos. Es importante destacar que el MST no forma parte del proceso de *clustering*, sino que se emplea únicamente como una herramienta de representación.

### 4. *Meta-clustering*

Aunque SOM puede utilizarse directamente como un método de *clustering*, asignando cada evento al nodo más cercano, esta aproximación suele generar un número elevado de grupos, lo que dificulta la interpretación. Por este motivo, se recomienda utilizar un número de nodos mayor que el número esperado de poblaciones celulares, con el fin de capturar en gran detalle la estructura local de los datos. Posteriormente, estos nodos se agrupan en un número reducido de *clusters* mediante un segundo paso de *meta-clustering*.

El paso de *meta-clustering* se lleva a cabo mediante *clustering* jerárquico por consenso, una técnica que evalúa la estabilidad de las particiones obtenidas a través de múltiples repeticiones del algoritmo. Para ello se utiliza la implementación *ConsensusClusterPlus* del entorno R [31], que permite generar agrupaciones robustas y reproducibles. Los grupos resultantes, denominados *meta-clusters*, se corresponden con las poblaciones celulares estimadas por el método.

Un ejemplo visual de esta representación se muestra en la Figura 3, tomada del artículo original de FlowSOM [12]. En ella se ilustra cómo los nodos del SOM se conectan entre sí mediante un MST, facilitando la interpretación estructural de los *clusters*.

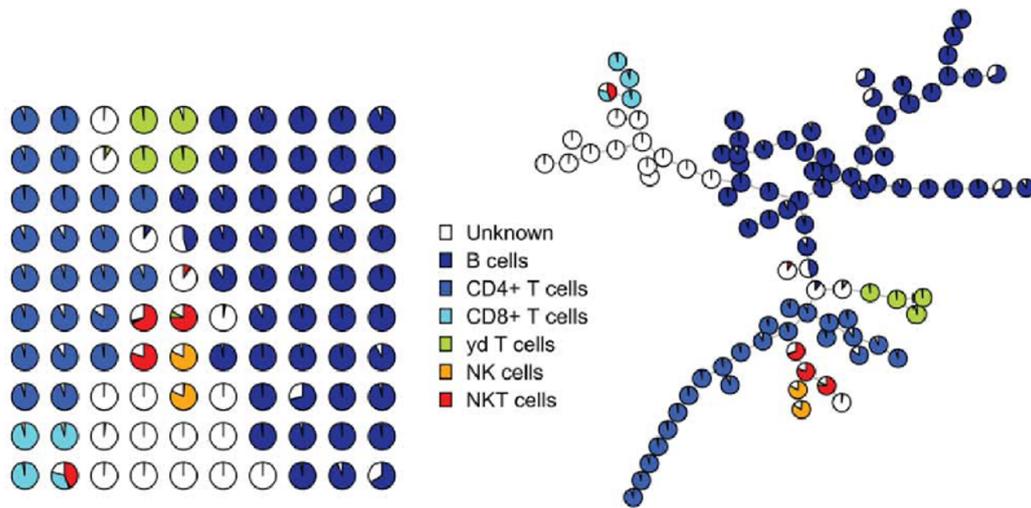


Figura 3: Comparativa con *manual gating*: red del mapa autoorganizado (SOM) y árbol de expansión mínima (MST). Ilustración tomada de la Figura 3 (A) de Van Gassen et al. [12]. El panel izquierdo muestra la red de nodos del SOM. En el panel derecho, los nodos están conectados mediante un MST, y se visualizan con diagramas circulares que indican la proporción de células de cada tipo asignadas a cada nodo según la clasificación de referencia obtenida por *manual gating*.

### 2.3.2. FlowMeans

FlowMeans es un algoritmo de *clustering* desarrollado específicamente para el análisis de datos de citometría de flujo [16]. Se basa en una extensión del algoritmo clásico de las k-medias, al que incorpora mejoras orientadas a capturar estructuras de datos no esféricas y a automatizar la selección del número de *clusters*.

El procedimiento tiene tres etapas principales:

#### 1. Estimación inicial del número de *clusters*

El algoritmo comienza proyectando los datos sobre sus componentes principales. En cada una de estas proyecciones, se estima la densidad de los datos mediante un núcleo gaussiano, aplicando el método propuesto por Duong et al. [32]. Las modas identificadas en cada dimensión se combinan para establecer una cota superior del número inicial de *clusters*.

Aunque la estimación puede dar lugar a *clusters* redundantes o solapados, se considera

una cota superior razonable, útil para evitar omitir poblaciones celulares relevantes. Esta sobreestimación se corrige en fases posteriores del algoritmo.

## 2. *Clustering* inicial y fusión de *clusters*

Una vez determinada la cota superior del número superior de *clusters*, se ejecuta una primera partición de los datos mediante una variante de k-medias. Esta etapa inicial tiende a generar un número elevado de grupos, que pueden incluir *clusters* solapados o redundantes. Con el fin de refinar la agrupación, se aplica un procedimiento de fusión iterativa entre pares de *clusters* basándose en una distancia generalizada de Mahalanobis. Concretamente, la distancia entre dos poblaciones X e Y se define como:

$$D(X, Y) = \min\{\sqrt{(\bar{X} - \bar{Y})S_X^{-1}(\bar{X} - \bar{Y})^T}; \sqrt{(\bar{X} - \bar{Y})S_Y^{-1}(\bar{X} - \bar{Y})^T}\} \quad (2)$$

donde  $\bar{X}$  y  $\bar{Y}$  son las medias de los *clusters* y  $S_X$  y  $S_Y$  sus matrices de varianzas-covarianzas. Los pares de *clusters* con una distancia baja, indicativa de un posible solapamiento, se fusionan. Este proceso se repite de forma iterativa hasta que se detecta un punto de cambio en la serie de distancias de fusión, utilizando para ello un modelo de regresión segmentada. El número óptimo de *clusters* se determina en el punto en el que el error del modelo alcanza su mínimo, lo que indica que las poblaciones restantes están claramente separadas.

## 3. Reducción de la sensibilidad a la inicialización

Uno de los principales inconvenientes del algoritmo k-medias es su dependencia de la inicialización aleatoria de los centroides, lo que puede llevar a soluciones locales. En flowMeans, este problema se aborda mediante una sobreestimación inicial del número de *clusters*, lo que permite captar estructuras más detalladas del espacio de datos desde el inicio.

Posteriormente, el proceso de fusión ajusta automáticamente el número de grupos, reduciendo así el impacto de las condiciones iniciales y mejorando la estabilidad del algoritmo frente a diferentes ejecuciones.

### 2.3.3. Selección del número de *clusters*

Una cuestión clave en la aplicación de algoritmos de *clustering* es la elección del número *clusters*, ya que influye directamente en la capacidad del modelo para clasificar con precisión las poblaciones presentes en el conjunto de datos. A pesar de que existen múltiples métodos para estimar este número, como los comparados por Kodinariya et al. [33], no existe un consenso

universal sobre cuál es el más adecuado [34]. Cada técnica presenta limitaciones distintas, cuya efectividad depende en gran medida de las características específicas de los datos.

Entre los métodos más usados se encuentra el método del codo (*elbow method*), el cual se puede atribuir a Thorndike [35], que busca un punto de inflexión en la curva de inercia *intra-cluster*. Sin embargo, algunos autores como Schubert [36] advierten sobre su uso excesivo y proponen el uso de otros métodos más robustos como el índice de Calinski–Harabasz [37] o el criterio de información Bayesiano (*Bayesian information criterion* - BIC) [38].

En este trabajo, dado que el objetivo no es comparar distintos métodos para la estimación del número de *clusters*, se ha optado por usar un número de *clusters* ligeramente superior al número de poblaciones anotadas. Esta decisión permite comparar con mayor claridad los resultados obtenidos mediante distintos métodos de *clustering*.

Adicionalmente, se ha empleado el algoritmo flowMeans en su modalidad automática, que estima el número óptimo de *clusters* de forma interna. Esta versión se ha empleado como referencia para contrastar sus resultados con los obtenidos al fijar el número de *clusters* de forma explícita. En adelante, se utilizará el término “Óptimo” para referirse a la versión con número fijado y “Automático” para la versión que lo estima. Esta comparación solo es posible con flowMeans, ya que FlowSOM no incluye un mecanismo automático de selección del número de *clusters*. No obstante, en el caso de FlowSOM se ha mantenido el tamaño por defecto de la red del SOM (una malla de 10×10), siguiendo las recomendaciones del paquete original.

## 2.4. Evaluación del rendimiento

Para comparar el rendimiento de los algoritmos de *clustering* presentados se ha seguido la metodología propuesta en el FlowCAP-I challenge [13], junto con las modificaciones que proponen Weber y Robinson [14]. En ambos enfoques, la evaluación parte de una tabla de contingencia 2x2. Para cada pareja formada por una población celular anotada mediante *manual gating* (clase real) y un *cluster* predicho por el algoritmo, se definen como verdaderos positivos (TP) los eventos correctamente asignados a dicha clase mediante ese *cluster*; falsos positivos (FP), los eventos asignados al *cluster* pero que no pertenecen a la clase; y falsos negativos (FN), los eventos que pertenecen a la clase pero no han sido asignados al *cluster* correspondiente. A partir de estas cantidades se calculan las métricas estándar de clasificación: precisión (*precision*), exhaustividad (*recall*) y el F-score (también denominado F-measure o F1-score), principal estadístico utilizado en este trabajo [39, 40].

Estas métricas se definen del siguiente modo:

$$Precision = \frac{TP}{TP + FP} \in [0, 1] \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \in [0, 1] \quad (4)$$

$$F\text{-score} = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \in [0, 1] \quad (5)$$

Desde el punto de vista interpretativo, la precisión mide la proporción de eventos asignados a un *cluster* que efectivamente pertenecen a la clase considerada, mientras que la exhaustividad cuantifica la proporción de eventos de la clase que han sido correctamente agrupados por el *cluster*. El F-score combina ambas dimensiones en una única medida, penalizando los casos en los que una de ellas sea muy baja. Así, un F-score de 1 indica una coincidencia perfecta entre el *cluster* y la clase, mientras que valores cercanos a 0 reflejan una clasificación deficiente.

Una vez calculadas las métricas de clasificación para cada pareja clase-*cluster*, es necesario resolver el problema de asignación, es decir, establecer una correspondencia entre *clusters* y poblaciones celulares para poder calcular un F-score global. Existen dos metodologías principales:

1. En el FlowCAP-I challenge [13], se asigna a cada población el *cluster* que maximiza su F-score individual. Este enfoque permite que varias poblaciones distintas compartan el mismo *cluster*, lo que puede ser útil en casos de solapamiento entre clases. El F-score global se obtiene como una media ponderada por el tamaño de cada población, según la expresión:

$$F(K, C) = \sum_{c_j \in C} \frac{|c_j|}{N} \max_{k_i \in K} \{F(k_i, c_j)\} \quad (6)$$

donde  $N$  es el número total de eventos,  $|c_j|$  el número de eventos de la clase  $j$ ,  $F(k_i, c_j)$  el F-score correspondiente a la pareja formada por la clase  $c_j$  y *cluster*  $k_i$ , y  $K$  y  $C$  el número total de *clusters* y clases, respectivamente.

2. Por su parte, Weber y Robinson [14] proponen utilizar el algoritmo del húngaro [41], que resuelve el problema de asignación de forma óptima a partir de la matriz de *F-scores*. Este enfoque impone una correspondencia uno a uno entre *clusters* y poblaciones, maximizando la suma total de *F-scores*. El F-score global se calcula como una media no

ponderada sobre las poblaciones asignadas:

$$F(K, C) = \sum_{c_j \in C} \frac{F(k_j, c_j)}{n} \quad (7)$$

donde  $n$  es el número de clases,  $F(k_j, c_j)$  el F-score correspondiente a la pareja formada por la clase  $c_j$  y *cluster*  $k_j$ , según la asignación obtenida mediante el húngaro.

Ambas metodologías presentan ventajas e inconvenientes. El enfoque del FlowCAP-I challenge permite una asignación más flexible, ya que permite asignar varias poblaciones a un mismo *cluster*. Esto puede ser útil cuando existen solapamientos entre poblaciones. Sin embargo, el cálculo del F-score como una media ponderada por el tamaño de cada población puede penalizar a las poblaciones raras que pueden ser de gran interés biológico. Por el contrario, la metodología de Weber y Robinson [14] impone una asignación uno a uno entre *clusters* y poblaciones, lo que puede resultar más equitativo en presencia de clases desbalanceadas, ya que no utiliza una media ponderada, pero impone restricciones que pueden dejar clases sin asignar si el número de *clusters* es insuficiente.

A lo largo de este trabajo, se hará referencia a estas metodologías como 'Máximo' (FlowCAP-I challenge) y 'Húngaro' (Weber y Robinson), respectivamente.

Además de las métricas descritas previamente, en este trabajo se ha calculado la exactitud (*accuracy*), como métrica complementaria, definida como la proporción total de eventos correctamente asignados. Aunque esta métrica no se contempla en los estudios originales, se incluye aquí para ofrecer una evaluación global adicional. No obstante, se reconoce que la *accuracy* puede ser menos informativa en contextos con clases desbalanceadas, ya que tiende a estar dominada por las poblaciones más numerosas.

Una vez resuelto el problema de asignación entre poblaciones reales y *clusters*—abordado mediante el algoritmo del húngaro o la estrategia del máximo—, se ha establecido un mapeo desde cada población del *manual gating* hacia su *cluster* correspondiente. Esta elección se debe a que, bajo el criterio del máximo, un mismo *cluster* puede estar asociado a múltiples poblaciones, lo que impide una asignación directa en la dirección opuesta. Finalmente, la *accuracy* se ha calculado como la proporción de eventos para los cuales el *cluster* asignado por el algoritmo de agrupamiento coincide con el *cluster* correspondiente a la población real de dicho evento, según el mapeo establecido.

Actualmente no existe un criterio claro sobre qué métrica debe emplearse para comparar algoritmos de *clustering*, ya que tanto el F-score como la *accuracy* presentan ventajas e inconvenientes [42]. La *accuracy* es ampliamente usada en el campo del aprendizaje automático, pero

no tiene en cuenta el desbalance entre clases, lo que puede dar lugar a resultados engañosos en presencia de poblaciones desiguales. Por el contrario, el F-score pondera tanto clases pequeñas como grandes, resultando más informativo aunque su interpretación sea menos directa.

Por último, se han calculado intervalos de confianza del 95 % para todas las métricas (*precisión*, *recall*, F-score y *accuracy*) recurriendo a técnicas de *bootstrap* sobre el conjunto de completo de eventos anotados [43]. Debido a la alta dimensionalidad de los conjuntos de datos y las limitaciones computacionales, se utilizó el esquema de *cheap bootstrap* propuesto por Lam y Liu [44], con 10 repeticiones. Bajo el supuesto  $p = o(n)$ , estos intervalos se estimaron utilizando la distribución  $t$  de Student, adaptada a entornos de alta dimensionalidad.

## 2.5. Estabilidad de los resultados

Tanto flowMeans como FlowSOM tienen un componente de aleatoriedad inherente a su funcionamiento, de forma que los resultados pueden variar entre ejecuciones si no se tiene en cuenta este aspecto. En este contexto, seleccionar una semilla para el generador de números aleatorios se convierte en un elemento clave para garantizar la reproducibilidad de los resultados.

En el caso de flowMeans, la aleatoriedad afecta a la elección inicial de los centroides, lo que puede dar lugar a distintas soluciones. Este efecto se puede mitigar parcialmente si se utiliza un número inicial de *clusters* mayor al número esperado de poblaciones, lo que favorece una mejor cobertura del espacio de datos desde el inicio.

Por su parte, FlowSOM incorpora aleatoriedad en dos fases: al inicializar el SOM, y en el *meta-clustering* posterior, donde se aplica *clustering* a los nodos del mapa. Esta doble fuente de aleatoriedad puede amplificar las diferencias entre ejecuciones.

Este aspecto fue evaluado por Weber y Robinson [14], quienes analizaron la variabilidad en los resultados al ejecutar cada algoritmo con 30 semillas distintas. Tal como se muestra en la Figura 4, los resultados obtenidos con FlowSOM presentan una mayor variabilidad entre ejecuciones con distintas semillas, en comparación con flowMeans.

