



## EJERCICIO TEMA 10

### Alumno:

*Realiza las distintas actividades de abajo copiando debajo de cada actividad los comandos necesarios. Para ello descárgate [la carpeta para realizar los ejercicios del Tema 10](#). Finalmente, convierte este documento a PDF y entrégalo a través del moodle.*

*1.- Utiliza el comando `gzip` para descomprimir el archivo `Homo_sapiens.GRCh38.dna.chromosome.12.fa.gz`.*

*2.- Utiliza el comando `bwa index` para generar los índices correspondientes al genoma `Homo_sapiens.GRCh38.dna.chromosome.12.fa`*

*3.- Utiliza la herramienta `bwa mem` para alinear los archivos `fastq` de la muestra `Sample1` frente al genoma que has indexado en el ejercicio 2. Convierte el archivo resultante a formato `BAM`, arregla las coordenadas de las parejas, ordénalo y marca los duplicados de PCR. Finalmente ordénalo e indéxalo.*

*4.- Utiliza la herramienta `macs3 callpeak` para generar un listado `BED` con las regiones con cobertura en la muestra.*

*5.- Utiliza la herramienta `bedtools coverage` para generar los contajes de lecturas del archivo `BAM` que corresponden a cada una de las regiones o picos identificados en el ejercicio 4.*

*6.- Modifica el script `Tema10.R` para generar un dataframe a partir de los archivos con extensión `.counts.txt` que contienen información de tres replicados biológicos de células deficientes en `ARID1A` (`H460_ARID1A`) y de células control (`H460_Empty`). El dataframe debe contener las regiones en filas y las distintas muestras en columnas.*

*7.- Modifica el script `Tema10.R` para generar un segundo dataframe con el nombre de las muestras en filas y una sola columna que contenga, en forma de factor, a qué grupo de estudio pertenece cada muestra.*

*8.- Carga la librería “`DESeq2`” para analizar los datos y genera un objetivo tipo `Deseq` utilizando la función “`DESeqDataSetFromMatrix`”*

*9.- Corre el análisis de `Deseq` utilizando la función “`DESeq`” sobre el objetivo generado en el ejercicio 8.*

*10.- Extrae los resultados del objeto `deseq` usando la función “`results`”.*

*11.- Extrae los resultados significativos como aquellos que presentan un  $\log_2\text{Foldchange} > 1$  o  $< -1$  y un  $\text{padj} \leq 0.05$ .*

*12.- Genera un archivo `BED` con las regiones que muestran cambios significativos.*

# Bioinformática y análisis de datos ómicos

© Ignacio Varela Egocheaga

Este material se publica bajo licencia Creative Commons CC BY-NC-SA 4.0



***13.- Introduce ese archivo BED en la herramienta GREAT de análisis de picos que puedes encontrar en el siguiente enlace y pega el resultado.  
<http://great.stanford.edu/public/html/>***