



## Actividad 2.2. Modelos lineales para el análisis de expresión diferencial

### Índice

Introducción.....	1
Casos de estudio .....	2
1. Efecto del tratamiento con SHAM1 en la expresión de células T-ALL.....	2
2. Comparación entre tres tipos de cáncer de mama .....	3
3. Comparación entre pacientes sanos y pacientes con endometriosis en un estudio de RNA-seq. ....	3
4. Efecto del programa génico termogénico durante la diferenciación de adipocitos .....	4
5. Bases moleculares de la desregulación de citocinas asociada a la edad en macrófagos estimulados con LPS.....	4

### Introducción

A continuación, se describen situaciones experimentales que se han estudiado utilizando tecnologías ómicas, principalmente microarrays, pero también RNA-seq. El objetivo de la actividad es que *investiguéis* alguno de los diseños experimentales habituales y su representación utilizando un modelo lineal adecuado. En cada caso, debéis identificar el **diseño experimental y el modelo lineal asociado**, es decir, **la matriz de diseño que lo describe y la matriz de contrastes utilizada para responder a las preguntas planteadas por los investigadores**. Después, **escoged únicamente uno de los casos propuestos** y responded las cuestiones que se plantean. La entrega de la actividad debe ser un documento breve con vuestros análisis, resultados y conclusiones.

# Casos de estudio

## 1. Efecto del tratamiento con SHAM1 en la expresión de células T-ALL

El conjunto de datos para el ejercicio está disponible en la entrada de la Serie GSE18198 en GEO. Consiste en el análisis de perfiles de expresión de líneas celulares humanas T-ALL tratadas con DMSO o SHAM1.

Las proteínas NOTCH regulan vías de señalización involucradas en la diferenciación, proliferación y muerte celular. La sobreactivación de la señalización Notch se ha observado en numerosos cánceres, y se ha estudiado extensamente en el contexto de la leucemia linfoblástica aguda de células T (T-ALL), donde más del 50% de los pacientes presentan mutaciones en NOTCH1. Identificar pequeñas moléculas moduladoras de la actividad de estas proteínas sería importante para entender el papel de las proteínas NOTCH en procesos biológicos malignos y normales. En este estudio, los investigadores estaban interesados en medir los cambios globales en la expresión génica tras el tratamiento de las líneas celulares T-ALL humanas HPB-ALL y KOPT-K1 con solo vehículo (DMSO) o SHAM1, un péptido modulador alfa-helicoidal derivado de la proteína coactivadora MAML1.

Por lo tanto, diseñaron un experimento que consiste en cultivos triplicados de células KOPT-K1 o HPB-ALL tratadas con DMSO (solamente) o SHAM1 (20  $\mu$ M) durante 24 horas. Se extrajo ARN total y se hibridó en microarrays de Affymetrix human U133 plus 2.0 (tres arrays por tratamiento y línea celular, un total de 12 arrays).

1. Describe el diseño experimental. Identifica los factores experimentales y sus niveles.
2. Escribe la matriz de diseño asociada con este diseño experimental.
3. Construye la matriz de contraste que se puede utilizar para responder a las siguientes preguntas:
  - a. Comparar el efecto de SHAM1 en la línea celular KOPT-K1: KOPT-K1 tratado con SHAM1 frente a KOPT-K1 tratado con DMSO (el vehículo)
  - b. Comparar el efecto de SHAM1 en la línea celular HPB-ALL: HPB-ALL tratado con SHAM1 frente a HPB-ALL tratado con DMSO.
  - c. La interacción: las diferencias entre los dos efectos anteriores.

## 2. Comparación entre tres tipos de cáncer de mama

Este estudio de caso se basa en un [artículo](#) cuyos datos están disponibles en GEO (Serie GSE1561).

Los investigadores investigaron tres tipos de tumores de cáncer de mama: apocrino (APO), basal (BAS) y luminal (LUMI). La clasificación se basa en la resistencia de los tumores a los receptores de estrógenos y andrógenos.

- Los tumores clasificados como “APO” son negativos para el receptor de estrógeno (ER-) y positivos para el receptor de andrógeno (AR+).
- Aquellos clasificados como “LUMI” son ER+ y AR+ y
- Aquellos clasificados como “BAS” son ER- y AR-.

La asignación de cada muestra a un grupo experimental se puede obtener en este [enlace](#).

Obviamente, este es un estudio observacional, pero su análisis también se puede realizar utilizando un enfoque de modelo lineal.

1. Identifica los factores experimentales y sus niveles.
2. Escribe la matriz de diseño asociada con este diseño de estudio.
3. Construye la matriz de contraste necesaria para comparar cada tipo de tumor con los otros dos, es decir:
  - a. “APO” vs “LUMI”
  - b. “APO” vs “BAS”
  - c. “LUMI” vs “BAS”

## 3. Comparación entre pacientes sanos y pacientes con endometriosis en un estudio de RNA-seq

Disponemos de 40 muestras de RNA-seq generadas para un estudio.

Tenemos 6 pacientes con endometriosis y 5 donantes sanos. En cada grupo tenemos 4 poblaciones celulares diferentes: A, B, C, D. Nos gustaría comparar cada población entre los individuos sanos y los afectados. Además, nos gustaría comparar las poblaciones C y D en los sanos (por un lado) y en los afectados (por otro).

1. Identifica los factores experimentales y sus niveles.
2. Escribe la matriz de diseño asociada con este diseño de estudio.
3. Construye la matriz de contraste necesaria para realizar las comparaciones requeridas.

## 4. Efecto del programa génico termogénico durante la diferenciación de adipocitos

Los datos para este estudio están disponibles en GEO (Serie GSE100924).

El estudio que generó los datos investigó la función del gen ZBTB7B. Este gen activa el programa termogénico durante la diferenciación de adipocitos marrones, regulando la expresión génica de grasa marrón a temperatura ambiente y después de la exposición al frío. El tamaño de la muestra del experimento es de 12 muestras, tres réplicas de cada grupo. Los microarrays utilizados para este experimento fueron del tipo “Mouse Gene 2.1” de Affymetrix, ahora Thermofisher, uno de los proveedores más populares de tecnología de microarrays.

En este ejemplo, queremos verificar el efecto de la desactivación de un gen (“KO frente a WT”) por separado para la temperatura fría y ambiente. También queremos probar si la eliminación del gen afecta la manera (distinta) en que la temperatura influye en la diferenciación de adipocitos.

1. Identifica los factores experimentales y sus niveles.
2. Escribe la matriz de diseño asociada con este diseño de estudio.
3. Construye la matriz de contraste necesaria para realizar las comparaciones requeridas.

## 5. Bases moleculares de la desregulación de citocinas asociada a la edad en macrófagos estimulados con LPS

Este estudio fue publicado en el *Journal of Leukocyte Biology* (DOI: 10.1189/jlb.0106024). Los datos se encuentran disponibles en GEO (Serie GSE68580). El objetivo del experimento que generó los datos fue comprender la base molecular de los procesos regulados por una molécula (citoquina) en ratones envejecidos. Para ello se realizó un análisis de microarrays, con ARN de ratones en reposo y estimulados con lipopolisacáridos (LPS) utilizando el chip de Affymetrix “Mouse Genome 430 2.0”.

1. Identifica los factores experimentales y sus niveles.
2. Escribe la matriz de diseño asociada con este diseño de estudio.
3. ¿Cuáles crees que podrían ser las “preguntas razonables” en este estudio? Construye la matriz de contraste necesaria para realizar las comparaciones requeridas.