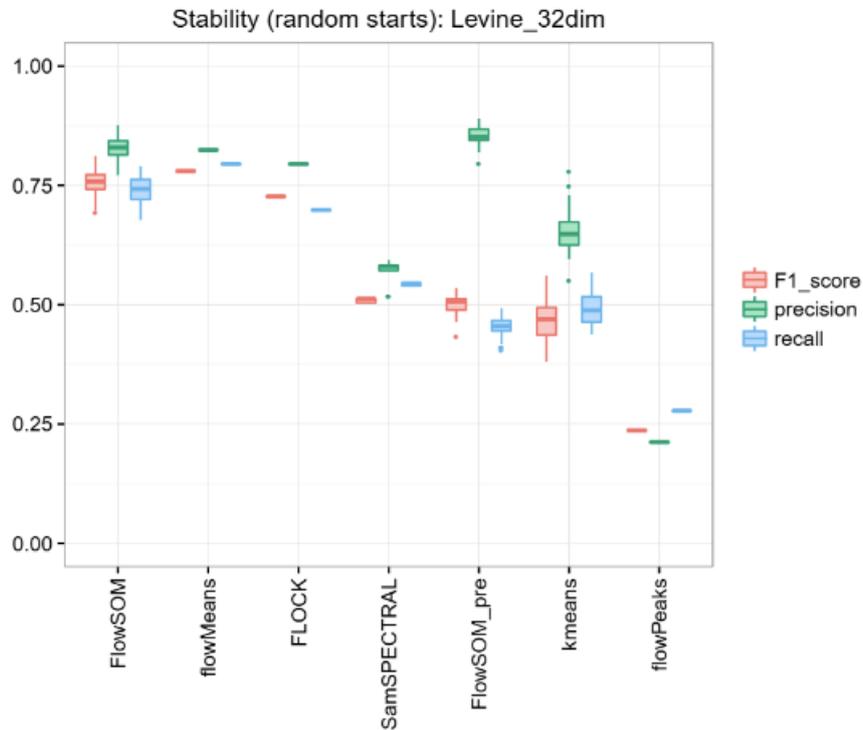


Figura 4: Estabilidad de los *clusters* al inicializarse con 30 semillas diferentes (Fuente: Figura S17, Weber y Robinson [14]).

Con base en estos resultados, en este trabajo se evalúa el efecto de la aleatoriedad en FlowSOM mediante diez ejecuciones con semillas distintas, aplicadas al conjunto de datos del IBGM, que presenta mayor dimensionalidad que los usados en el estudio original. Esta evaluación permitirá analizar la variabilidad de los *F-scores* entre ejecuciones, así como las diferencias en estabilidad entre las distintas poblaciones celulares.

A su vez, cabe destacar que, según Weber y Robinson [14] las distintas poblaciones pueden presentar valores muy distintos de F-score. Por tanto, el análisis planteado prestará atención a las diferencias de rendimiento por población. Este enfoque no es aplicable a los conjuntos de datos del FlowCAP-I challenge, debido a la inconsistencia en el etiquetado entre las distintas repeticiones del experimento.

## 2.6. Clasificación jerárquica

La clasificación celular presenta una estructura jerárquica en la que una misma célula puede estar incluida en distintos niveles de especificidad biológica. En la Figura 5 se muestra una simplificación de la jerarquía de poblaciones celulares anotadas en el conjunto de datos del IBGM; la clasificación completa puede consultarse en el esquema del Apéndice A. Esta estructura permite reflejar relaciones de inclusión entre distintos tipos celulares.

A diferencia de los conjuntos del FlowCAP-I challenge, que solo contienen una etiqueta numérica por evento, el conjunto del IBGM conserva esta estructura jerárquica, lo que permite aplicar una metodología de clasificación multinivel. Para ello, se ha utilizado el algoritmo FlowSOM de forma iterativa, aplicada sucesivamente sobre subconjuntos definidos según distintos niveles de la jerarquía.

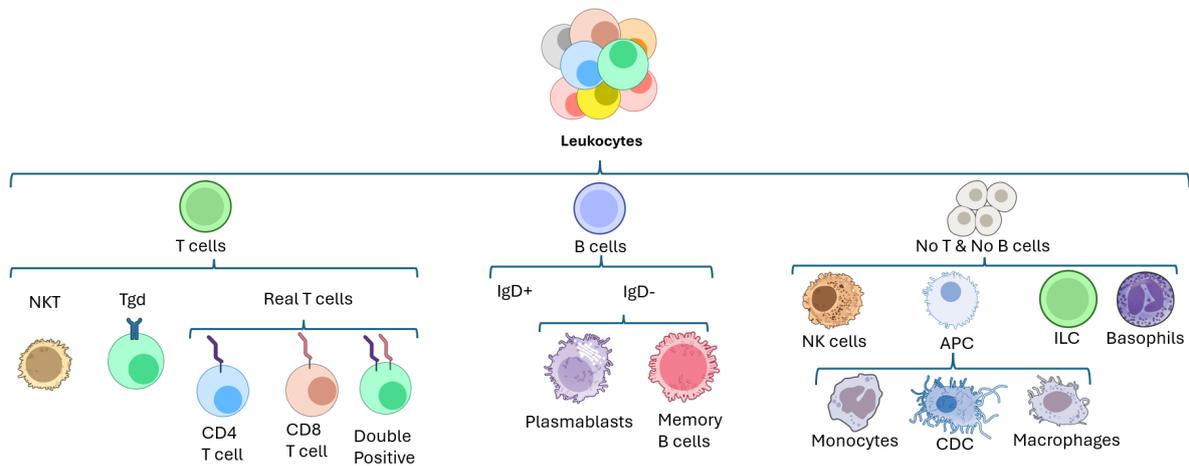


Figura 5: Esquema de la jerarquía de las poblaciones celulares del conjunto IBGM, elaborado a partir de ilustraciones de BioArt [45].

En una primera etapa, se realizó una primera clasificación en tres categorías principales: *T cells*, *B cells* y *No T & No B cells*, utilizando como criterio de asignación el *cluster* con mayor F-score individual para cada población. A partir de esta partición inicial, se generaron subconjuntos de datos para cada grupo, sobre los que se volvió a aplicar FlowSOM de forma independiente.

La subclasificación por grupo fue:

- Para las *T cells*, se llevó a cabo una segunda clasificación en tres subgrupos: *NKT*, *Tgd* y *Real T cells*. Sobre estos últimos se aplicó una tercera etapa de clasificación para distinguir

entre las poblaciones  $DP$ ,  $CD4^+$  y  $CD8^+$ , seguidas de FlowSOM por separado sobre las dos últimas.

- En el caso de las *B cells*, se hizo una segunda clasificación en dos subgrupos:  $IgD^+$  e  $IgD^-$ , posteriormente se aplicó FlowSOM de forma independiente a  $IgD^-$ .
- En el grupo de las *No T & No B cells*, se realizó una clasificación en cuatro grupos: *NK cells*, *APC*, *ILC* y *Basophils*. El análisis posterior se centró en los subconjuntos de *NK cells* y *APC*.

Esta clasificación jerárquica se evaluó bajo dos enfoques distintos en cuanto a los marcadores utilizados:

1. Todos los marcadores disponibles, para aprovechar al máximo la información contenida en los datos.
2. Marcadores esenciales para la diferenciación celular, según los criterios utilizados en el *manual gating* original. Este enfoque busca una mayor similitud con el análisis convencional realizados por expertos.

## 2.7. Implementación computacional

Todos los resultados presentados en este trabajo se han realizado utilizando el lenguaje de programación R. Para el tratamiento de los archivos de citometría de flujo, en formato .fcs, se ha empleado la librería *FlowCore* [46], disponible públicamente en Bioconductor [26]. Asimismo, se han empleado las implementaciones disponibles en Bioconductor [26] de los algoritmos de *flowMeans* [47] y *FlowSOM* [48], a través de los paquetes que llevan sus mismos nombres. Por último, se ha usado la implementación del algoritmo del húngaro incluida en la librería *clue* [49]. El código completo empleado en el análisis se encuentra disponible en mi repositorio de GitHub.

Los algoritmos se ejecutaron en un equipo del IBGM con 64GB de RAM y un procesador Intel Core i7-9700 con 8 núcleos, de forma que los tiempos de ejecución obtenidos sean comparables. Por otro lado, la evaluación y análisis de resultados se llevaron a cabo en un equipo portátil con 16GB de RAM y un procesador Intel Core i7-1185G7 con 4 núcleos.

## 3. Resultados

### 3.1. Comparativa de rendimiento: flowMeans frente FlowSOM

Los resultados de los algoritmos flowMeans y FlowSOM se resumen en las Tablas 1, 2 y 3, que recogen respectivamente los valores de F-score, *accuracy* y tiempo de ejecución para los tres conjuntos de datos utilizados. Las métricas se han calculado bajo las dos estrategias de asignación descritas en la metodología: el criterio del F-score máximo y el algoritmo del húngaro.

Para cada conjunto de datos se han considerado dos configuraciones por algoritmo:

- Una configuración óptima, en la que el número de *clusters* se ha fijado manualmente en función del número de poblaciones objetivo conocidas.
- En el caso de flowMeans, una segunda configuración automática, en la que el número de *clusters* es estimado internamente por el algoritmo.

Concretamente:

- En Normal Donor se fijaron 7 *clusters*, uno por cada población anotada. La versión automática de flowMeans produjo entre 7 y 12 *clusters* según la repetición. Tal como se indicó en la metodología, algunas repeticiones fueron descartadas por errores de singularidad.
- En West Nile Virus se fijaron 4 *clusters*. En modo automático, flowMeans generó entre 2 y 7 *clusters*.
- En el conjunto del IBGM, dada su mayor complejidad y dimensionalidad, se fijaron 60 *clusters*, valor ligeramente superior al número de poblaciones conocidas. La versión automática de flowMeans generó 31 *clusters*.

#### 3.1.1. Comparativa de F-score

La Tabla 1 muestra los valores globales de F-score para cada combinación de algoritmo, estrategia de asignación y conjunto de datos. Se observa que:

- Para los conjuntos del FlowCAP-I challenge (Normal Donor y West Nile Virus), flowMeans obtiene sistemáticamente mejores *F-scores* que FlowSOM, especialmente bajo el criterio del F-score máximo.
- En el conjunto del IBGM, que presenta mayor complejidad y dimensionalidad, el rendimiento de FlowSOM supera ligeramente al de flowMeans.

Además, puede apreciarse que el método de asignación tiene un impacto importante en los resultados: los *F-scores* calculados mediante el algoritmo del húngaro son considerablemente más bajos, en algunos casos reduciéndose a la mitad respecto a los obtenidos por el criterio del máximo. Esto refleja la mayor exigencia del algoritmo del húngaro, al forzar una asignación uno a uno entre *clusters* y poblaciones.

En cualquier caso, cabe destacar que los valores absolutos del F-score en el conjunto de datos del IBGM son notablemente más bajos que en los conjuntos del FlowCAP-I challenge, independientemente del algoritmo o del método de asignación empleado. Este resultado pone de manifiesto la mayor dificultad del problema en escenarios de alta dimensionalidad y mayor complejidad biológica.

|                        | FlowMeans                  |                            | FlowSOM                    |
|------------------------|----------------------------|----------------------------|----------------------------|
|                        | Automático                 | Óptimo                     | Óptimo                     |
| <b>Normal Donor</b>    |                            |                            |                            |
| Máximo                 | 0.9104<br>[0.9098, 0.9111] | 0.9042<br>[0.9007, 0.9076] | 0.8090<br>[0.8063, 0.8117] |
| Húngaro                | 0.4693<br>[0.4674, 0.4712] | 0.4036<br>[0.3978, 0.4095] | 0.2409<br>[0.2356, 0.2462] |
| <b>West Nile Virus</b> |                            |                            |                            |
| Máximo                 | 0.8830<br>[0.8826, 0.8835] | 0.8849<br>[0.8843, 0.8855] | 0.7881<br>[0.7873, 0.7889] |
| Húngaro                | 0.6949<br>[0.6931, 0.6966] | 0.6616<br>[0.6608, 0.6624] | 0.5908<br>[0.5889, 0.5927] |
| <b>IBGM</b>            |                            |                            |                            |
| Máximo                 | 0.2684<br>[0.2680, 0.2688] | 0.2841<br>[0.2836, 0.2846] | 0.3479<br>[0.3474, 0.3484] |
| Húngaro                | 0.1307<br>[0.1301, 0.1314] | 0.1680<br>[0.1671, 0.1688] | 0.2149<br>[0.2139, 0.2158] |

Tabla 1: F-score para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

### 3.1.2. Comparativa del *accuracy*

La Tabla 2 presenta los valores de *accuracy*, calculados como la proporción total de eventos correctamente asignados. Como es esperable, los valores de *accuracy* son más altos que los *F-scores*, ya que esta métrica no penaliza el desbalance entre clases.

En los conjuntos de menor dimensionalidad (Normal Donor y West Nile Virus), flowMeans y FlowSOM obtienen valores de *accuracy* similares. En Normal Donor, flowMeans presenta una ligera ventaja en ambas estrategias de asignación, mientras que en West Nile Virus, FlowSOM alcanza una mayor *accuracy* cuando se utiliza el criterio del F-score máximo, pero es superado por flowMeans al aplicar el algoritmo del húngaro.

Es importante subrayar que, a pesar de ser superiores a los *F-scores*, los valores de *accuracy* en el conjunto del IBGM siguen siendo bajos en términos absolutos. Esto indica que, incluso a nivel global, la asignación de eventos a las poblaciones reales es limitada, lo que refuerza la complejidad inherente al análisis automático en entornos de alta dimensionalidad.

|                        | FlowMeans                  |                            | FlowSOM                    |
|------------------------|----------------------------|----------------------------|----------------------------|
|                        | Automático                 | Óptimo                     | Óptimo                     |
| <b>Normal Donor</b>    |                            |                            |                            |
| Máximo                 | 0.9021<br>[0.9011, 0.9030] | 0.9206<br>[0.9158, 0.9254] | 0.9117<br>[0.9078, 0.9157] |
| Húngaro                | 0.8909<br>[0.8901, 0.8917] | 0.8986<br>[0.8950, 0.9023] | 0.8303<br>[0.8269, 0.8336] |
| <b>West Nile Virus</b> |                            |                            |                            |
| Máximo                 | 0.8496<br>[0.8488, 0.8505] | 0.8442<br>[0.8433, 0.8451] | 0.9390<br>[0.9386, 0.9395] |
| Húngaro                | 0.8306<br>[0.8301, 0.8312] | 0.8307<br>[0.8300, 0.8315] | 0.7933<br>[0.7924, 0.7943] |
| <b>IBGM</b>            |                            |                            |                            |
| Máximo                 | 0.3315<br>[0.3284, 0.3345] | 0.3282<br>[0.3276, 0.3288] | 0.3855<br>[0.3790, 0.3921] |
| Húngaro                | 0.2409<br>[0.2372, 0.2445] | 0.2392<br>[0.2368, 0.2416] | 0.2990<br>[0.2982, 0.2999] |

Tabla 2: *Accuracy* para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

### 3.1.3. Comparativa de tiempos de ejecución

La Tabla 3 muestra los tiempos de ejecución requeridos por los algoritmos flowMeans y FlowSOM al aplicarse sobre los tres conjuntos de datos analizados. Los resultados evidencian una clara ventaja computacional de FlowSOM, especialmente en escenarios de alta dimensionalidad.

FlowSOM es notablemente más eficiente que flowMeans en todos los casos, con diferencias especialmente marcadas en el conjunto del IBGM. Mientras que flowMeans requiere más de 14 horas en modo óptimo y más de 17 en automático, FlowSOM completa la tarea en menos de tres minutos.

Esta disparidad en el rendimiento computacional resulta especialmente relevante considerando que los conjuntos de datos del FlowCAP-I challenge fueron generados hace más de una década y presentan una dimensionalidad moderada. En cambio, el conjunto IBGM refleja con mayor fidelidad el volumen y la complejidad de los datos utilizados actualmente en citometría de flujo, lo que refuerza la utilidad práctica de FlowSOM en contextos reales.

|                  | <b>Normal Donor</b> | <b>West Nile Virus</b> | <b>IBGM</b> |
|------------------|---------------------|------------------------|-------------|
| <b>FlowMeans</b> |                     |                        |             |
| Automático       | 00:12:29            | 00:01:28               | 17:28:45    |
| Óptimo           | 00:12:08            | 00:01:25               | 14:23:36    |
| <b>FlowSOM</b>   |                     |                        |             |
| Óptimo           | 00:01:33            | 00:00:41               | 00:02:34    |

Tabla 3: Tiempo de ejecución para los distintos algoritmos en formato hh:mm:ss.

