

# APRENDIZAJE AUTOMÁTICO APLICADO AL ANÁLISIS DE DATOS PROTEÓMICOS PARA DEFINIR NUEVOS ENDOTIPOS DE ARTROSIS DE RODILLA

Patricia Quaranta<sup>1</sup>, Patricia Fernandez-Puente<sup>1,2</sup>, Diego Fernandez-Edreira<sup>3</sup>, Jose Linares-Blanco<sup>3</sup>, Lucia Lourido<sup>1</sup>, Cristina Ruiz Romero<sup>1,4</sup>, Natividad Oreiro<sup>5</sup>, M. Teresa Silva Diaz<sup>5</sup>, Mercedes Freire Gonzalez<sup>5</sup>, Alfonso Soto Gonzalez<sup>6</sup>, Marisa Crespo<sup>7</sup>, Jose Luis Diaz Diaz<sup>8</sup>, Enrique Miguez Rey<sup>9</sup>, Alvaro Mena de Cea<sup>9</sup>, Guillermo Vazquez<sup>10</sup>, Francisco J Blanco-Garcia<sup>1,5</sup>, Valentina Calamia<sup>1</sup>.

<sup>1</sup>Instituto de Investigación Biomédica de A Coruña (INIBIC); <sup>2</sup>Centro Interdisciplinar de Química y Biología (CICA). Universidad de A Coruña (UDC); <sup>3</sup>Machine Learning in Life Sciences Lab. Dept. of Computer Science and Information Technologies, Universidad de A Coruña (CITIC); <sup>4</sup>CIBER-BBN; <sup>5</sup>Reumatología, Hospital Universitario A Coruña (HUAC); <sup>6</sup>Endocrinología, Hospital Universitario A Coruña (HUAC); <sup>7</sup>Cardiología, Hospital Universitario A Coruña (HUAC); <sup>8</sup>Medicina Interna, Hospital Universitario A Coruña (HUAC); <sup>9</sup>Enfermedades Infecciosas, Hospital Universitario A Coruña (HUAC); <sup>10</sup>Avances en Telemedicina e Informática Sanitaria, Hospital Universitario A Coruña

Email: patricia.quaranta.diaz@sergas.es; valentina.calamia@sergas.es

GR Grupo de Investigación de Reumatología



## INTRODUCCIÓN

Varios estudios in vitro e in vivo (en modelos animales) han demostrado que los diversos fenotipos clínicos de los pacientes con artrosis (OA) podrían estar asociados a alteraciones específicas del metabolismo de los condrocitos, el único tipo celular presente en el cartílago articular. Sin embargo, los fenotipos resultantes aún no se han caracterizado claramente y se necesitan más investigaciones al respecto. Teniendo en cuenta que una mejor caracterización de estos fenotipos clínicos tiene el potencial de realizar contribuciones impactantes al desarrollo de intervenciones más eficaces en la OA, planteamos la hipótesis de que las comorbilidades en la artrosis podrían determinar la aparición de fenotipos clínicos distintos asociados con firmas moleculares específicas (endotipos de OA).

## OBJETIVOS

El objetivo principal de este trabajo es identificar distintos endotipos de OAr asociados con perfiles proteicos específicos mediante el análisis de datos proteómicos y el uso de aprendizaje automático.

## MÉTODOS

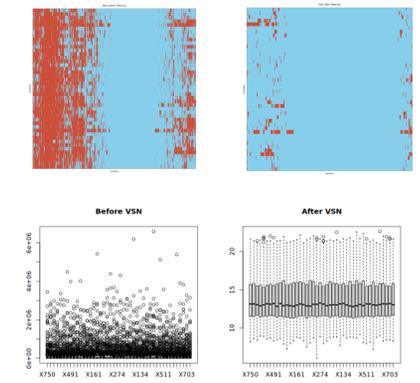
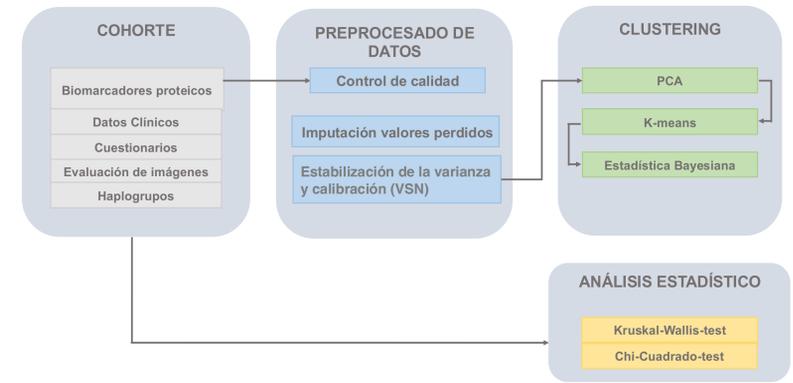


Fig.1: Pre-procesamiento datos crudos.

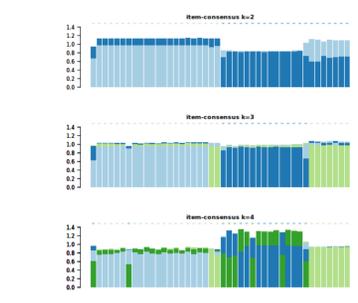


Fig.2: Análisis consenso y de cluster.

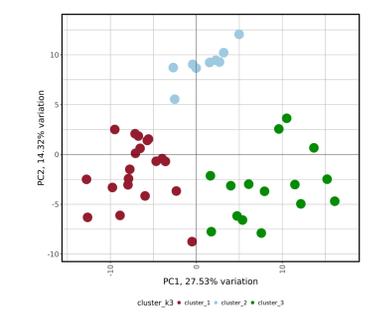


Fig.3: Visualización de los 3 clusters.

