



Análisis de bioestadística aplicada a la expresión génica en Los Simpson

Estadística y R para Ciencias de la Salud

Ainhoa Artetxe, Alba Xiaohe Elias, Alejandra Martín, Alejandro Pascual y Ana Sofía Santos Tedim



Repository de
GitHub

INTRODUCCIÓN

La obesidad es un fenotipo complejo y multifactorial. Diversos genes implicados en la regulación del apetito, la señalización neuroendocrina y el balance energético han sido asociados al desarrollo de obesidad, aunque su expresión puede variar entre individuos. El objetivo de este estudio fue explorar la variabilidad de la expresión génica asociada a la obesidad mediante técnicas multivariantes.

METODOLOGÍA

Se analizó un conjunto de datos correspondiente a 59 individuos, de la familia Simpson, con información clínica y de expresión de 37 genes relacionados con obesidad. Tras la limpieza de los datos y la eliminación de valores perdidos y genes con varianza nula, se evaluó la normalidad de la expresión génica mediante el test de Shapiro-Wilk. Posteriormente, se realizó un análisis de componentes principales (PCA) con los datos centrados y escalados. Se llevaron a cabo visualizaciones de genes e individuos, análisis de clustering, mapas de calor de correlación Spearman y modelos de regresión logística para explorar asociaciones entre las componentes principales y la obesidad.

RESULTADOS

Figura 1. PCA de genes y clustering de genes (k = 3). Biplot del análisis de PCA que representa los genes proyectados sobre PC1 y PC2 con clustering de genes en tres grupos.

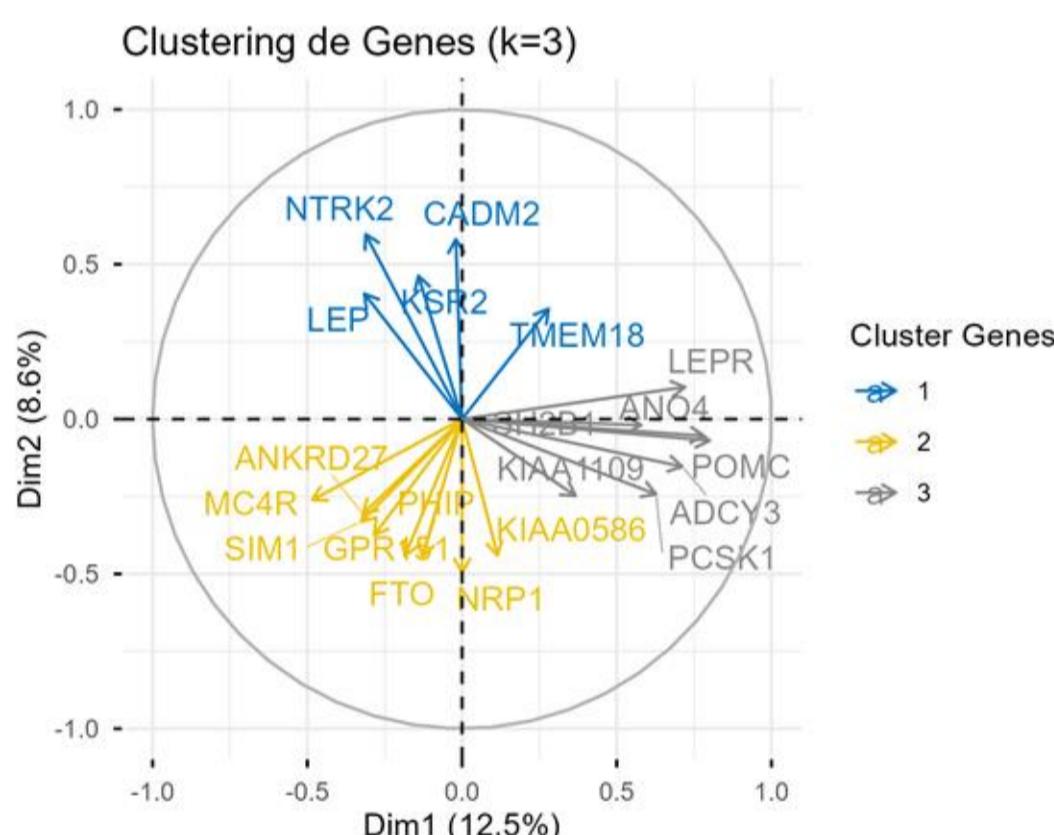


Figura 3. PCA de individuos según categoría de IMC. Proyección de los individuos en el espacio definido por PC1 y PC2, coloreados según la categoría de IMC.