En resumen, flowMeans y FlowSOM ofrecen resultados equiparables a nivel de su F-score, cuando hablamos de sus formas óptimas y resolviendo el problema de asignación por el máximo. No obstante, el coste computacional de flowMeans es significativamente mayor al de FlowSOM, sin conseguir una mejora significativa en los valores del F-score.

### 3.1.4. Análisis por subpoblaciones celulares

La Figura 6 muestra los valores de F-score obtenidos para cada una de las 51 poblaciones celulares anotadas en el conjunto de datos del IBGM, ordenadas de mayor a menor abundancia. Esta visualización permite evaluar el rendimiento de los algoritmos flowMeans y FlowSOM no

solo a nivel global, sino también a nivel de subpoblación, destacando en qué casos se obtiene una clasificación más precisa o más deficiente.

Tal como era esperable, los *F-scores* más altos se concentran en las poblaciones más abundantes, mientras que los valores tienden a disminuir notablemente en las poblaciones minoritarias. Esta tendencia se mantiene independientemente del algoritmo utilizado o del método de asignación aplicado (máximo o húngaro), lo que pone de manifiesto las dificultades inherentes a la detección automática de poblaciones raras en contextos de alta dimensionalidad.

Además, los resultados confirman que el criterio del F-score máximo tiende a beneficiar el rendimiento estimado en estas poblaciones, en comparación con el algoritmo del húngaro, que impone restricciones más exigentes en la asignación.

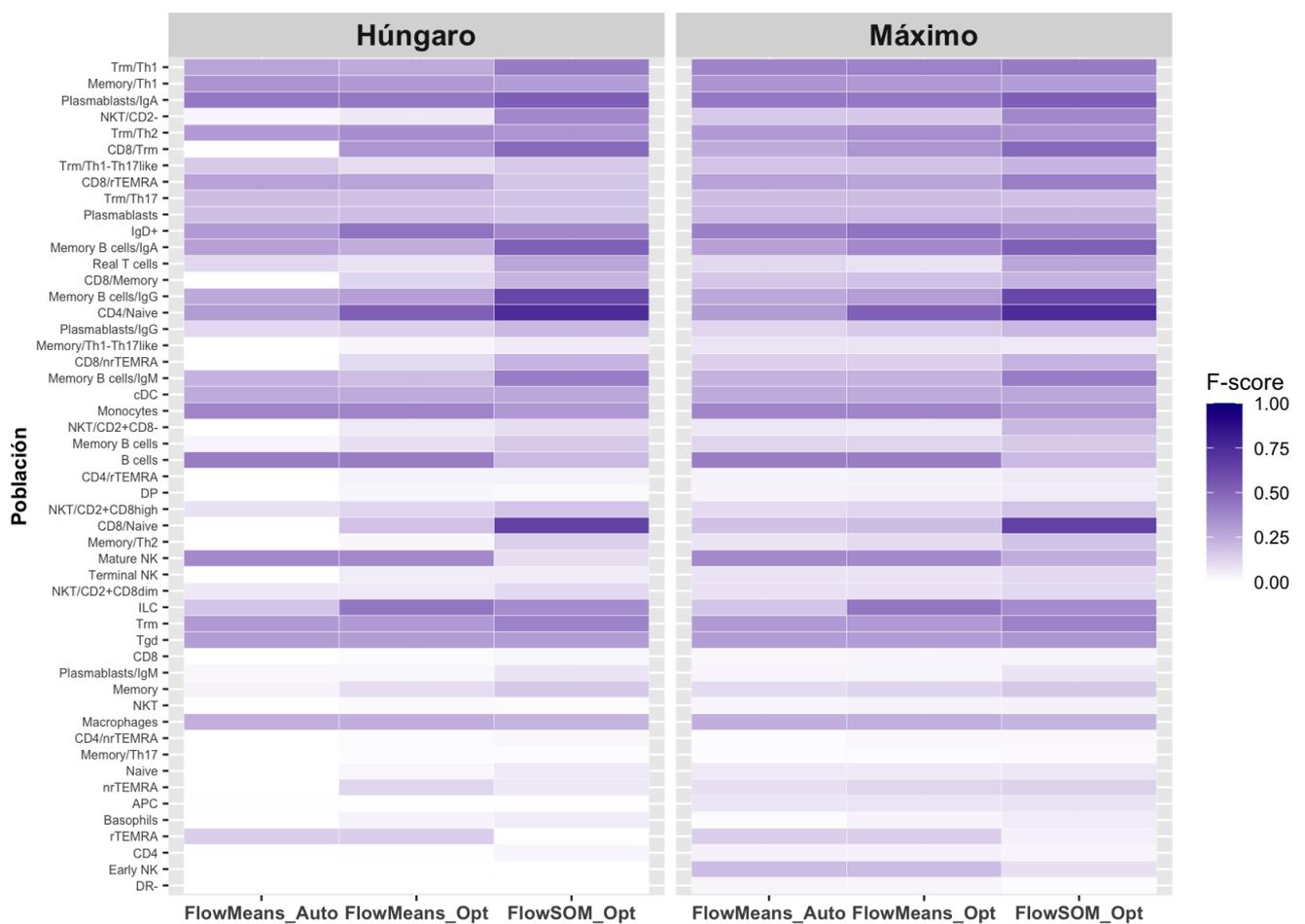


Figura 6: Evolución del F-score por poblaciones, ordenadas de mayor a menor abundancia, para cada método de asignación y algoritmo utilizado.

Los mapas de expresión por *cluster* (*heatmaps*), generados a partir de los marcadores celulares, están disponibles en el repositorio de GitHub. Estos gráficos permiten una interpretación biológica más detallada de los *clusters* obtenidos.

### 3.2. Evaluación de la estabilidad frente a la semilla aleatoria

Tal como se indicó en la sección de metodología, la evaluación de la estabilidad frente a la aleatoriedad se ha centrado exclusivamente en el algoritmo FlowSOM aplicado al conjunto de datos del IBGM. Esta decisión se debe, por un lado, a la falta de etiquetado consistente en las repeticiones del FlowCAP-I challenge, que impide realizar evaluaciones comparables bajo diferentes semillas aleatorias. Por otro lado, el conjunto del IBGM representa un escenario más realista por su alta dimensionalidad y complejidad biológica, donde el impacto de la aleatoriedad puede ser más relevante.

Además, aunque tanto FlowSOM como flowMeans incorporan elementos aleatorios en su funcionamiento, estudios previos como el de Weber y Robinson [14] han mostrado que FlowSOM tiende a presentar mayor variabilidad entre ejecuciones, lo que refuerza el interés de centrar el análisis en este algoritmo.

Tras la ejecución del algoritmo FlowSOM con diez semillas distintas sobre el conjunto IBGM, se observa que las métricas de rendimiento presentan una variabilidad limitada, lo que indica un comportamiento relativamente estable frente a la aleatoriedad.

La Figura 7 resume gráficamente la distribución de los valores obtenidos para la *precision*, *recall*, F-score y *accuracy* bajo ambos métodos de asignación (máximo y algoritmo del húngaro), mientras que la Tabla 4 recoge sus valores estimados junto con los intervalos de confianza al 95 % construidos mediante *cheap bootstrap* [44].

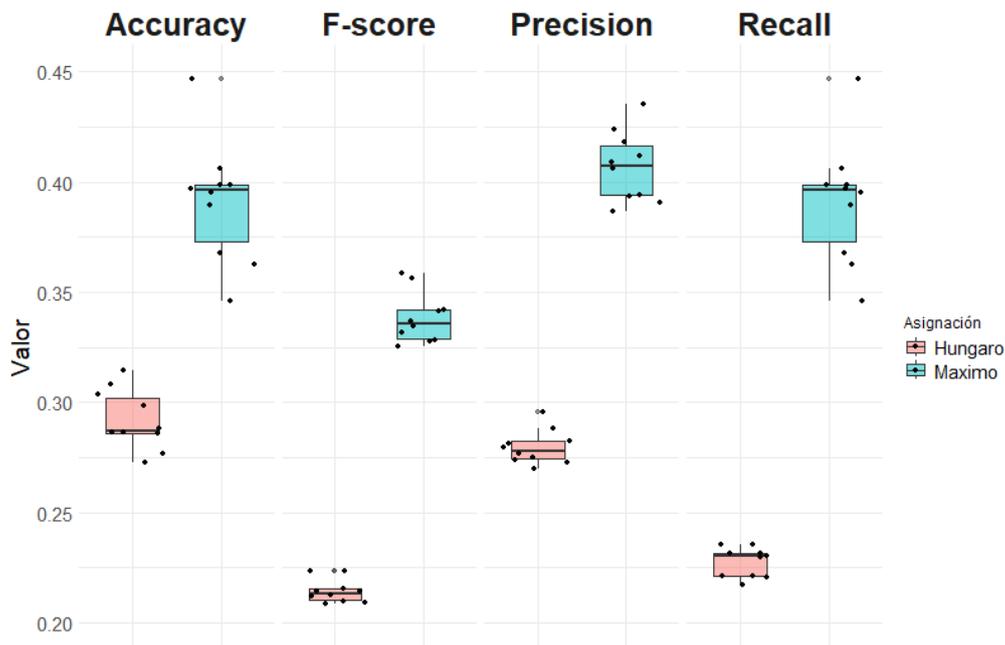


Figura 7: Distribución de las métricas de rendimiento (*precision*, *recall*, F-score y *accuracy*) obtenidas tras ejecutar FlowSOM con diez semillas distintas sobre el conjunto IBGM. Se comparan los resultados según el tipo de asignación empleado (máximo o húngaro).

En todos los casos, las métricas obtenidas mediante el criterio del F-score máximo son superiores a las derivadas del algoritmo del húngaro. No obstante, los resultados muestran que la variabilidad entre ejecuciones es mayor bajo el criterio del máximo, mientras que los resultados asociados al algoritmo del húngaro son más consistentes pero sistemáticamente más bajos.

Estos patrones sugieren que, aunque FlowSOM introduce cierto grado de aleatoriedad en la inicialización y en la fase de *meta-clustering*, su impacto sobre el rendimiento final es relativamente acotado, al menos en un contexto de alta dimensionalidad como el del conjunto del IBGM. La elección del método de asignación influye más en los resultados que la semilla en sí.

|                  | <b>Máximo</b>              | <b>Húngaro</b>             |
|------------------|----------------------------|----------------------------|
| <b>F-score</b>   | 0.3383<br>[0.3261, 0.3583] | 0.2143<br>[0.2087, 0.2234] |
| <b>Accuracy</b>  | 0.3909<br>[0.3499, 0.4376] | 0.2920<br>[0.2735, 0.3132] |
| <b>Precision</b> | 0.4069<br>[0.3877, 0.4328] | 0.2796<br>[0.2705, 0.2942] |
| <b>Recall</b>    | 0.3909<br>[0.3499, 0.4376] | 0.2273<br>[0.2178, 0.2942] |

Tabla 4: Valores promedio de *precision*, *recall*, F-score y *accuracy* obtenidos por FlowSOM en 10 ejecuciones independientes con semillas distintas sobre el conjunto IBGM. Se muestran los valores medios junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* [44]. Se comparan los resultados bajo los métodos de asignación por máximo y por algoritmo del húngaro.

### 3.3. Clasificación jerárquica con FlowSOM

En este apartado se evalúa el rendimiento del algoritmo FlowSOM cuando se aplica de forma jerárquica sobre subconjuntos definidos por niveles biológicos de especificidad celular, de acuerdo con la estructura descrita en el apéndice A. Respecto al número de *clusters*, se ha tomado un valor ligeramente superior al número de poblaciones conocidas, para capturar grupos residuales del modelo de nivel superior. Estas poblaciones residuales se han incluido en la categoría “Otro” para no comprometer la interpretación de las clases principales. Este enfoque permite no solo reducir la complejidad del *clustering* en cada una de las etapas, sino también reflejar mejor la organización natural de las poblaciones inmunológicas.

El análisis se ha realizado tanto utilizando todos los marcadores disponibles como empleando únicamente aquellos relevantes para cada nivel, tal como se hace en el *manual gating*. Los resultados obtenidos en ambos casos permiten valorar la utilidad de incorporar conocimiento biológico en el resultado del análisis.

#### 3.3.1. Resultados utilizando marcadores relevantes por nivel

La Tabla 5 presenta los valores globales de F-score y *accuracy* obtenidos en cada uno de los niveles del análisis jerárquico, tras aplicar FlowSOM sobre los subconjuntos celulares

correspondientes y emplear únicamente los marcadores relevantes. Se muestran también los intervalos de confianza al 95 %, obtenidos mediante *bootstrap* con el enfoque *cheap bootstrap* descrito en la sección de metodología [44].

En general, se observa que la clasificación es más precisa en los niveles superiores de la jerarquía, como el nivel general o los grupos amplios de *T cells* y *B cells*. Sin embargo, a medida que se profundiza en subpoblaciones más específicas (como  $CD4^+$ ,  $CD8^+$ , *Plasmablasts* o *NKT*), el rendimiento disminuye, lo cual es esperable debido a su menor tamaño y mayor solapamiento con otras clases. Aun así, el enfoque jerárquico permite mantener niveles razonables de F-score en muchas de estas subpoblaciones.

| Grupo                        | F-score                    |                            | Accuracy                   |                            |
|------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
|                              | Máximo                     | Húngaro                    | Máximo                     | Húngaro                    |
| <b>General</b>               | 0.9871<br>[0.9869, 0.9872] | 0.9650<br>[0.9645, 0.9656] | 0.9873<br>[0.9871, 0.9874] | 0.9873<br>[0.9871, 0.9874] |
| <b>T cells</b>               | 0.9715<br>[0.9710, 0.9720] | 0.8349<br>[0.8317, 0.8381] | 0.9635<br>[0.9629, 0.9640] | 0.9635<br>[0.9629, 0.9640] |
| NKT                          | 0.6082<br>[0.6047, 0.6116] | 0.3826<br>[0.3798, 0.3854] | 0.5154<br>[0.5121, 0.5188] | 0.4782<br>[0.4755, 0.4809] |
| Tgd                          | 0.4004<br>[0.3911, 0.4097] | 0.3484<br>[0.3334, 0.3633] | 0.3995<br>[0.3556, 0.4434] | 0.3275<br>[0.3190, 0.3359] |
| Real T cells                 | 0.9069<br>[0.9061, 0.9077] | 0.5516<br>[0.5502, 0.5529] | 0.8915<br>[0.8905, 0.8925] | 0.8804<br>[0.8795, 0.8814] |
| CD4 <sup>+</sup>             | 0.6025<br>[0.6017, 0.6033] | 0.3570<br>[0.3550, 0.3590] | 0.6615<br>[0.6511, 0.6719] | 0.5857<br>[0.5848, 0.5866] |
| CD8 <sup>+</sup>             | 0.4514<br>[0.4499, 0.4528] | 0.2805<br>[0.2788, 0.2823] | 0.8045<br>[0.8027, 0.8063] | 0.4406<br>[0.4483, 0.4527] |
| <b>B cells</b>               | 0.9169<br>[0.9156, 0.9183] | 0.7002<br>[0.6967, 0.7037] | 0.9208<br>[0.9196, 0.9219] | 0.9208<br>[0.9196, 0.9219] |
| IgD <sup>-</sup>             | 0.9491<br>[0.9486, 0.9497] | 0.9439<br>[0.9432, 0.9446] | 0.9039<br>[0.9029, 0.9049] | 0.9039<br>[0.9029, 0.9049] |
| Plasmablasts                 | 0.5931<br>[0.5915, 0.5946] | 0.3506<br>[0.3483, 0.3530] | 0.5752<br>[0.5723, 0.5780] | 0.4000<br>[0.3979, 0.4019] |
| Memory B cells               | 0.7096<br>[0.6935, 0.7257] | 0.5899<br>[0.5774, 0.6024] | 0.2217<br>[0.2037, 0.2396] | 0.2008<br>[0.1872, 0.2144] |
| <b>No T &amp; No B cells</b> | 0.5870<br>[0.5846, 0.5894] | 0.3707<br>[0.3523, 0.3892] | 0.8825<br>[0.8790, 0.8860] | 0.5393<br>[0.5168, 0.5618] |
| NK cells                     | 0.4035<br>[0.3987, 0.4084] | 0.3192<br>[0.3039, 0.3344] | 0.2372<br>[0.2334, 0.2410] | 0.2372<br>[0.2250, 0.2495] |
| APC                          | 0.5569<br>[0.5537, 0.5600] | 0.2080<br>[0.2024, 0.2135] | 0.5523<br>[0.5485, 0.5561] | 0.2757<br>[0.1960, 0.3555] |

Tabla 5: F-score y *accuracy* en el análisis jerárquico con FlowSOM empleando solo los marcadores relevantes por nivel. Se muestran los valores estimados junto con los intervalos *bootstrap* de confianza al 95 %.

