

Los datos para el análisis

El objetivo del estudio presentado es pretende hallar genes diferencialmente expresados entre varios tipos de tumores de cancer de mama: apocrinos (APO), basales (BAS) y luminales (LUMI).

La clasificación se basa en la resistencia de los tumores a los receptores de estrógenos y de andrógenos.

- Los tumores clasificados como "APO" son negativos para los receptores de estrógenos (ER-) y positivos para los receptores de andrógenos (AR+).
- los clasificados como "LUMI" son ER+ y AR+ y
- los clasificados como "BAS" son ER- y AR-.

Los datos originales así como información acerca del estudio pueden obtenerse del Gene Expression Omnibus (GEO) serie GSE1561 en el siguiente enlace

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1561>

El artículo que describe el estudio es: <http://www.ncbi.nlm.nih.gov/pubmed/15897907>

Mi propuesta es que para el análisis no utiliceis todos los datos sino que extraigais una muestra aleatoria de 15 arrays (5 de cada grupo) lo que asegurará que no tengais problemas de memoria.

La asignación de cada muestra a un grupo experimental podeis obtenerla de este enlace:

<http://www.ncbi.nlm.nih.gov/geo/gds/profileGraph.cgi?gds=1329>