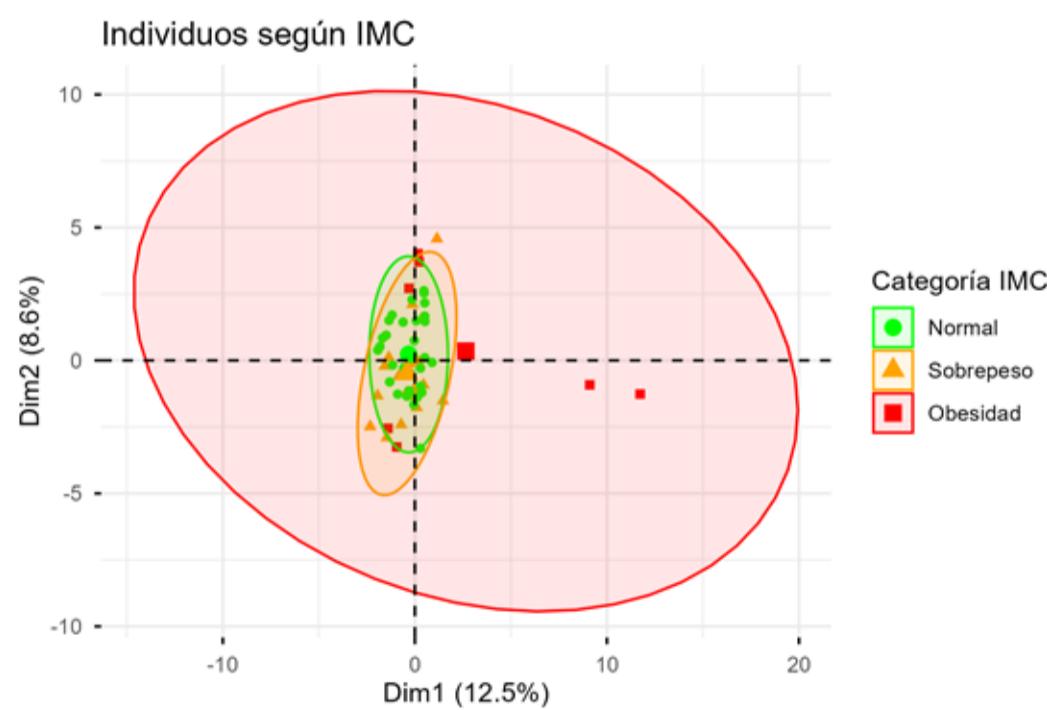


Figura 2. Heatmap de correlación genes - PCA. Muestra las correlaciones de Spearman entre la expresión de los genes analizados y las 6 primeras componentes principales.