### 3.3.2. Resultados utilizando todos los marcadores disponibles

En la Tabla 6 se recogen los resultados obtenidos al aplicar el mismo análisis jerárquico, pero empleando todos los marcadores del conjunto de datos del IBGM, sin filtrar según su relevancia para cada nivel. No se incluyen valores para el nivel *NK cells*, ya que en el nivel anterior *No T & No B cells* solo se identifican *APC* y *Basophils*.

Al comparar con los resultados anteriores, se aprecia que el uso exclusivo de los marcadores relevantes suele mejorar las métricas de rendimiento, especialmente en niveles bajos como *CD4<sup>+</sup>*, *CD8<sup>+</sup>* o *Plasmablasts*. Esto sugiere que la efectividad de este filtrado depende del grado de solapamiento y especificidad de las poblaciones implicadas.

| Grupo                        | F-score                    |                            | Accuracy                   |                            |
|------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
|                              | Máximo                     | Húngaro                    | Máximo                     | Húngaro                    |
| <b>General</b>               | 0.6888<br>[0.6882, 0.6894] | 0.3122<br>[0.3113, 0.3130] | 0.9513<br>[0.9506, 0.9520] | 0.7046<br>[0.7000, 0.7094] |
| <b>T cells</b>               | 0.6934<br>[0.6925, 0.6943] | 0.3112<br>[0.3095, 0.3128] | 0.7023<br>[0.7012, 0.7035] | 0.6162<br>[0.6127, 0.6197] |
| NKT                          | 0.1207<br>[0.1196, 0.1217] | 0.0402<br>[0.0390, 0.0413] | 0.0850<br>[0.0844, 0.0856] | 0.0626<br>[0.0618, 0.0633] |
| Tgd                          | 0.3888<br>[0.3688, 0.4088] | 0.2088<br>[0.1498, 0.2679] | 0.3075<br>[0.2810, 0.3341] | 0.2297<br>[0.1794, 0.2800] |
| Real T cells                 | 0.6081<br>[0.6072, 0.6090] | 0.2022<br>[0.2016, 0.2029] | 0.5958<br>[0.5947, 0.5969] | 0.4368<br>[0.4361, 0.4375] |
| CD4 <sup>+</sup>             | 0.3171<br>[0.3162, 0.3180] | 0.1456<br>[0.1446, 0.1465] | 0.4457<br>[0.4445, 0.4469] | 0.2157<br>[0.2140, 0.2173] |
| CD8 <sup>+</sup>             | 0.3209<br>[0.3194, 0.3224] | 0.0773<br>[0.0765, 0.0781] | 0.1790<br>[0.1780, 0.1800] | 0.0689<br>[0.0675, 0.0703] |
| <b>B cells</b>               | 0.2937<br>[0.2929, 0.2945] | 0.1597<br>[0.1574, 0.1621] | 0.2426<br>[0.2418, 0.2435] | 0.2053<br>[0.1986, 0.2120] |
| IgD <sup>-</sup>             | 0.3500<br>[0.3493, 0.3509] | 0.3981<br>[0.3972, 0.3991] | 0.1622<br>[0.1612, 0.1633] | 0.1622<br>[0.1611, 0.1633] |
| Plasmablasts                 | 0.5281<br>[0.5267, 0.5296] | 0.1785<br>[0.1775, 0.1795] | 0.1225<br>[0.1207, 0.1242] | 0.0917<br>[0.0906, 0.0928] |
| Memory B cells               | 0.3379<br>[0.3351, 0.3406] | 0.1157<br>[0.1134, 0.1180] | 0.5151<br>[0.5129, 0.5172] | 0.2884<br>[0.2621, 0.3147] |
| <b>No T &amp; No B cells</b> | 0.5500<br>[0.5406, 0.5595] | 0.3112<br>[0.2764, 0.3460] | 0.2376<br>[0.2263, 0.2489] | 0.2376<br>[0.2177, 0.2576] |
| NK cells                     | -                          |                            |                            |                            |
| APC                          | 0.4595<br>[0.4436, 0.4753] | 0.2409<br>[0.2209, 0.2608] | 0.2986<br>[0.2809, 0.3163] | 0.2182<br>[0.2018, 0.2346] |

Tabla 6: F-score y *accuracy* en el análisis jerárquico con FlowSOM empleando todos los marcadores disponibles. Se muestran los valores estimados junto con los intervalos *bootstrap* de confianza al 95 %.

### 3.3.3. Visualización de resultados: gráficos *Star Pie*

Para ilustrar la distribución de las poblaciones asignadas a cada *cluster* en los distintos niveles del análisis jerárquico, se presentan a continuación los gráficos *Star Pie* obtenidos en los

niveles *General* y *T cells*, tanto para el modelo con marcadores relevantes como para el modelo con todos los marcadores. Estos niveles han sido seleccionados por su valor representativo: el nivel *General* permite visualizar la asignación global de las principales poblaciones celulares, mientras que el nivel *T cells*, dada su complejidad y riqueza jerárquica, permite apreciar con mayor detalle el impacto de la selección de marcadores y del método de asignación. Los gráficos correspondientes al resto de niveles jerárquicos, para ambas metodologías, se incluyen en el Apéndice E.

En estos gráficos, cada nodo representa una neurona del mapa auto-organizado de Flow-SOM y contiene sectores circulares que indican el porcentaje de eventos, asignados a dicho nodo, pertenecientes a cada población de referencia. El color del contorno del nodo indica la agrupación superior obtenida tras el paso de *meta-clustering*. Estas representaciones permiten evaluar visualmente la coherencia del *clustering* con las poblaciones reales.

En el modelo representado en la Figura 8, que corresponde a nivel *General* e incluye únicamente los marcadores esenciales, se observa una asignación coherente entre los dos métodos de asignación: el grupo de *B cells* se asocia al *cluster 1*, las *T cells* al *cluster 3* y las *No T & No B cells* al *cluster 2*, tanto bajo el criterio del F-score máximo como con el algoritmo del húngaro.

Por el contrario, en el modelo con todos los marcadores (Figura 9), el criterio del máximo agrupa los *B cells* y *T cells* en el *cluster 1*, asignando el grupo *No T & No B* al *cluster 2*. La asignación del algoritmo del húngaro coincide con esta distribución, salvo por los *B cells*, que pasan a asignarse al *cluster 3*.

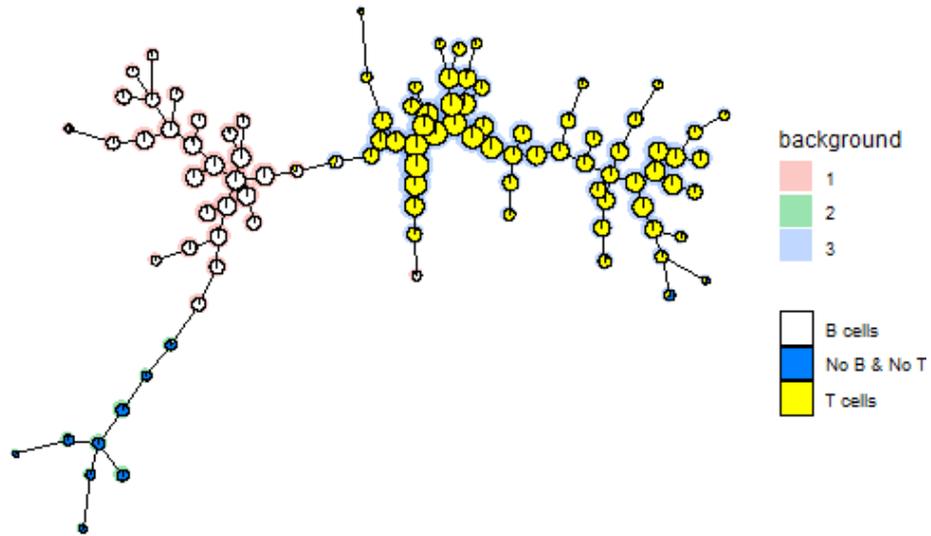


Figura 8: *Star Pie* del nivel *General* incluyendo solo los marcadores esenciales para la caracterización celular. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*.

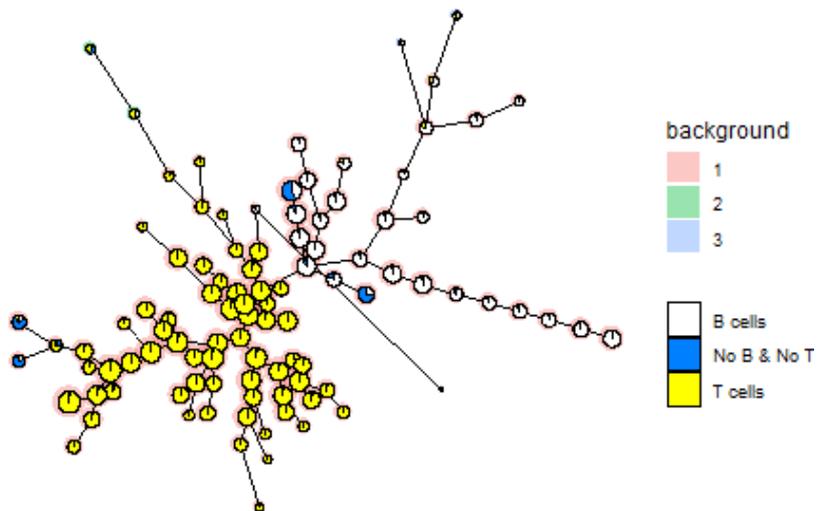


Figura 9: *Star Pie* del nivel *General* incluyendo todos los marcadores. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*.

En el análisis del nivel *T cells*, los resultados también varían según los marcadores considerados. Con solo los marcadores esenciales (Figura 10), ambos métodos de asignación producen resultados idénticos: las *Real T cells* se asignan al *cluster 1*, las *NKT* al *cluster 2* y las *Tgd* al *cluster 3*. En cambio, en el modelo con todos los marcadores (Figura 11), el criterio del máximo asigna todos los grupos al *cluster 1*, salvo las *Tgd*, que se agrupan en el *cluster 4*. En este mismo escenario, el algoritmo del húngaro asigna las *Real T cells* al *cluster 1*, las *NKT* al *cluster 2* y las *Tgd* al *cluster 4*.

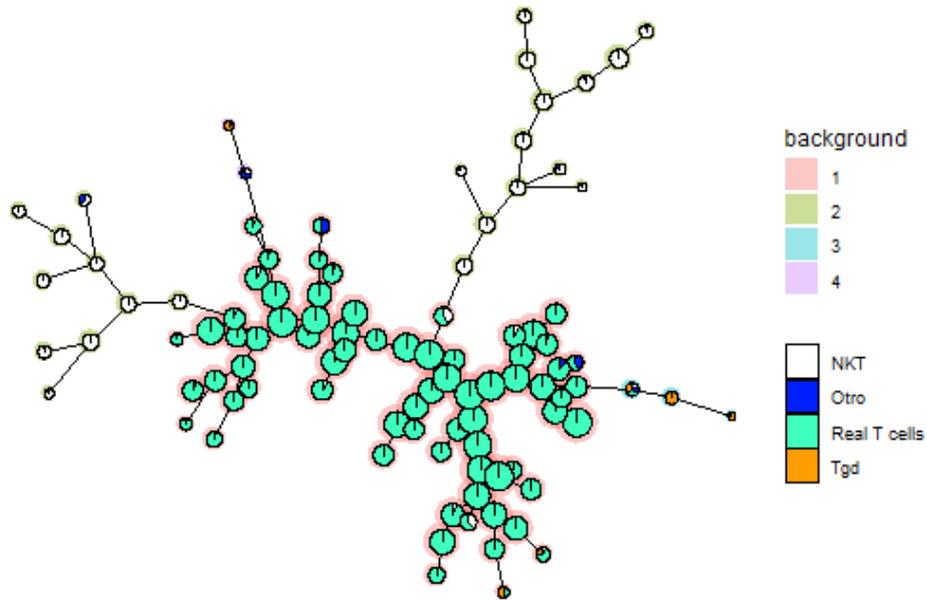


Figura 10: *Star Pie* del nivel *T cells* incluyendo todos los marcadores esenciales para la caracterización celular. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*.

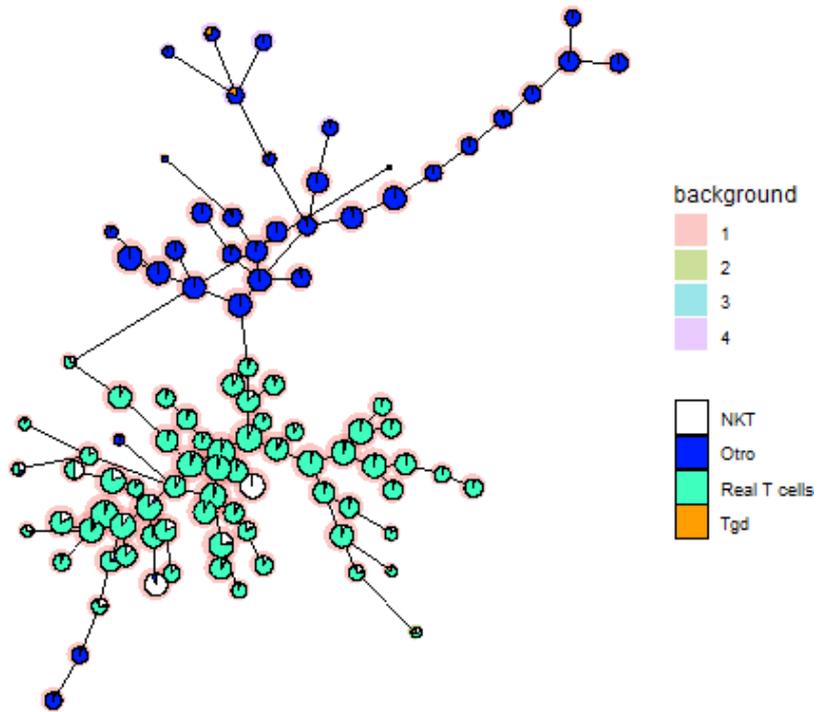


Figura 11: *Star Pie* del nivel *T cells* incluyendo todos los marcadores. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*.

## 4. Discusión

Los resultados obtenidos permiten reflexionar sobre el papel que tienen los algoritmos de *clustering* en el análisis automático de datos de citometría de flujo, especialmente en contexto de alta dimensionalidad y clasificación celular compleja. A continuación se discuten los principales hallazgos estructurados según los aspectos evaluados.

### 4.1. Comparativa entre FlowSOM y flowMeans

La comparación entre FlowSOM y flowMeans revela un comportamiento consistente dentro de cada conjunto de datos. En los conjuntos de menor complejidad (Normal Donor y West Nile Virus), flowMeans ofrece un rendimiento competitivo. Sin embargo, en el conjunto del IBGM, más representativo de un contexto real por su elevada dimensionalidad y número de poblaciones, FlowSOM alcanza mejores resultados con una clara ventaja computacional. En este sentido, flowMeans requiere tiempos de ejecución muy elevados, lo que limita su aplicabilidad práctica. Por el contrario, FlowSOM mantiene tiempos muy reducidos incluso en los escenarios más exigentes. Además, la versión automática de flowMeans obtiene resultados similares a su versión óptima, lo que indica que el método interno de estimación del número de *clusters* es razonablemente adecuado.

