Peak annotation practical

Tool 1: PAVIS (https://manticore.niehs.nih.gov/pavis2/)

To fit the PAVIS input format requirement, you first need to cut the first four columns of the bed file with your peaks (output of MACS2 callpeak)
For this, use **Cut columns from a table** in Galaxy:

1. Set **Cut columns** to c1,c2,c3,c4 and **execute**.
2. Download the resulting file on your computer and adapt its name (no brackets, .bed extension)

Then on the PAVIS web page:

1. Choose the right **Species/Genome Assembly/Gene Set**: Ensembl_GRCm38/mm10 all genes
2. Choose **The query peak file to be annotated**
3. Make sure the **File format** is set to UCSC_BED
4. SUBMIT

The output shows a **PAVIS Annotation Report** with a table and graphical representation of the peak location distribution. It also has a link to **The Full Annotation File**. The Full Annotation File contains the category (Intron, Upstream, Downstream...), Ensembl gene ID, gene symbol, strand and distance to TSS for each peak that was associated with a gene.

Tool 2: GREAT (http://great.stanford.edu/public/html/)

To fit the GREAT input format requirement, you need to cut the first four columns of the bed file with your peaks (output of MACS2 callpeak), but you also have to change the name of the chromosome from X to chrX.
For this, use **Replace text in a specific column** from **previous cut file (tool1)** in Galaxy:

1. Set **Find pattern** to X
2. Set **Replace with** to chrX (case is important) and **execute**

Then in the GREAT web page:

1. Choose the right **Species Assembly** : Mouse: GRCm38 (UCSC mm10, Dec. 2011)
2. Choose the bed file to annotate for the **Test regions**
3. **Submit**