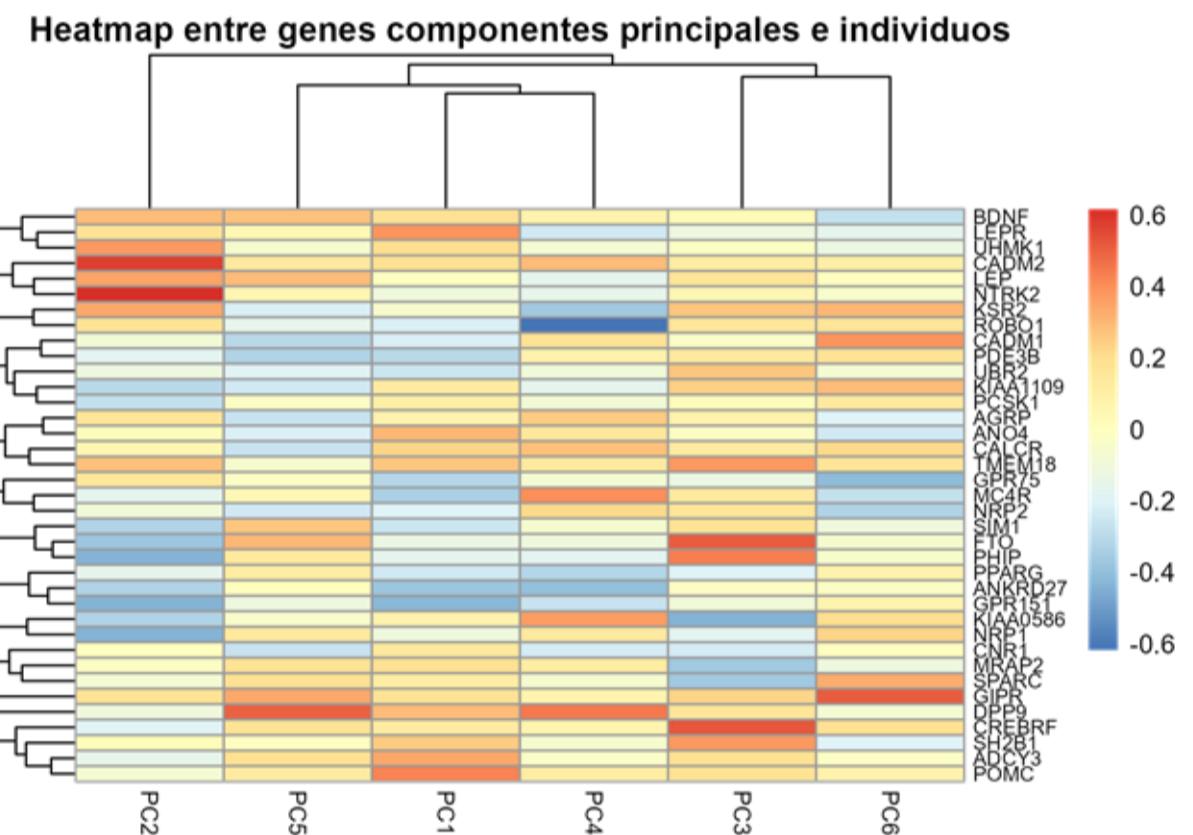
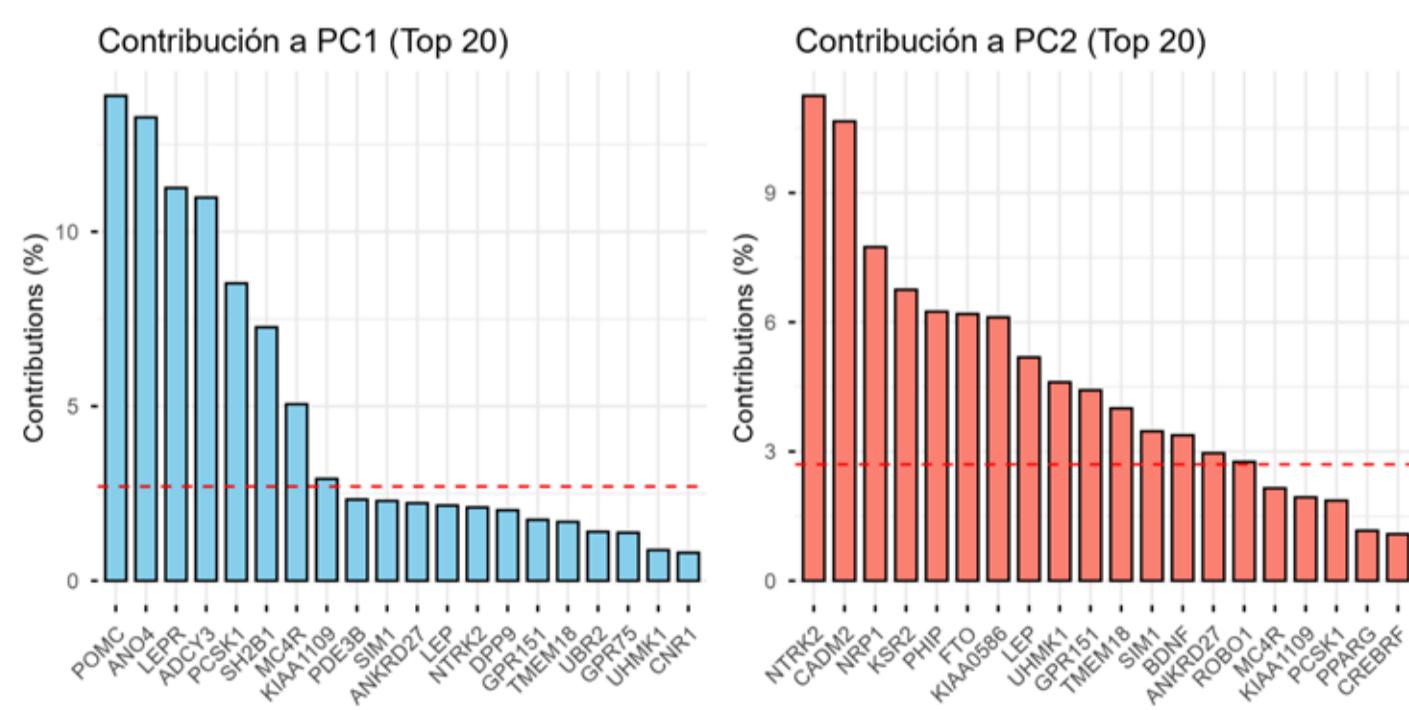


Figura 4. Contribución relativa de los 20 primeros genes a PC1 y PC2. La línea discontinua indica la contribución media esperada.



DISCUSIÓN

Los análisis multivariantes muestran que la variabilidad de la expresión génica asociada a la obesidad se distribuye entre múltiples componentes, sin un eje dominante. Las componentes principales PC1 y PC2, están influenciadas por genes clave implicados en la regulación del apetito y la señalización neuroendocrina, como por ejemplo *POMC*, *LEPR*, *ADCY3*, *NTRK2*, *CADM2* y *NRP1*.

El clustering de genes y las correlaciones con las componentes principales sugieren la existencia de módulos transcriptómicos relacionados con procesos biológicos compartidos, como el control de la ingesta de alimentos y el metabolismo energético.

La proyección de los individuos según el IMC muestra un alto solapamiento entre categorías, lo que indica una marcada heterogeneidad molecular y la ausencia de un perfil transcriptómico único asociado a la obesidad.

CONCLUSIONES

El análisis multivariante de la expresión génica permitió identificar patrones globales asociados a procesos biológicos clave en la obesidad, como la regulación del apetito y el metabolismo energético.

La elevada heterogeneidad observada entre individuos refuerza la concepción de la obesidad como un fenotipo biológicamente complejo.

Este estudio pone de manifiesto la utilidad del PCA y de los enfoques exploratorios para el análisis de datos transcriptómicos en el contexto de enfermedades complejas.