Los resultados obtenidos para el conjunto de datos del IBGM son notablemente inferiores a los presentados por Weber y Robinson [14]. Sin embargo, esta diferencia puede contextualizarse teniendo en cuenta varios factores. En primer lugar, el conjunto IBGM contiene en torno a 1.5 millones de eventos, mientras que los conjuntos evaluados en el FlowCAP-I challenge [13] y en el estudio de Weber y Robinson [14] se sitúan entre 100 000 y 500 000 eventos. En segundo lugar, el conjunto IBGM está completamente anotado (100 %), mientras que los anteriores lo están sólo parcialmente (alrededor del 40–70 %). Finalmente, también es importante destacar el número y nivel de detalle de las poblaciones: el conjunto IBGM contiene 51 poblaciones celulares definidas con gran precisión, frente a las 10–30 poblaciones de los conjuntos anteriores. Estas diferencias justifican la mayor complejidad del conjunto IBGM y explican, en parte, el menor rendimiento relativo de los algoritmos en este escenario.

Por otro lado, el análisis por población revela que las poblaciones más numerosas tienden a clasificarse con mayor precisión, lo cual tiene un efecto directo sobre el F-score global, especialmente cuando se emplean estrategias de asignación que no ponderan por tamaño.

## 4.2. Impacto del método de asignación

Uno de los aspectos metodológicos clave del presente estudio es la comparación entre dos estrategias de asignación de poblaciones a *clusters*: el criterio del F-score máximo y el algoritmo del húngaro.

Los resultados muestran de forma consistente que el criterio del máximo ofrece un rendimiento superior al del algoritmo del húngaro en todos los conjuntos y algoritmos evaluados. Esta diferencia también había sido señalada previamente en estudios como los de Weber y Robinson [14], y se refuerza aquí con una evaluación sistemática sobre múltiples métricas.

El algoritmo del húngaro maximiza la suma total del F-score, pero impone una correspondencia uno a uno entre *clusters* y poblaciones, lo cual puede conducir a asignaciones artificiales o biológicamente incoherentes. En particular, puede emparejar una población a un *cluster* donde apenas hay eventos de ese tipo si el F-score agregado es mayor. Por el contrario, el criterio del máximo permite que varias poblaciones se asignen a un mismo *cluster*, lo que es más adecuado en caso de poblaciones similares o parcialmente solapadas, como ocurre, por ejemplo, con subtipos de *NKT* que se diferencian únicamente en el nivel de expresión del marcador CD8.

Además, el F-score calculado con el algoritmo del húngaro se basa en una media no ponderada, otorgando el mismo peso a poblaciones grandes y pequeñas. Esto penaliza la calidad global cuando hay muchas poblaciones minoritarias difíciles de identificar. El criterio del máximo, al no requerir esa correspondencia uno a uno, logra valores de F-score más estables y, en general, más cercanos a una asignación realista.

Esta diferencia también se refleja en otras métricas como la *accuracy*, que es sistemáticamente menor con el algoritmo del húngaro debido a estas asignaciones forzadas. Por tanto, este trabajo refuerza la idea de que la elección del método de asignación tiene un impacto notable en la evaluación y debe ser cuidadosamente considerada al comparar algoritmos de *clustering*.

## 4.3. Variabilidad por la semilla aleatoria en FlowSOM

La ejecución de FlowSOM con distintas semillas aleatorias introduce cierta variabilidad en los resultados. No obstante, esta variabilidad es relativamente contenida y consistente con lo reportado por Weber y Robinson [14]. En particular, se observa que el criterio del máximo es algo más sensible a los cambios de semilla que el algoritmo del húngaro, probablemente porque el primero depende más directamente de la composición concreta de los *clusters* en términos de frecuencia.

En cualquier caso, los intervalos de confianza obtenidos mediante *bootstrap* muestran que la variabilidad inter semillas no altera las conclusiones principales del análisis. Este resultado aporta robustez al uso de FlowSOM en escenarios reales.

Este análisis de estabilidad se ha realizado únicamente con FlowSOM por varias razones. En primer lugar, dado que FlowSOM es considerablemente más eficiente computacionalmente, resulta factible ejecutar múltiples réplicas, algo inviable con flowMeans en conjuntos de alta dimensionalidad como el del IBGM. En segundo lugar, flowMeans solo presenta aleatoriedad al elegir los centroides iniciales, problema que solventa al seleccionar un número de *clusters* iniciales superior al número de poblaciones esperadas, por lo que no existe un efecto semilla comparable con el de FlowSOM.

#### **4.4. Clasificación jerárquica con FlowSOM**

En este trabajo, el análisis jerárquico se ha aplicado exclusivamente con FlowSOM debido a su bajo coste computacional y a la flexibilidad que ofrece su estructura de mapas auto-organizados, que permite extraer subconjuntos celulares de forma intuitiva.

La estrategia jerárquica basada en la aplicación secuencial de FlowSOM a subconjuntos celulares ha demostrado ser eficaz, especialmente cuando se utilizan únicamente los marcadores esenciales en cada nivel. Esta selección conduce a una mejora clara del rendimiento del algoritmo.

Los resultados obtenidos en el nivel *General* son notablemente superiores a los alcanzados con la versión óptima de FlowSOM. Esto pone de manifiesto el potencial de este enfoque para mejorar la identificación celular, especialmente cuando la clasificación se hace de forma escalonada, empezando por distinguir grandes grupos y profundizando posteriormente en subpoblaciones más específicas.

Una de las claves de esta mejora radica en el uso exclusivo de los marcadores esenciales para cada nivel jerárquico, lo cual permite una caracterización más precisa de las poblaciones celulares implicadas. De hecho, al restringir el análisis a dichos marcadores, se consiguen resultados que reproducen con gran fidelidad los obtenidos mediante el *manual gating*. En contraste, cuando se utilizan todos los marcadores, la calidad de la asignación disminuye de forma notable: algunos tipos celulares, como *No B & No T cells*, aparecen dispersos entre varios nodos y el *meta-clustering* agrupa poblaciones muy diferentes, como *B cells* y *T cells*, en un único *cluster*, perdiendo especificidad. En este contexto, aumentar el número de *meta-clusters* podría ser una estrategia útil para mejorar la diferenciación entre poblaciones.

Asimismo, los resultados reflejan que, a medida que se profundiza en la jerarquía, el rendimiento tiende a disminuir. Esta caída puede atribuirse al incremento del porcentaje de células no pertenecientes a las poblaciones diana dentro de cada subconjunto, lo que dificulta la correcta identificación de las clases minoritarias. A pesar de ello, incluso en los niveles más profundos, los modelos jerárquicos que emplean solo los marcadores esenciales siguen mostrando un F-score superior al de FlowSOM óptimo.

En conjunto, estos hallazgos confirman que la selección informada de marcadores resulta clave para mejorar el rendimiento del *clustering* automático. La incorporación de conocimiento experto sobre caracterización celular y/o la integración de estrategias de reducción de dimensionalidad podría mejorar los resultados obtenidos mediante enfoques jerárquicos.

## 5. Conclusiones y líneas futuras

Este trabajo ha comparado dos algoritmos de *clustering*, aplicados al análisis automático de datos de citometría de flujo: FlowSOM y flowMeans, evaluando su rendimiento, estabilidad y capacidad para detectar poblaciones celulares en distintos escenarios, incluyendo datos reales de alta dimensionalidad.

Los resultados obtenidos permiten extraer varias conclusiones relevantes:

- FlowSOM ofrece una clara ventaja computacional respecto a flowMeans, siendo especialmente notable en conjuntos complejos como el del IBGM. Esta eficiencia convierte a FlowSOM en una herramienta más adecuada para aplicaciones reales en citometría moderna.
- Respecto a rendimiento, flowMeans supera a FlowSOM en conjuntos más simples, como los del FlowCAP-I challenge. Sin embargo, FlowSOM ofrece mejores resultados en datos más realistas y desafiantes, como los del IBGM, especialmente cuando se incorpora conocimiento experto a través del análisis jerárquico y la selección de marcadores relevantes.
- La comparación entre métodos de asignación revela que el criterio del F-score máximo produce resultados más coherentes desde el punto de vista biológico, que el algoritmo del húngaro, que impone una correspondencia uno a uno entre clases reales y *clusters*.
- Le extensión jerárquica del algoritmo FlowSOM demuestra que la clasificación mejora considerablemente cuando se emplean únicamente los marcadores relevantes para cada nivel. Este enfoque reduce la confusión entre poblaciones celulares similares y permite una asignación más precisa, incluso en niveles de gran detalle.

En conjunto, estos resultados refuerzan la idea de que una estrategia híbrida, que combine algoritmos eficientes con conocimiento experto, es fundamental para mejorar la clasificación automática en citometría de flujo.

### 5.1. Líneas de trabajo futuras

A partir de los resultados obtenidos y las limitaciones observadas al realizar este trabajo, se proponen a continuación posibles líneas de investigación futuras que permitirían profundizar y ampliar este trabajo:

- Aplicación del enfoque propuesto a nuevos conjuntos de datos, con poblaciones no estudiadas previamente o con anotaciones más parciales, para evaluar su robustez en esos escenarios.
- Evaluación de otros algoritmos de *clustering* jerárquico, incluyendo alternativas a Flow-SOM, que permitan comparar el impacto del modelo base en la estructura multinivel.
- Incorporación de técnicas automáticas de selección de marcadores o reducción de dimensionalidad dentro del flujo jerárquico.
- Estudio de la estabilidad de los resultados en otras condiciones de variabilidad, como diferentes preprocesamientos, tamaños del mapa autoorganizado o estrategias de *meta-clustering*.
- Desarrollo de herramientas visuales y métricas adaptadas para facilitar la interpretación biológica de los resultados jerárquicos y mejorar la comunicación entre perfiles computacionales y clínicos.

## Referencias

- [1] Howard M Shapiro. *Practical flow cytometry*. John Wiley & Sons, 2005.
- [2] Katherine M McKinnon. «Flow cytometry: an overview». En: *Current protocols in immunology* 120.1 (2018), págs. 5-1.
- [3] Josef Spidlen et al. «Data file standard for flow cytometry, version FCS 3.1». En: *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry* 77.1 (2010), págs. 97-100.
- [4] ISAC. *International Society for Advancement of Cytometry*. <https://isac-net.org/>.  
Accedido: 14/03/2025.
- [5] Mack J Fulwyler. «Electronic separation of biological cells by volume». En: *Science* 150.3698 (1965), págs. 910-911.
- [6] Mario Roederer et al. «8 color, 10-parameter flow cytometry to elucidate complex leukocyte heterogeneity». En: *Cytometry: The Journal of the International Society for Analytical Cytology* 29.4 (1997), págs. 328-339.
- [7] Patrick Autissier et al. «Evaluation of a 12-color flow cytometry panel to study lymphocyte, monocyte, and dendritic cell subsets in humans». En: *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry* 77.5 (2010), págs. 410-419.
- [8] Andrew J Konecny et al. «OMIP-102: 50-color phenotyping of the human immune system with in-depth assessment of T cells and dendritic cells». En: *Cytometry Part A* (2024).
- [9] Yvan Saeys, Sofie Van Gassen y Bart N Lambrecht. «Computational flow cytometry: helping to make sense of high-dimensional immunology data». En: *Nature Reviews Immunology* 16.7 (2016), págs. 449-462.
- [10] Peng Qiu et al. «Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE». En: *Nature biotechnology* 29.10 (2011), págs. 886-891.

- [11] Yu Qian et al. «Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data». En: *Cytometry Part B: Clinical Cytometry* 78.S1 (2010), S69-S82.
- [12] Sofie Van Gassen et al. «FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data». En: *Cytometry Part A* 87.7 (2015), págs. 636-645.
- [13] Nima Aghaeepour et al. «Critical assessment of automated flow cytometry data analysis techniques». En: *Nature methods* 10.3 (2013), págs. 228-238.
- [14] Lukas M Weber y Mark D Robinson. «Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data». En: *Cytometry Part A* 89.12 (2016), págs. 1084-1096.
- [15] Xiao Liu et al. «A comparison framework and guideline of clustering methods for mass cytometry data». En: *Genome biology* 20 (2019), págs. 1-18.
- [16] Nima Aghaeepour et al. «Rapid cell population identification in flow cytometry data». En: *Cytometry Part A* 79.1 (2011), págs. 6-13.
- [17] Lukas Weber. *FlowRepository - Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry (CyTOF) data (ID: FR-FCM-ZZPH)*. <http://flowrepository.org/id/FR-FCM-ZZPH>. Accedido: 06/03/2025, las especificaciones para la descarga de datos son muy concretas.
- [18] John P Nolan y Danilo Condello. «Spectral flow cytometry». En: *Current protocols in cytometry* 63.1 (2013), págs. 1-27.
- [19] Greg Finak et al. «Optimizing transformations for automated, high throughput analysis of flow cytometry data». En: *BMC bioinformatics* 11 (2010), págs. 1-13.
- [20] Andrea Cossarizza et al. «Guidelines for the use of flow cytometry and cell sorting in immunological studies». En: *European journal of immunology* 49.10 (2019), págs. 1457-1973.

- [21] Hannah den Braanker, Margot Bongenaar y Erik Lubberts. «How to prepare spectral flow cytometry datasets for high dimensional data analysis: a practical workflow». En: *Frontiers in immunology* 12 (2021), pág. 768113.
- [22] Gerjen H Tinnevelt et al. «High-throughput single cell data analysis—a tutorial». En: *Analytica Chimica Acta* 1185 (2021), pág. 338872.
- [23] Rita Folcarelli et al. «Transformation of multicolour flow cytometry data with OTflow prevents misleading multivariate analysis results and incorrect immunological conclusions». En: *Cytometry Part A* 101.1 (2022), págs. 72-85.
- [24] Annelies Emmaneel et al. «PeacoQC: Peak-based selection of high quality cytometry data». En: *Cytometry Part A* 101.4 (2022), págs. 325-338.
- [25] Anil K Jain. «Data clustering: 50 years beyond K-means». En: *Pattern recognition letters* 31.8 (2010), págs. 651-666.
- [26] Bioconductor Open Source Software for Bioinformatics. *Bioconductor Software Packages*. <https://bioconductor.org/packages/3.20/bioc/>. Accedido: 06/03/2025.
- [27] Teuvo Kohonen. «The self-organizing map». En: *Proceedings of the IEEE* 78.9 (1990), págs. 1464-1480.
- [28] Ron Wehrens y Lutgarde MC Buydens. «Self-and super-organizing maps in R: the Kohonen package». En: *Journal of Statistical Software* 21 (2007), págs. 1-19.
- [29] Roland Coghetto. «Chebyshev distance». En: *Formalized Mathematics* (2016).
- [30] V Kevin M Whitney. «Algorithm 422: minimal spanning tree [H]». En: *Communications of the ACM* 15.4 (1972), págs. 273-274.
- [31] M Wilkerson y P Waltman. «ConsensusClusterPlus: ConsensusClusterPlus». En: *R Package. version 1.22* (2013).
- [32] Tarn Duong et al. «Feature significance for multivariate kernel density estimation». En: *Computational Statistics & Data Analysis* 52.9 (2008), págs. 4225-4242.

- [33] Trupti M Kodinariya, Prashant R Makwana et al. «Review on determining number of Cluster in K-Means Clustering». En: *International Journal* 1.6 (2013), págs. 90-95.
- [34] Glenn W Milligan y Martha C Cooper. «An examination of procedures for determining the number of clusters in a data set». En: *Psychometrika* 50 (1985), págs. 159-179.
- [35] Robert L Thorndike. «Who belongs in the family?» En: *Psychometrika* 18.4 (1953), págs. 267-276.
- [36] Erich Schubert. «Stop using the elbow criterion for k-means and how to choose the number of clusters instead». En: *ACM SIGKDD Explorations Newsletter* 25.1 (2023), págs. 36-42.
- [37] Tadeusz Caliński y Jerzy Harabasz. «A dendrite method for cluster analysis». En: *Communications in Statistics-theory and Methods* 3.1 (1974), págs. 1-27.
- [38] Gideon Schwarz. «Estimating the dimension of a model». En: *The annals of statistics* (1978), págs. 461-464.
- [39] Ethan Zhang y Yi Zhang. «F-Measure». En: *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009, págs. 1147-1147. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9\_483. URL: [https://doi.org/10.1007/978-0-387-39940-9\\_483](https://doi.org/10.1007/978-0-387-39940-9_483).
- [40] Yutaka Sasaki et al. «The truth of the F-measure». En: *Teach tutor mater* 1.5 (2007), págs. 1-5.
- [41] Harold W Kuhn. «The Hungarian method for the assignment problem». En: *Naval research logistics quarterly* 2.1-2 (1955), págs. 83-97.
- [42] Steven A Hicks et al. «On evaluation metrics for medical applications of artificial intelligence». En: *Scientific reports* 12.1 (2022), pág. 5979.
- [43] Thomas J DiCiccio y Bradley Efron. «Bootstrap confidence intervals». En: *Statistical science* 11.3 (1996), págs. 189-228.

