username:
password:

**RNA-Seq hands-on tutorial using Chipster: Parathyroid dataset**

Eija Korpelainen & Maria Lehtivaara, CSC –IT Center for Science, Finland, chipster@csc.fi

1. Start Chipster

Go to **http://chipster.csc.fi/**, and **Launch Chipster**. Log in.

2. Open a session

Select **Open Example session** and **course_ RNAseq_parathyroid**. Inspect the session description and the phenodata: This data contains 12 samples from 3 different patients, 2 different treatments and 2 different time points.

3. Study the effect of different factors with PCA

Select the **parathyroid_counts.tsv** and run the tool **Quality control / PCA and heatmap of samples with DESeq2**. Run the tool three times:
a) Column describing groups = group
b) Column describing groups = patient
c) Column describing groups = time.
Compare the **pca-deseq2.pdf** files (rename the files and detach them for easier comparison).
-Which factor shows most variance amongst the samples?

4. Analyze differential expression with DESeq2

Select the file **parathyroid_counts.tsv** and run the tool **RNA-seq / Differential expression using DESeq2**.
-How many differentially expressed genes do you get?

Repeat the run so that you set **Column describing additional experimental factor = patient**.
-How many differentially expressed genes do you get? And if you set the P-value cutoff to 0.1?
-How many genes are removed by the automatic independent filtering (check summary.txt)?

5. Analyze differential expression with edgeR

Let's run edgeR three times, adding one more effect (factor) on each run.
a) Select the file **parathyroid_counts.tsv** and run the tool **RNA-seq / Differential expression using edgeR for multivariate experiments**, and set the effects so that the **Main effect 1 = group**, and leave the other two effect fields EMPTY for now. Filter out genes that are not expressed in at least **3 samples.**
b) Run as above but set also **Main effect 2 = time**.
c) Run as above but set also **Main effect 3 = patient**.

6. Filter the results

Select the three **edger-glm.tsv** files and run the tool **Utilities / Filter table by column value** for them setting the parameters as follows:

**Column to filter by** = **FDR-as.factor(group)2**
**Does the first column have a title = no.**

- Which edgeR run produced most DE genes? How many are they? And if you filter with FDR < 0.1?
-How would you see the number of differentially expressed genes between the different time points?

7. Compare the result lists

Select the **de-list-deseq2.tsv** and **filtered-ngs-result.tsv** (FDR < 0.1) and select the **Venn-diagram** as visualization method.
-How many genes do the lists have in common?

8. Annotate the results

In order to see the gene symbols and descriptions of the DE genes, choose the **filtered-ngs-result.tsv** (FDR < 0.1) and the tool **Utilities / Annotate Ensembl IDs**. Set **species = human**.