Expander v7.1– hands-on session:

Launching Expander:

In your Expander directory double click on **Expander.bat**.

- Before starting please create a directory named "results" in the Expander directory.

Human CAL51 cell line analysis:

In this section, we will analyze a dataset published by Rashi-Elkeles, Warnatz and Elkon et al. (Science Signaling 2014) in which HiSEq 200 Illumina RNA-Seq data was used to measure time-series expression profiles from CAL51 cell line under IR-induction for 2 biological replicates with 5 samples each – Control (0h), Control (4h), IR (4h), Control (8h) and IR (8h).

1. Loading the data:

- From the "File" menu, select "Load Data" → "Expression Data" → "Tabular Data File".
- In the "Load Study" dialog box, make sure the Organism is set to "**Human**". Use the "Browse" button and search for the file *CAL51_IR_exp.groups.ensid.txt*.
- Add an ID conversion file using "Browse" button and search for the file *ens_ent_map.txt*.
- Make sure that the "Data type" is set to "Relative Intensities".
- Press "OK" in the "Load Study" dialog box.
- If a question pops up asking if you would like to download data for human, press Ok and wait for the download process to end.

2. Preprocessing the data:

- From the "Preprocessing" menu, select "Filter Probes" → "Fold Change". In the dialog box select probes that change in 2 folds in at least 2 conditions. Press "OK". A message will pop-up saying the number of probes will remain. Press "Yes".
- From the "Preprocessing" menu, select "Standardization" → "Mean 0 and Variance 1".

3. Using t-test statistics to detect differential expression:

- From the "Supervised Grouping" menu, select "Differential Expression" → "t-Test". Click "Yes" for performing T-test on standardized values.
  - In the "Requested type of change" combo-box select "Differential".
  - Press the "Select" button for the "Group 1 conditions".
  - In the "Select Conditions" dialog box, check (mark) 'Control' samples (checkbox) press "OK".
  - Press the "Select" button for the "Group 2 conditions".
  - In the "Select Conditions" dialog box, check (mark) 'IR' samples (checkbox) and press "OK".
  - Press "Ok".
- Look at the resulting groups on the left.
- Click on the pattern chart of the up-regulated group. A table will appear on the right pane. It contains the relevant t-test results.
- Click on the p-value column title to sort the probes according to their p-values.
- Click on the t-score column title, results will be sorted according to their t-score.
- Select the "Expression Matrix" tab to view the expression matrix of the up-regulated probes.

- From the “File” menu, select “Export to text”.
- Save the results to “*ttest1.1.txt*”
- Open the directory using Windows Explorer and browse through it to see the content of this output file.

4. Performing hierarchical clustering on the data:

- From the “Unsupervised Grouping” menu, select “Hierarchical Clustering”>>”Cluster”.
- In the dialog box, select “Average” as linkage type. Press the “OK” button.
- Look at the resulting display.
- Use the “Zoom in” and "Zoom out" buttons from tool bar to change zoom, and get a closer or more general view of the expression patterns arranged in the hierarchical clustering order.
- From the “File” menu, select “Save As Image”. In the dialog box select only "Dendrogram with matrix" Save the results to folder “*Hier1.1*”.
- Open the directory using Windows Explorer and browse through it to see the content of this output file.

5. Identifying enriched GO categories within the clusters:

- From the “Enrichment Analysis” menu, select “Functional Analysis” → “TANGO”. Make sure that in the “Functional Analysis” dialog box “T-test 1.1” is selected as the grouping solution on which analysis is to be performed. Press the “OK” button.
- Look at the resulting display (diagrams and enrichment table). Which functional enrichment has been detected?
- Click on one of the columns in the enrichment diagram. The “Enrichment Info” dialog box will appear.
- Click on one of the Gene IDs in the table. An internet browser will open, containing information about the corresponding gene, from the NCBI-Gene site.

6. Identifying enriched KEGG pathways within the clusters:

- From the “Enrichment Analysis” menu, select “Pathway Analysis” → “KEGG…”. Make sure that in the “Pathway Analysis” dialog box “T-test 1.1” is selected as the grouping solution on which analysis is to be performed. Select "**Original GE data**" as background genes. Press the “OK” button.
- Look at the resulting display (diagrams and enrichment table). Which pathway enrichments have been detected? Is "p53 signaling pathway" among the pathways found?
- Click on one of the columns in the enrichment diagram. The “Enrichment Info” dialog box will appear.

7. Browsing results for up regulated genes:

- Go back to “T-test 1.1” tab.
- Select “Up regulated” from the list on the left. The entire set of results produced for up regulated genes will appear on the right. Browse through the results.
- Open the “Expression Matrix” tab to view the columns added on the left for each enrichment that has been detected.

8. GSEA using gene-groups, generated in Expander, as gene sets

- Repeat step 1. Do not continue current session.
- Select "Enrichment analysis" → "GSEA". In the dialog box select the option "Merge Probes by Gene IDs". In the "Average Probes" select "Average over current values" and press "Ok".
- In the GSEA dialog box – press on "Rank by GE data" and choose the test condition "C.IR.4h.a.logR0" in the drop down list.
- Press "Ok"
- Look on the left table. Click on "q-value" column to sort the "Wikipathways" gene sets from the lowest q-value. Click on the first row. A visualization will appear on the right panel showing the enrichment score, hits and the ranks of the genes.
- Click on the "Leading edge table" on the right panel. Press on the "Hit" column to sort be genes located at the leading edge (i.e., marked with 1 in the "Hit" column).

9. Loading ChIP-seq data

- Load ChIP-Seq data file via *ChIP-Seq Analysis→Load ChIP-Seq Data.*
- A window will pop-up to ask if to continue with the current session, press "Yes".
- Click on the "Browse" button and search for the file *p53_t2h_analysis_peaks_hg19_IR_induced_filtered.bed*.
- A dialog box will appear asking whether you wish to download required data, press the "Yes" button. Required data should be loaded within 1-5 minutes. A ChIP-Seq visualization tab should appear.
- Look on the top left "ChIP-Seq Info" panel. How many genes have a peak hit within the search area?
- Peaks that didn't hit a gene are considered "Intergenic", how many are there according to the pie chart?
- Select "Regions hits enrichment" tab on the right panel. Look on the "Upstream TSS" bar. It is expected compared to random hits that peaks will fall upstream of the TSS (p-value = 0).
- Select "Peaks Annotations" tab on the right panel. A table will be displayed showing the annotations of the peaks.

10. Identifying GO categories within the ChIP-Seq set of genes with a hit

- Select *ChIP-Seq Analysis*→Functional analysis→TANGO…
- In the dialog box, perform analysis on "ChIP-Seq Data 1".
- In the "Background Set" panel, select "**Original GE data**". This option will take the set of background genes from the loaded "GE Data 11" tab.
- Press "Ok"
- Do you see any involvement of p53 transcription factor (TF)?

References

1) S. Rashi-Elkeles, H.J. Warnatz, R. Elkon, A. Kupershtein, Y. Chobod, A. Paz, V. Amstislavskiy, M. Sultan, H. Safer, W. Nietfeld, H. Lehrach, R. Shamir, M.L. Yaspo and Y. Shiloh. Parallel Profiling of the Transcriptome, Cistrome, and Epigenome in the Cellular Response to Ionizing Radiation. Science Signaling Vol. 7, 325, RS3, 2014.