- [44] Henry Lam y Zhenyuan Liu. «Bootstrap in high dimension with low computation». En: *International Conference on Machine Learning*. PMLR. 2023, págs. 18419-18453.
- [45] NIH BioArt. *NIAID Visual & Medical Arts*. <https://bioart.niaid.nih.gov/discover?bioartCategory=500041>. Accedido: 29/05/2025.
- [46] B Ellis et al. *flowCore: flowCore: Basic structures for flow cytometry data*. R package version 2.16.0. 2024. DOI: 10.18129/B9.bioc.flowCore. URL: <https://bioconductor.org/packages/flowCore>.
- [47] Nima Aghaeepour. *flowMeans: Non-parametric Flow Cytometry Data Gating*. R package version 1.64.0. 2024. DOI: 10.18129/B9.bioc.flowMeans. URL: <https://bioconductor.org/packages/flowMeans>.
- [48] Sofie Van Gassen et al. «FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data». En: *Cytometry Part A* 87.7 (2015), págs. 636-645. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/cyto.a.22625>.
- [49] Kurt Hornik. «A CLUE for CLUster Ensembles». En: *Journal of Statistical Software* 14.12 (sep. de 2005). DOI: 10.18637/jss.v014.i12.

## Apéndice A: Clasificación jerárquica de las poblaciones celulares en los datos IBGM

A continuación de cada nivel del esquema se encuentra el porcentaje de células de ese tipo que hay. Al tratarse de una clasificación jerárquica no siempre se llega hasta el máximo nivel disponible.

### 1. T cells (70.61 %)

#### *a)* NKT (8.74 %)

- 1) CD2<sup>-</sup> (6.26 %)
- 2) CD2<sup>+</sup>CD8<sup>high</sup> (0.81 %)
- 3) CD2<sup>+</sup>CD8<sup>dim</sup> (0.37 %)
- 4) CD2<sup>+</sup>CD8<sup>+</sup> (1.09 %)

#### *b)* Tgd (1.00 %)

- 1) Trm (0.30 %)
- 2) Memory (0.22 %)
- 3) Naive (0.09 %)
- 4) nrTEMRA (0.08 %)
- 5) rTEMRA (0.04 %)

#### *c)* T cells (Real) (60.87 %)

##### 1) CD8<sup>+</sup> (14.62 %)

- a'* Trm (5.47 %)
- b'* Memory (2.34 %)
- c'* Naive (0.60 %)
- d'* nrTEMRA (1.26 %)
- e'* rTEMRA (4.80 %)

##### 2) CD4<sup>+</sup> (42.99 %)

- a'* Trm: Th2 (5.98 %), Th1 (14.36 %), Th17 (4.49 %), Th1/Th17-like (5.38 %)
- b'* Memory: Th2 (0.57 %), Th1 (7.86 %), Th17 (0.15 %), Th1/Th17-like (1.29 %)
- c'* Naive (1.68 %)
- d'* nrTEMRA (1.01 %)
- e'* rTEMRA (0.16 %)

- 3) DP (CD4<sup>+</sup>CD8<sup>+</sup>) (0.89 %)
- 2. B cells (25.25 %)
  - a) IgD<sup>+</sup> (3.71 %)
  - b) IgD<sup>-</sup> (20.49 %)
    - 1) Plasmablasts (13.29 %): IgA (7.41 %), IgG (1.38 %), IgM (0.23 %)
    - 2) Memory B cells (7.20 %): IgA (2.78 %), IgG (2.12 %), IgM (1.21 %)
- 3. No T & No B cells (4.14 %)
  - a) Basophils (0.06 %)
  - b) NK cells (DR<sup>-</sup>) (1.09 %)
    - 1) Early NK (0.57 %)
    - 2) Mature NK (0.04 %)
    - 3) Terminal NK (0.47 %)
  - c) ILC (0.35 %)
  - d) APC (2.63 %)
    - 1) Macrophages (0.20 %)
    - 2) Monocytes (1.17 %)
    - 3) cDC (1.20 %)



## Apéndice B: Resultados adicionales para la comparativa de rendimiento, flowMeans frente a FlowSOM

|                        | FlowMeans                  |                            | FlowSOM                    |
|------------------------|----------------------------|----------------------------|----------------------------|
|                        | Automático                 | Óptimo                     | Óptimo                     |
| <b>Normal Donor</b>    |                            |                            |                            |
| Máximo                 | 0.9404<br>[0.9398, 0.9410] | 0.9116<br>[0.9076, 0.9157] | 0.7735<br>[0.7710, 0.7761] |
| Húngaro                | 0.4719<br>[0.4703, 0.4736] | 0.4101<br>[0.4015, 0.4187] | 0.2421<br>[0.2297, 0.2546] |
| <b>West Nile Virus</b> |                            |                            |                            |
| Máximo                 | 0.9483<br>[0.9478, 0.9487] | 0.9524<br>[0.9519, 0.9529] | 0.7363<br>[0.7354, 0.7371] |
| Húngaro                | 0.7391<br>[0.7372, 0.7410] | 0.6886<br>[0.6875, 0.6897] | 0.6530<br>[0.6489, 0.6571] |
| <b>IBGM</b>            |                            |                            |                            |
| Máximo                 | 0.2823<br>[0.2804, 0.2841] | 0.3008<br>[0.2998, 0.3018] | 0.4228<br>[0.4205, 0.4251] |
| Húngaro                | 0.1680<br>[0.1671, 0.1690] | 0.2102<br>[0.2063, 0.2142] | 0.2810<br>[0.2796, 0.2824] |

Tabla B.1: *Precision* para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

|                        | <b>FlowMeans</b>           |                            | <b>FlowSOM</b>             |
|------------------------|----------------------------|----------------------------|----------------------------|
|                        | Automático                 | Óptimo                     | Óptimo                     |
| <b>Normal Donor</b>    |                            |                            |                            |
| Máximo                 | 0.9021<br>[0.9011, 0.9030] | 0.9206<br>[0.9158, 0.9254] | 0.9117<br>[0.9078, 0.9157] |
| Húngaro                | 0.5395<br>[0.5318, 0.5472] | 0.4384<br>[0.4315, 0.4453] | 0.2858<br>[0.2777, 0.2938] |
| <b>West Nile Virus</b> |                            |                            |                            |
| Máximo                 | 0.8496<br>[0.8488, 0.8505] | 0.8442<br>[0.8433, 0.8451] | 0.9390<br>[0.9385, 0.9395] |
| Húngaro                | 0.7000<br>[0.6979, 0.7022] | 0.6693<br>[0.6677, 0.6709] | 0.5866<br>[0.5847, 0.5885] |
| <b>IBGM</b>            |                            |                            |                            |
| Máximo                 | 0.3315<br>[0.3284, 0.3345] | 0.3282<br>[0.3276, 0.3288] | 0.3855<br>[0.3790, 0.3921] |
| Húngaro                | 0.1418<br>[0.1397, 0.1439] | 0.1712<br>[0.1670, 0.1753] | 0.2265<br>[0.2248, 0.2281] |

Tabla B.2: *Recall* para los distintos algoritmos. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

## Apéndice C: Resultados adicionales de la evaluación de la estabilidad

| F-score          |                  | Accuracy         |                  |
|------------------|------------------|------------------|------------------|
| Máximo           | Húngaro          | Máximo           | Húngaro          |
| 0.3257           | 0.2085           | 0.3462           | 0.2765           |
| [0.3248, 0.3265] | [0.2080, 0.2091] | [0.3444, 0.3480] | [0.2755, 0.2775] |
| 0.3413           | 0.2145           | 0.3971           | 0.2983           |
| [0.3409, 0.3417] | [0.2128, 0.2162] | [0.3965, 0.3977] | [0.2976, 0.2990] |
| 0.3562           | 0.2141           | 0.4063           | 0.3080           |
| [0.3552, 0.3572] | [0.2127, 0.2154] | [0.4052, 0.4074] | [0.3039, 0.3121] |
| 0.3588           | 0.2234           | 0.4467           | 0.3148           |
| [0.3579, 0.3598] | [0.2222, 0.2247] | [0.4443, 0.4491] | [0.3124, 0.3172] |
| 0.3421           | 0.2232           | 0.3986           | 0.3035           |
| [0.3415, 0.3427] | [0.2221, 0.2244] | [0.3940, 0.4032] | [0.3031, 0.3039] |
| 0.3348           | 0.2125           | 0.3624           | 0.2881           |
| [0.3341, 0.3354] | [0.2116, 0.2133] | [0.3608, 0.3640] | [0.2876, 0.2886] |
| 0.3280           | 0.2098           | 0.3896           | 0.2859           |
| [0.3274, 0.3287] | [0.2083, 0.2112] | [0.3877, 0.3915] | [0.2852, 0.2866] |
| 0.3370           | 0.2155           | 0.3954           | 0.2862           |
| [0.3363, 0.3377] | [0.2143, 0.2168] | [0.3944, 0.3964] | [0.2821, 0.2903] |
| 0.3318           | 0.2093           | 0.3987           | 0.2862           |
| [0.3310, 0.3325] | [0.2085, 0.2101] | [0.3893, 0.4081] | [0.2848, 0.2876] |
| 0.3278           | 0.2123           | 0.3677           | 0.2726           |
| [0.3266, 0.3289] | [0.2116, 0.2129] | [0.3655, 0.3699] | [0.2716, 0.2736] |

Tabla C.1: F-score y *accuracy* para la ejecución de FlowSOM óptimo, cada fila corresponde a cada una de las diez semillas. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

| <b>Precision</b> |                  | <b>Recall</b>    |                  |
|------------------|------------------|------------------|------------------|
| Máximo           | Húngaro          | Máximo           | Húngaro          |
| 0.4088           | 0.2728           | 0.3462           | 0.2209           |
| [0.4081, 0.4095] | [0.2700, 0.2757] | [0.3444, 0.3481] | [0.2190, 0.2229] |
| 0.3937           | 0.2740           | 0.3971           | 0.2316           |
| [0.3928, 0.3947] | [0.2721, 0.2758] | [0.3965, 0.3977] | [0.2296, 0.2336] |
| 0.4238           | 0.2751           | 0.4063           | 0.2203           |
| [0.4226, 0.4250] | [0.2663, 0.2838] | [0.4052, 0.4074] | [0.2170, 0.2236] |
| 0.4117           | 0.2959           | 0.4467           | 0.2353           |
| [0.4099, 0.4136] | [0.2925, 0.2993] | [0.4443, 0.4491] | [0.2316, 0.2390] |
| 0.3942           | 0.2883           | 0.3986           | 0.2317           |
| [0.3929, 0.3956] | [0.2869, 0.2896] | [0.3940, 0.4031] | [0.2300, 0.2333] |
| 0.4062           | 0.2699           | 0.3624           | 0.2305           |
| [0.4055, 0.4070] | [0.2690, 0.2707] | [0.3608, 0.3640] | [0.2290, 0.2320] |
| 0.3905           | 0.2827           | 0.3896           | 0.2210           |
| [0.3898, 0.3912] | [0.2815, 0.2839] | [0.3878, 0.3915] | [0.2188, 0.2231] |
| 0.4183           | 0.2794           | 0.3954           | 0.2297           |
| [0.4175, 0.4190] | [0.2741, 0.2846] | [0.3944, 0.3964] | [0.2266, 0.2328] |
| 0.3869           | 0.2766           | 0.3987           | 0.2171           |
| [0.3836, 0.3901] | [0.2752, 0.2780] | [0.3894, 0.4081] | [0.2126, 0.2217] |
| 0.4355           | 0.2813           | 0.3677           | 0.2354           |
| [0.4342, 0.4367] | [0.2792, 0.2833] | [0.3655, 0.3699] | [0.2311, 0.2397] |

Tabla C.2: *Precision y recall* para la ejecución de FlowSOM óptimo, cada fila corresponde a cada una de las diez semillas. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

## Apéndice D: Resultados adicionales del análisis jerárquico

| Grupo                        | Precision                  |                            | Recall                     |                            |
|------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
|                              | Máximo                     | Húngaro                    | Máximo                     | Húngaro                    |
| <b>General</b>               | 0.9872<br>[0.9871, 0.9874] | 0.9856<br>[0.9853, 0.9859] | 0.9872<br>[0.9871, 0.9874] | 0.9470<br>[0.9460, 0.9479] |
| <b>T cells</b>               | 0.9678<br>[0.9672, 0.9684] | 0.8782<br>[0.8756, 0.8807] | 0.9765<br>[0.9761, 0.9768] | 0.8006<br>[0.7973, 0.8038] |
| NKT                          | 0.7319<br>[0.7281, 0.7356] | 0.4610<br>[0.4459, 0.4761] | 0.5307<br>[0.5273, 0.5340] | 0.4686<br>[0.4441, 0.4473] |
| Tgd                          | 0.3472<br>[0.3041, 0.3904] | 0.3592<br>[0.3371, 0.3813] | 0.5774<br>[0.5139, 0.6410] | 0.4853<br>[0.4588, 0.5119] |
| Real T cells                 | 0.9206<br>[0.9199, 0.9213] | 0.6556<br>[0.6535, 0.6575] | 0.9179<br>[0.9169, 0.9188] | 0.5414<br>[0.5401, 0.5427] |
| CD4 <sup>+</sup>             | 0.6008<br>[0.5830, 0.6186] | 0.5600<br>[0.5559, 0.5641] | 0.7044<br>[0.6935, 0.7152] | 0.3563<br>[0.3530, 0.3596] |
| CD8 <sup>+</sup>             | 0.3246<br>[0.3231, 0.3261] | 0.4023<br>[0.3858, 0.4187] | 0.8439<br>[0.8421, 0.8457] | 0.3290<br>[0.3273, 0.3308] |
| <b>B cells</b>               | 0.9096<br>[0.9081, 0.9111] | 0.7630<br>[0.7575, 0.7685] | 0.9301<br>[0.9289, 0.9313] | 0.6795<br>[0.6766, 0.6823] |
| IgD <sup>-</sup>             | 0.9319<br>[0.9311, 0.9328] | 0.9344<br>[0.9335, 0.9353] | 0.9683<br>[0.9679, 0.9688] | 0.9551<br>[0.9543, 0.9558] |
| Plasmablasts                 | 0.6783<br>[0.6603, 0.6962] | 0.5777<br>[0.5641, 0.5913] | 0.8174<br>[0.8044, 0.8303] | 0.6466<br>[0.6388, 0.6543] |
| Memory B cells               | 0.7306<br>[0.7264, 0.7389] | 0.5190<br>[0.5166, 0.5213] | 0.7509<br>[0.7485, 0.7533] | 0.5456<br>[0.5439, 5474]   |
| <b>No T &amp; No B cells</b> | 0.4685<br>[0.4649, 0.4720] | 0.3441<br>[0.3215, 0.3666] | 0.9003<br>[0.8968, 0.9039] | 0.4126<br>[0.3842, 0.4410] |
| NK cells                     | 0.2770<br>[0.2732, 0.2808] | 0.2852<br>[0.2729, 0.2974] | 0.7954<br>[0.7900, 0.8008] | 0.5049<br>[0.4872, 0.5227] |
| APC                          | 0.4264<br>[0.4225, 0.4302] | 0.5176<br>[0.4379, 0.5974] | 0.9309<br>[0.9285, 0.9333] | 0.2887<br>[0.2848, 0.2925] |