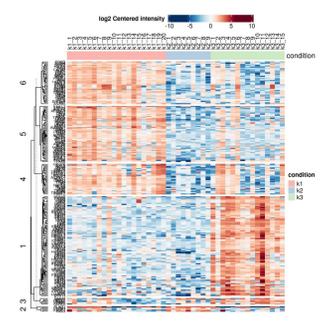


Fig.4: Heatmap de los 3 clusters.

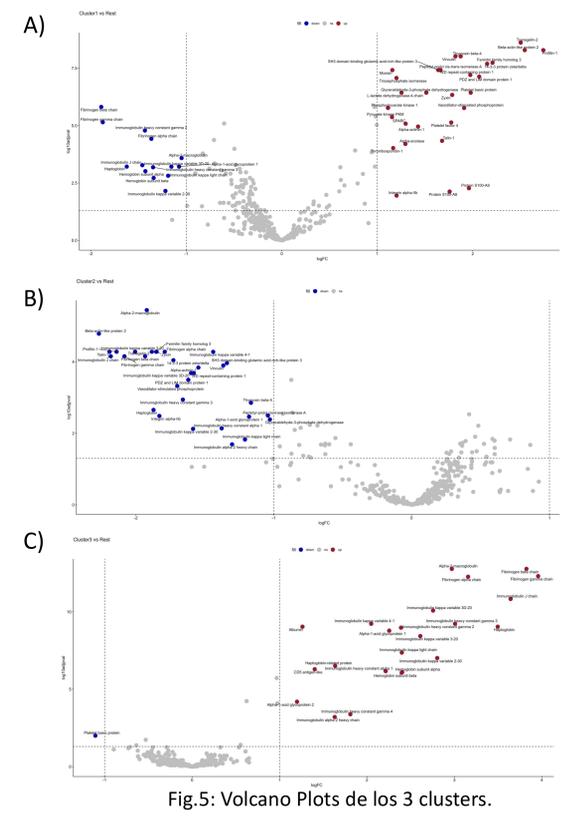
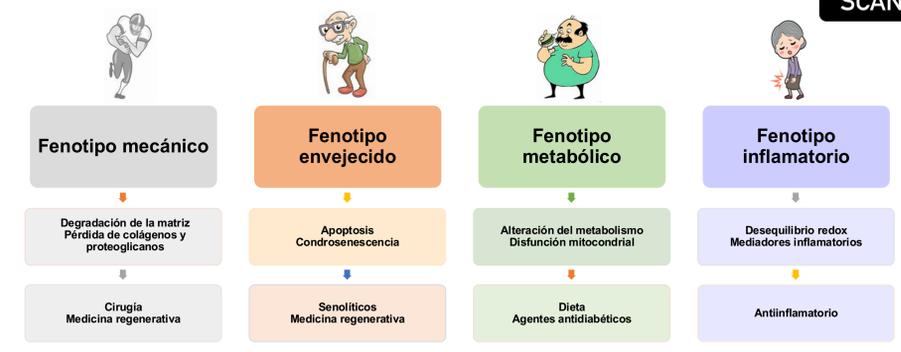


Fig.5: Volcano Plots de los 3 clusters.



## RESULTADOS

Tras el análisis por espectrometría de masas al azar de las muestras de plasma, se identificaron 1.075 proteínas. Los datos crudos se filtraron en función del número de péptidos únicos detectados y el número de los valores perdidos; tras este filtro se mantuvieron 395 proteínas para la imputación y la normalización (Fig.1). El análisis de consenso de clusters reveló que cuando el número de clusters aumenta, la estabilidad de los clusters disminuye a medida que los clusters se dividen en grupos más pequeños. Como se puede observar en la figura 2, los cluster 2 y 3 son los más estables (valores de consenso =1), mientras que el cluster 4 presentaba más variabilidad (valores de consenso <1). Tras el análisis del número de clusters, se fijó en 3 el número óptimo de cluster quedando así distribuidas las muestras entre los 3 grupos: Cluster 1 (n=20, rojo), Cluster 2 (n=9, azul) y Cluster 3 (n=15, verde) (Fig.3). El análisis de expresión diferencial se realizó utilizando la estadística Empírica de Bayes, ajustando por sexo, con corrección de pruebas múltiples aplicada mediante el método Benjamini-Hochberg (Fig.4,5). El cluster 1 se caracterizó por 43 proteínas alteradas (Fig.5A), de las cuales 29 están incrementadas y asociadas principalmente con la glucólisis aeróbica y la activación, señalización y agregación plaquetaria. El cluster 2 presentó 32 proteínas, todas disminuidas (Fig.5C). Por último, el cluster 3 se caracterizó por 24 proteínas, de las cuales 23 están aumentadas y relacionadas con procesos inmunitarios (Fig.5C). Cuatro biomarcadores que definen el endotipo 1 se validaron mediante inmunoensayos. Estos biomarcadores fueron TSP1 (p = 0,0026, AUC = 0,77), SPARC (p = 0,0047, AUC = 0,78), PF4 (p = 0,0015, AUC = 0,79) y CXCL7 (p = 0,0001, AUC = 0,878). Para las 4 proteínas se obtuvieron concentraciones más altas en las muestras clasificadas como endotipo 1 respecto a las muestras de los endotipos 2 y 3, lo cual confirmó los resultados obtenidos por espectrometría de masas.

## CONCLUSIÓN

La combinación de enfoques computacionales con la proteómica representa un importante avance en la comprensión de la heterogeneidad de la OA y en la mejora del desarrollo de terapias dirigidas. Su relevancia traslacional radica en la optimización del diseño de ensayos clínicos y la mejora de la eficacia del tratamiento para los pacientes con OA de rodilla.