Expander v7.1.1 new features – hands-on session:

1) Loading GE data:

   - From the "File" menu, select "New Session" → "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) GSEA using gene-groups, generated in Expander, as gene sets:
   - 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).

3) 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.

4) GSEA on ChIP-Seq Data as a single gene set:
   - Repeat section 2 with the following change – Under "Collection Group" select "Grouping solution" and choose "ChIP-Seq Data 1" from the drop down list.
   - Do you see that a large fraction of the ChIP-Seq genes (135/371) appear in the leading edge?

5) 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)?

6) ChIP-Seq distribution within GE data:
   - Select ChIP-Seq Analysis→Integration with expression data→ChIP-Seq vs. GE analysis
   - Select "C.IR.4h.a.logR0" as a test condition and Press "Ok".
   - Are the ChIP-Seq genes with a hit tend to be ranked high according to the selected condition? Is it significant?

7) Loading GE data:
   - Repeat section 1. Click "**No**" in the "continue current session" dialog box.

8) Preprocessing the GE 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".

9) Performing K-Means clustering on the data:

   - From the "Unsupervised Grouping" menu, select "K-Means Clustering".
   - In the dialog box, type 6 clusters. Press the "OK" button.
   - Look at the resulting groups on the left.
   - Click on the pattern cluster chart with >300 probes. The number of this cluster will be defined as "x". A table will appear on the right pane. It contains the relevant K-Means results.
   - Select the "Expression Matrix" tab to view the expression matrix of the selected cluster.

10) AMADEUS motif finding on K-Means clusters:
   - Select "Enrichment Analsysis"→"Promoter Analysis"→"AMADEUS".
   - In the dialog box, in the "perform analysis on" drop-down list select "K-Means 1.1" clustering solution. In the "Background set" choose "**Original GE data**".
   - Press "Ok".
   - Click on one of the bars inside diagram x (the cluster we defined in section 9).
   - Amadeus visualization will pop-up showing all motifs found.
   - Do you see "p53" TF in the first ranked motif?

11) Loading ChIP-Seq data:

- Repeat section 3. Click "**Yes**" in the "continue current session" dialog box.

12) Identify K-Means clusters that are enriched with genes that are closest to ChIP-Seq peaks:

- Select ChIP-Seq Analysis→Integration with expression data→ChIP-Seq Enrichment.
- In the dialog box, in the "perform analysis on" drop-down list select "K-Means 1.1" clustering solution.
- Press "Ok".
- A new tab will be generated showing clusters that contain a significantly high number of genes closest to peak.
- You should see only a single cluster enriched with ChIP-Seq genes. Is this the cluster of genes that have a positive expression under IR treatment?

13) Generating ChIP-Seq peaks sequences for AMADEUS De-Novo motif finding:

- Select ChIP-Seq Analysis→Motif analysis→Fetch ChIP-Seq sequences.
- Change the "Top number of peaks" to 500.
- Press "Ok".
- A dialog box will appear asking whether you wish to download required genome data, press the "Yes" button. Required data should be loaded within 5-15 minutes.
- After downloading the genome data and creating the peaks sequences an AMADEUS dialog box would appear. Press "Ok".
- Click on the single bar on the diagram. Amadeus visualization will be shown.
- You should see a single motif linked with p53 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.