Tabla D.1: *Precision* y *recall* en el análisis jerárquico con FlowSOM empleando solo los marcadores relevantes por nivel. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

| Grupo                        | Precision                   |                            | Recall                     |                            |
|------------------------------|-----------------------------|----------------------------|----------------------------|----------------------------|
|                              | Máximo                      | Húngaro                    | Máximo                     | Húngaro                    |
| <b>General</b>               | 0.5761<br>[0.5754, 0.5768]  | 0.5073<br>[0.5026, 0.5120] | 0.9513<br>[0.9510, 0.9517] | 0.3519<br>[0.3514, 0.3524] |
| <b>T cells</b>               | 0.5610<br>[0.5599, 0.5622]  | 0.3115<br>[0.3079, 0.3150] | 0.9911<br>[0.9908, 0.9913] | 0.4776<br>[0.4730, 0.4821] |
| NKT                          | 0.0655<br>[0.0648, 0.0661]  | 0.0230<br>[0.0223, 0.0238] | 0.9737<br>[0.9725, 0.9750] | 0.2147<br>[0.2128, 0.2166] |
| Tgd                          | 0.4707<br>[0.4442, 0.4973]  | 0.2222<br>[0.1719, 0.2726] | 0.5821<br>[0.5599, 0.6042] | 0.3684<br>[0.2502, 0.4867] |
| Real T cells                 | 0.4745<br>[0.34734, 0.4756] | 0.1678<br>[0.1671, 0.1685] | 0.9419<br>[0.9415, 0.9424] | 0.2596<br>[0.2590, 0.2601] |
| CD4 <sup>+</sup>             | 0.2487<br>[0.2475, 0.2499]  | 0.1710<br>[0.1694, 0.1727] | 0.7465<br>[0.7458, 0.7472] | 0.1721<br>[0.1712, 0.1729] |
| CD8 <sup>+</sup>             | 0.2032<br>[0.2022, 0.2042]  | 0.0643<br>[0.0629, 0.0657] | 0.8692<br>[0.8366, 0.9018] | 0.1566<br>[0.1560, 0.1571] |
| <b>B cells</b>               | 0.1941<br>[0.1932, 0.1949]  | 0.2531<br>[0.2464, 0.2598] | 0.9605<br>[0.9598, 0.9611] | 0.3577<br>[0.3562, 0.3593] |
| IgD <sup>-</sup>             | 0.2699<br>[0.2689, 0.2710]  | 0.3258<br>[0.3247, 0.3269] | 0.7877<br>[0.7863, 0.7891] | 0.7488<br>[0.7475, 0.7502] |
| Plasmablasts                 | 0.3917<br>[0.3899, 0.3934]  | 0.1370<br>[0.1359, 0.1381] | 0.9068<br>[0.9052, 0.9083] | 0.3024<br>[0.3011, 0.3038] |
| Memory B cells               | 0.2111<br>[0.2090, 0.2132]  | 0.1180<br>[0.0917, 0.1443] | 0.9971<br>[0.9965, 0.9976] | 0.2510<br>[0.2498, 0.2521] |
| <b>No T &amp; No B cells</b> | 0.4156<br>[0.4043, 0.4270]  | 0.2277<br>[0.2077, 0.2476] | 0.8134<br>[0.7990, 0.8278] | 0.6214<br>[0.3629, 0.8798] |
| NK cells                     | -                           |                            |                            |                            |
| APC                          | 0.3455<br>[0.3278, 0.3632]  | 0.1876<br>[0.1712, 0.2040] | 0.7172<br>[0.6872, 0.7472] | 0.3439<br>[0.3010, 0.3869] |

Tabla D.2: *Precision* y *recall* en el análisis jerárquico con FlowSOM empleando todos los marcadores disponibles. Se muestran los valores estimados junto con sus intervalos de confianza al 95 %, calculados mediante el método *cheap bootstrap* propuesto por Lam y Liu [44].

## Apéndice E: Gráficos *Star Pie* del análisis jerárquico completo

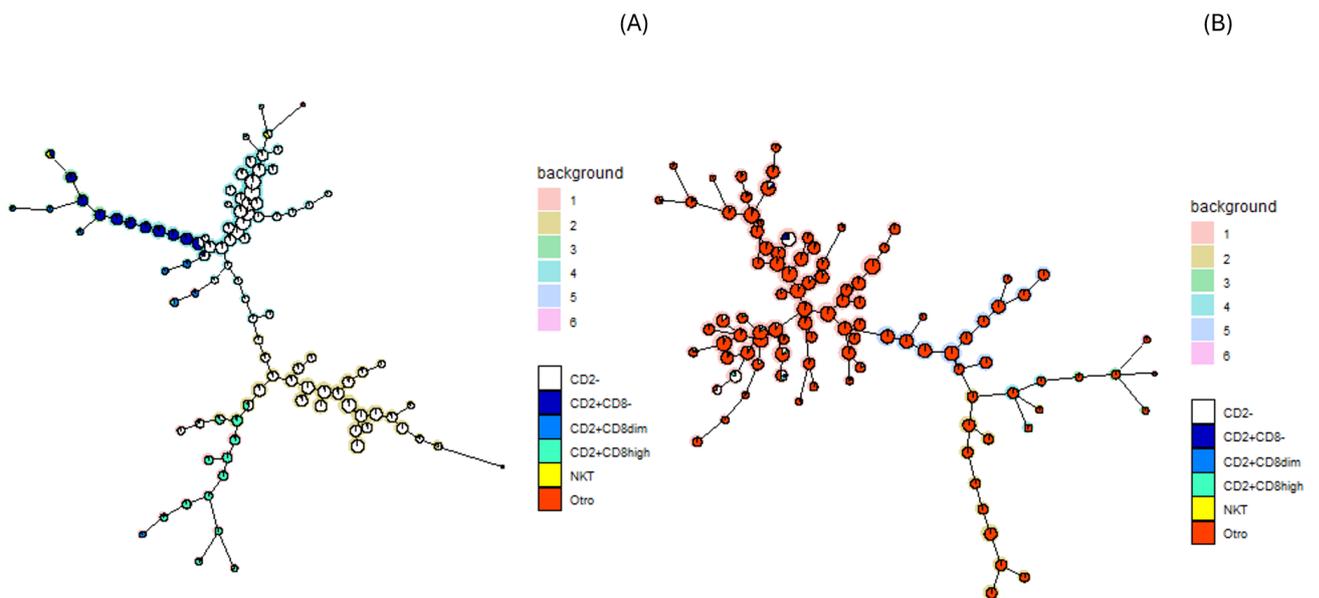


Figura E.1: *Star Pie plot* del nivel *NKT*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

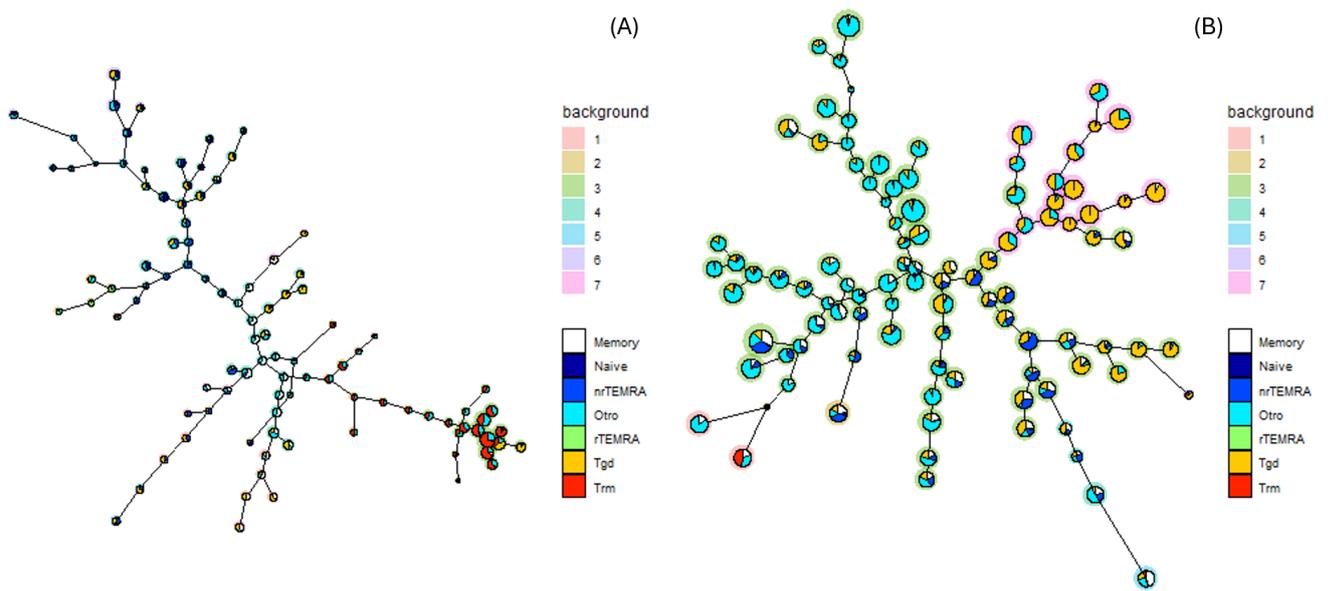


Figura E.2: *Star Pie plot* del nivel *Tgd*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

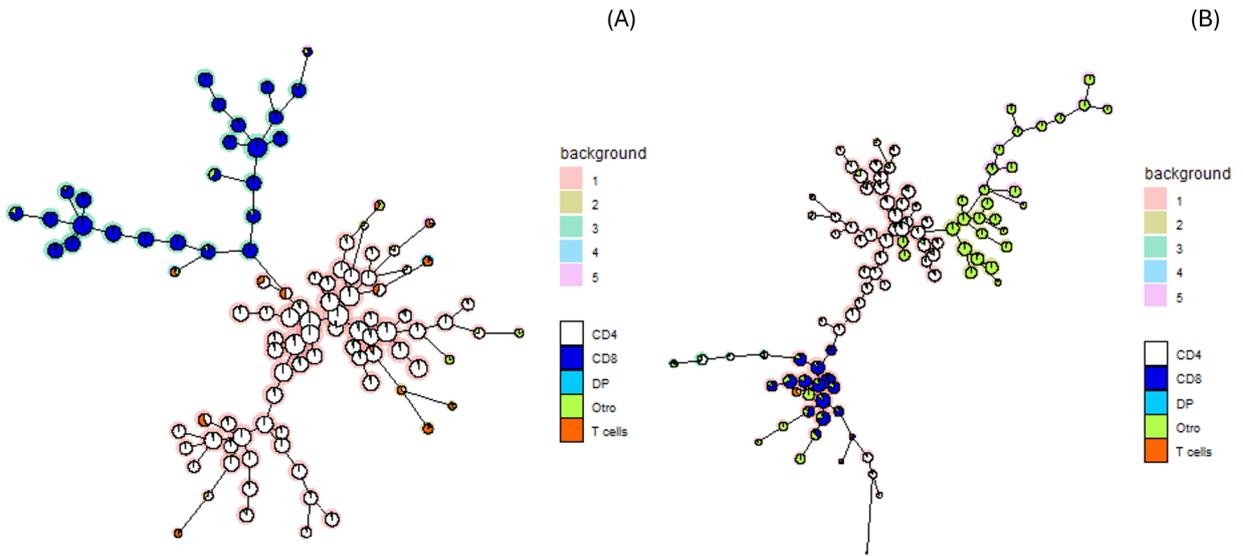


Figura E.3: *Star Pie plot* del nivel *Real T cells*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

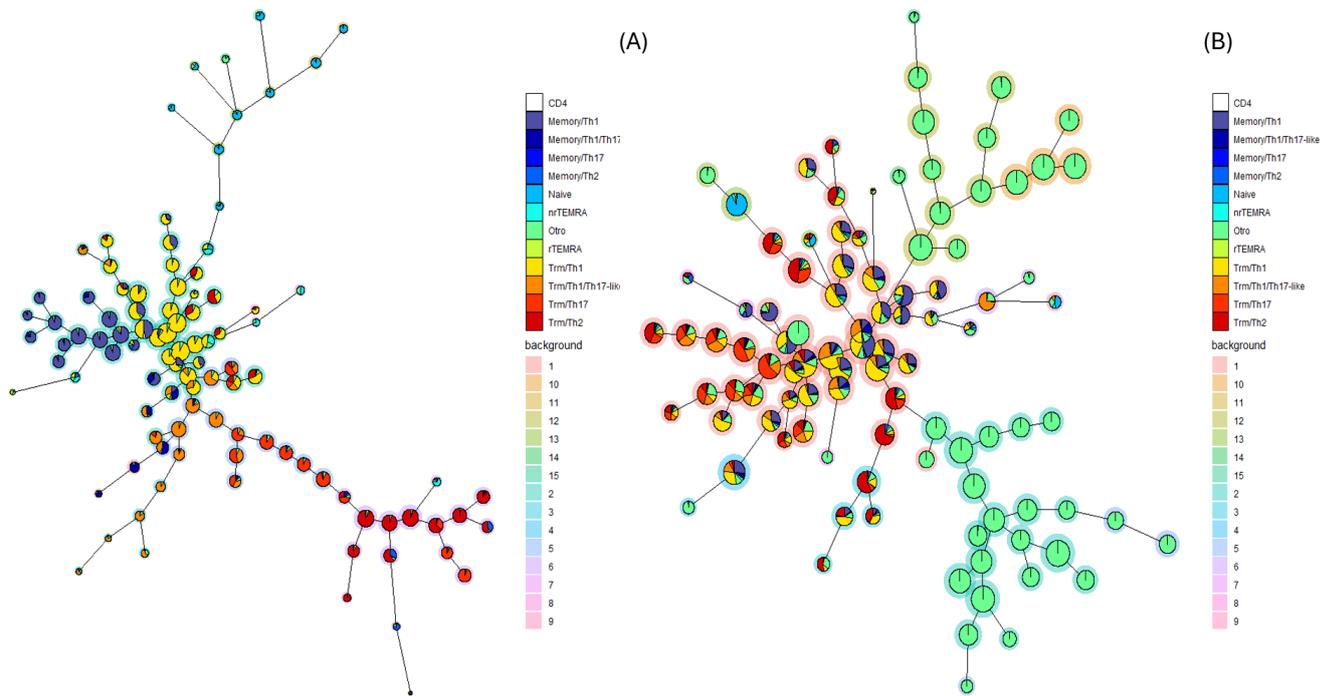


Figura E.4: *Star Pie plot* del nivel  $CD4^+$ . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

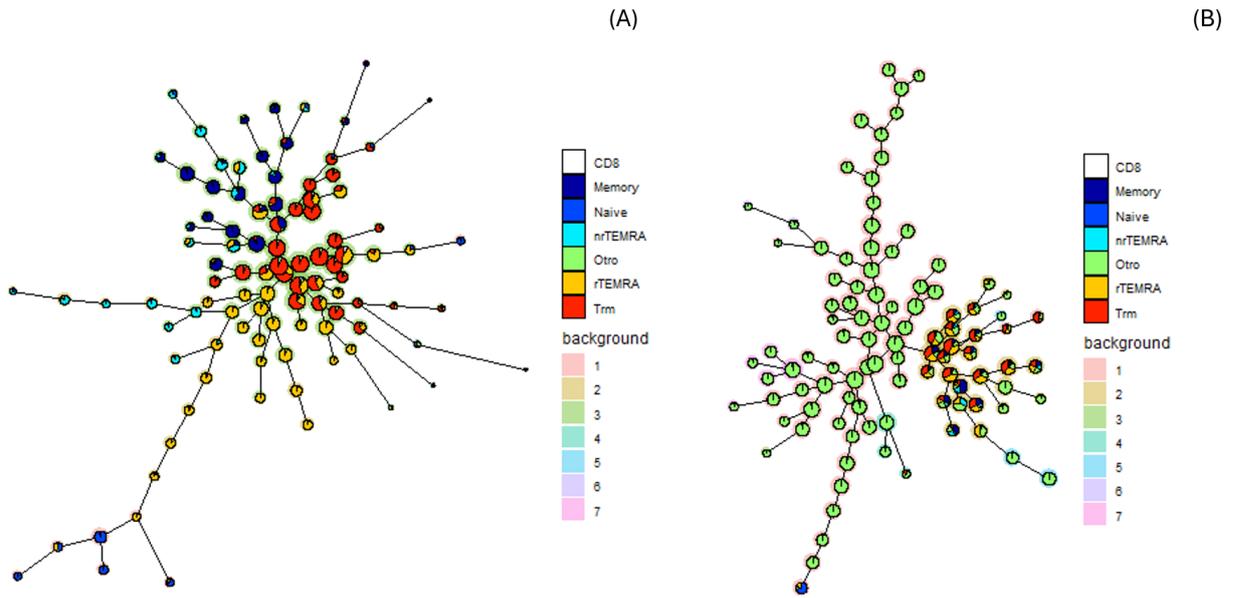


Figura E.5: *Star Pie plot* del nivel  $CD8^+$ . Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

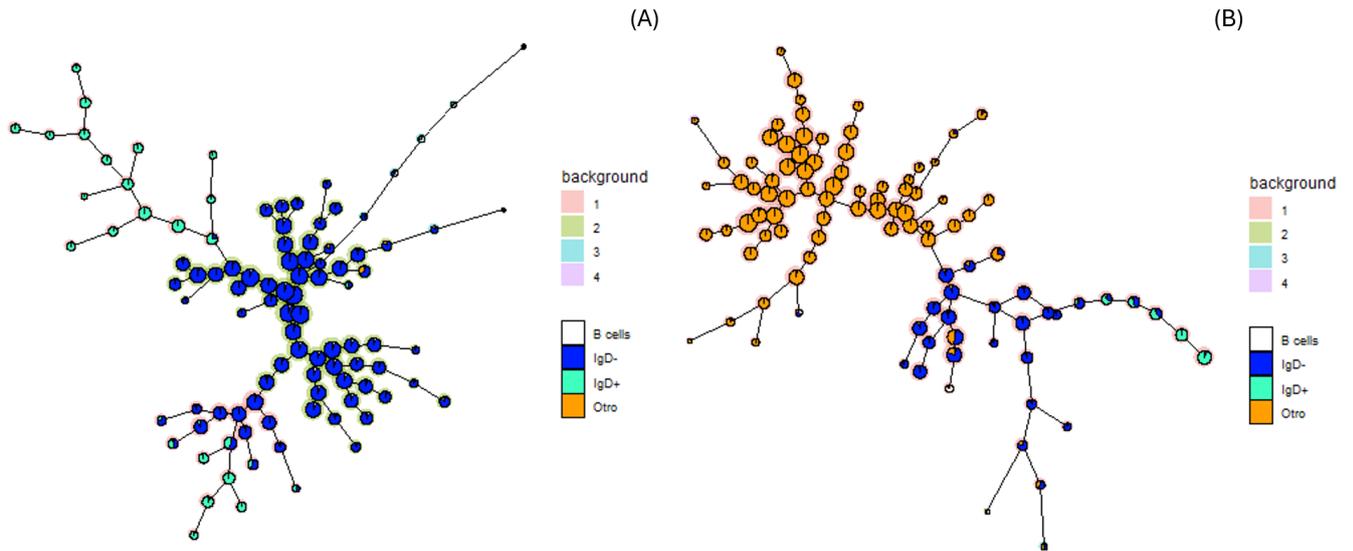


Figura E.6: *Star Pie plot* del nivel *B cells*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

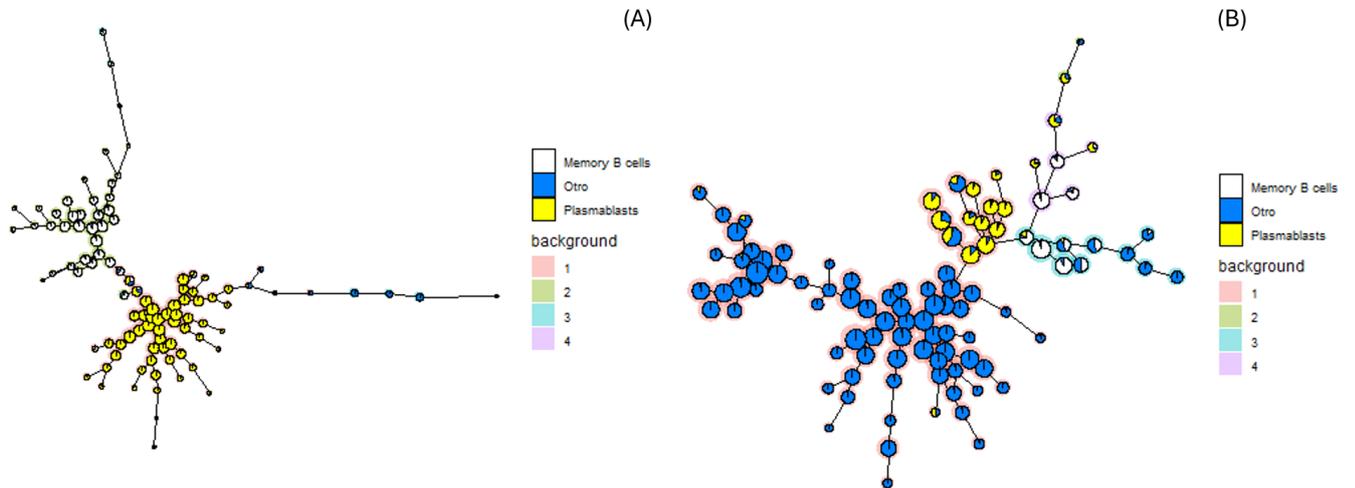


Figura E.7: *Star Pie plot* del nivel *IgD<sup>-</sup>*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

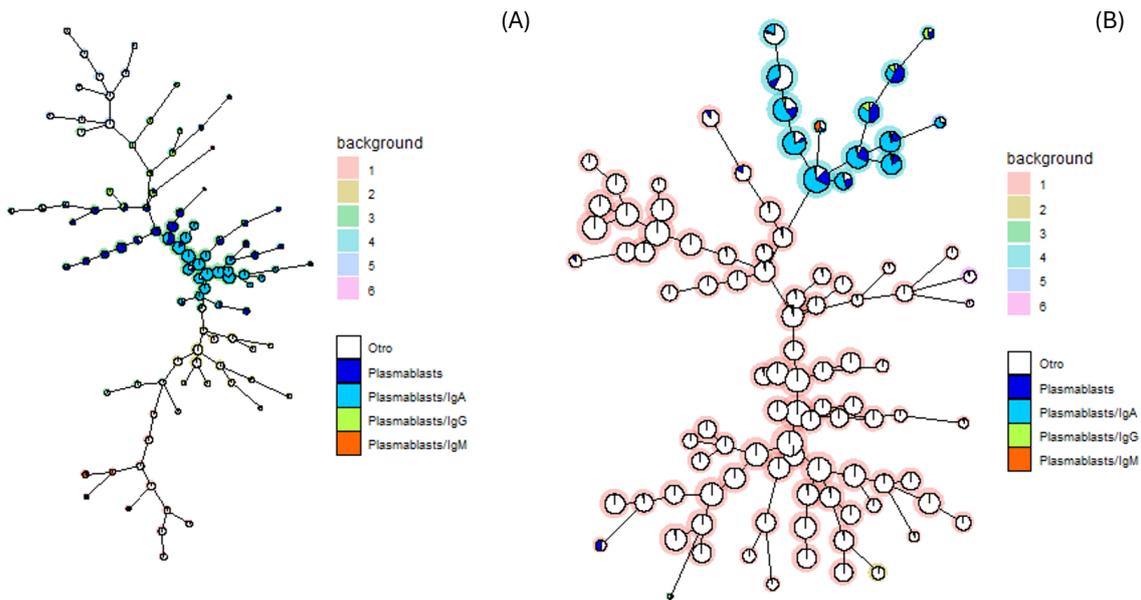


Figura E.8: *Star Pie plot* del nivel *Plasmablasts*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

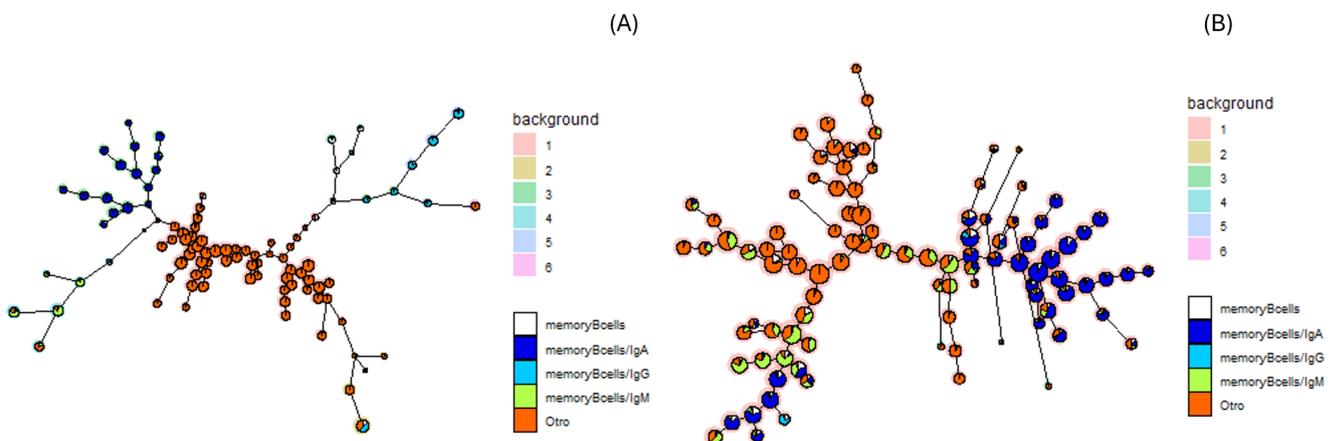


Figura E.9: *Star Pie plot* del nivel *Memory B cells*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

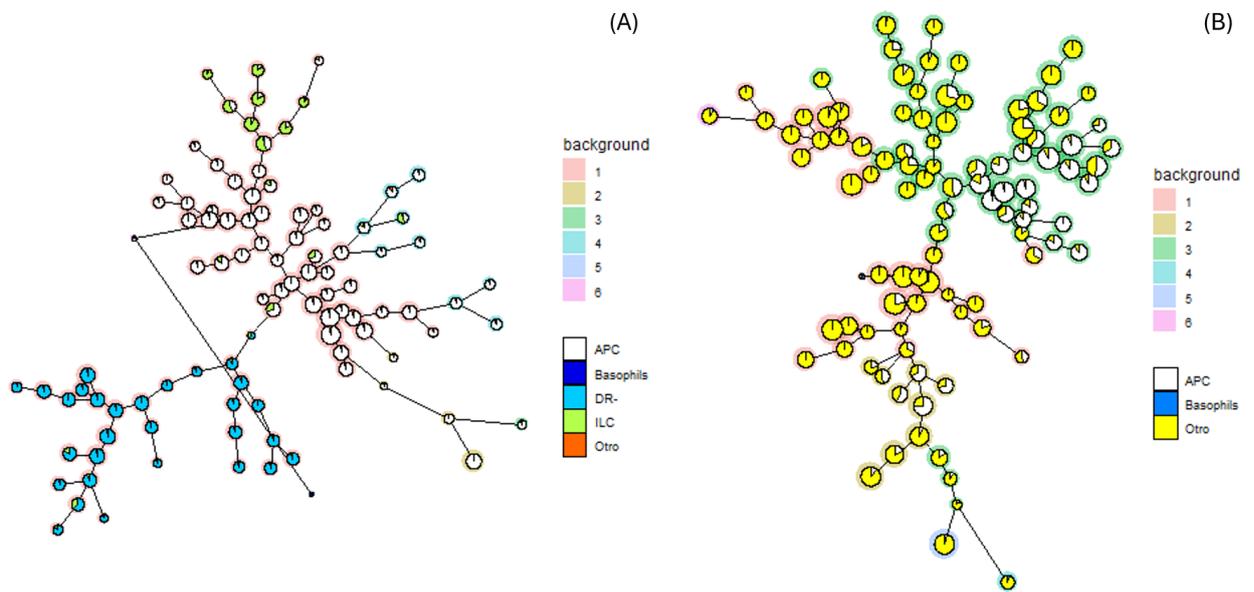


Figura E.10: *Star Pie plot* del nivel *No T & No B cells*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.

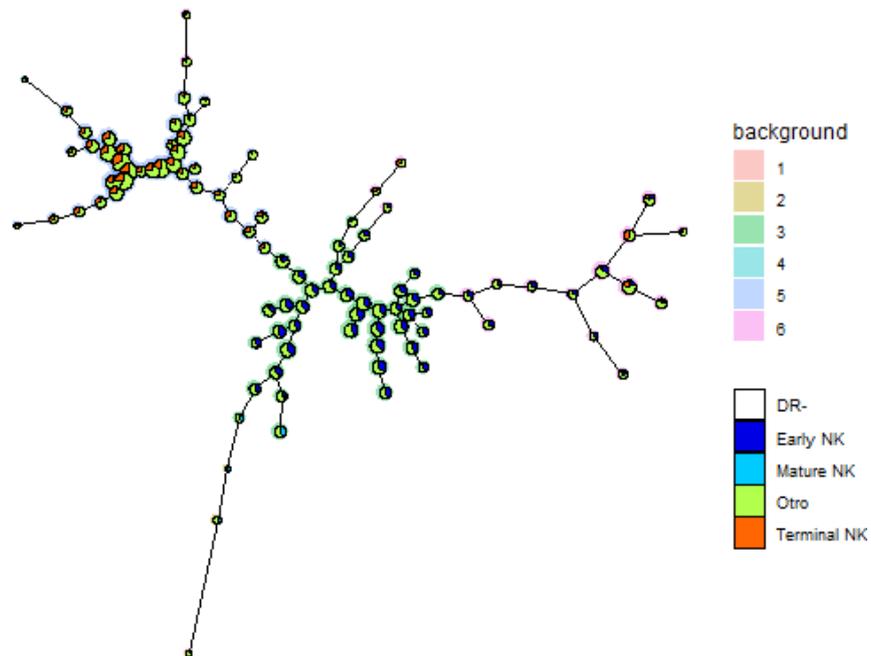


Figura E.11: *Star Pie plot* del nivel *NK cells*, para el modelo que incluye los marcadores esenciales. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*.

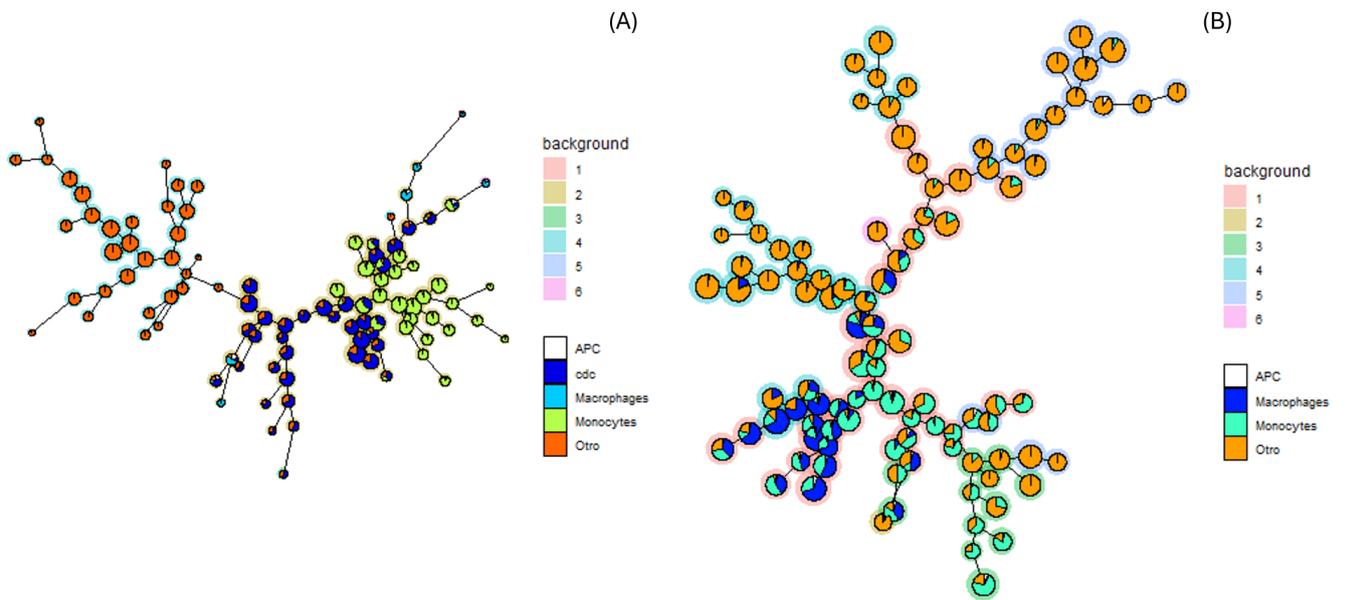


Figura E.12: *Star Pie plot* del nivel *APC*. Se muestra el resultado obtenido tras aplicar FlowSOM, representando el color del contorno del nodo la agrupación superior obtenida tras el paso de *meta-clustering*. (A) modelo que incluye los marcadores esenciales. (B) modelos con todos los marcadores disponibles.