Hands-on GSEA

by Dr. Ester Feldmesser, March 2020

Part 1 - GSEA hands-on

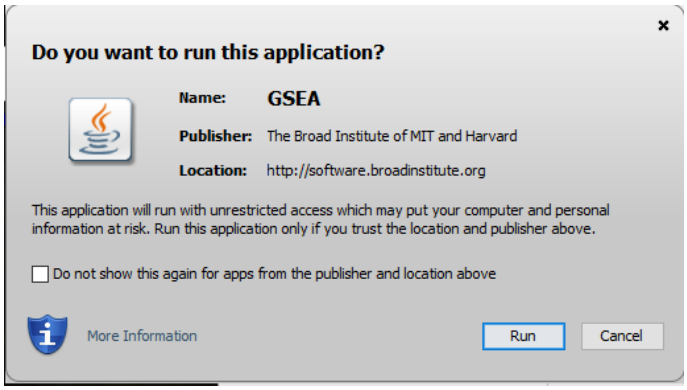
The data for this exercise was taken from the paper by Umansky, Groner et al called "Runx1 Transcription Factor Is Required for Myoblasts Proliferation during Muscle Regeneration" (http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005457#abstract1). In this paper the function of Runx1 in muscle regeneration is investigated. One of the analysis performed was RNA-seq comparing mice lacking dystrophin and muscle Runx1 (*mdx*$^{-}$*/Runx1*$^{f/f}$) to mice lacking dystrophin and having muscle Runx1 (*mdx*$^{-}$*/Runx1*$^{L/L}$). The mdx mice are a model of Duchenne muscular dystrophy. In contrast to human, the mdx mice model (with wt Runx1) regenerates constantly the muscle.

The RNA-seq data was analyzed and an input file was prepared. The input file is called "MdxVsMdxKO_Capital.rnk" (the KO is for Runx1) and is located http://dors.weizmann.ac.il/course/GSEA/. Please download it to your PC, by right clicking and choosing "Save link as...". This is a text file and can be opened with Excel. The suffix in the name is part of the file format required for the analysis we are performing by GSEA. There are two columns in the file: the gene name and the fold change; and it is sorted according to fold changes. The file includes all the genes that had more than 10 reads in at least one sample. Note that the GSEA software was developed for human data, therefore the mouse gene names were converted into their human orthologues.

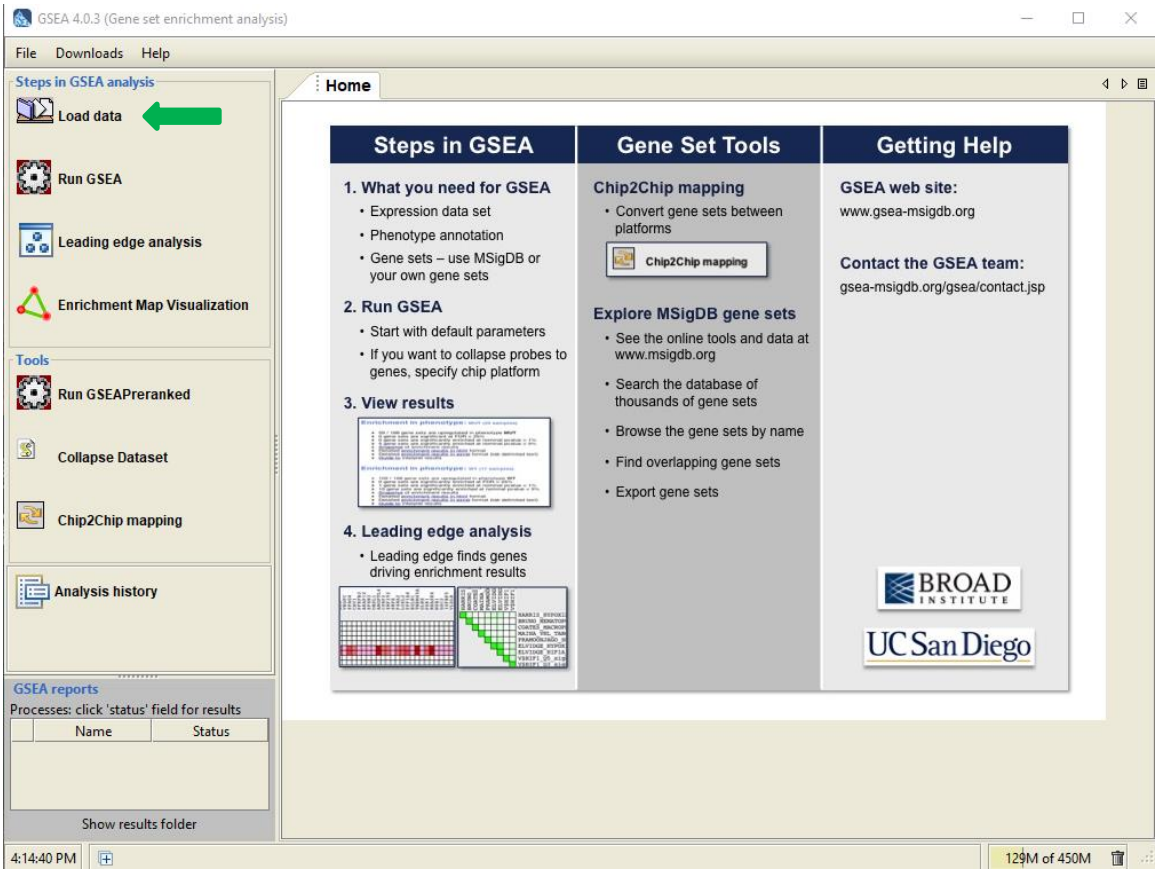
Our workflow:

1. Open GSEA by double clicking in the icon at the desktop:

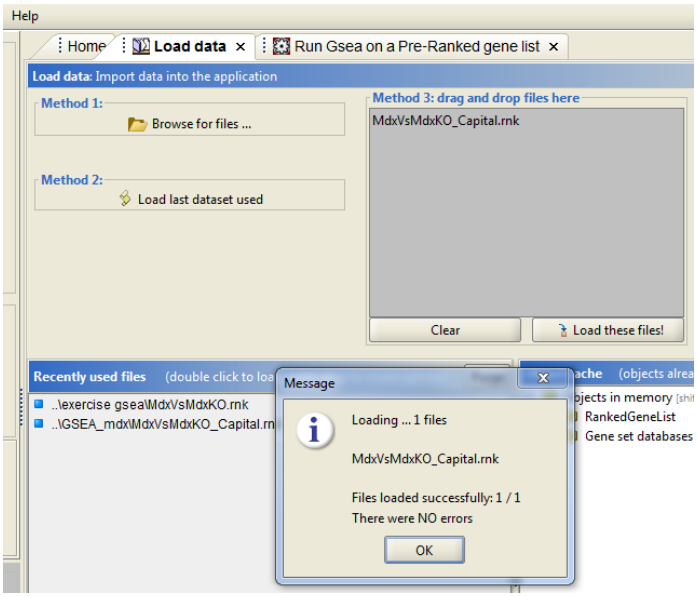
2. When the following dialog box appears, click on Run.



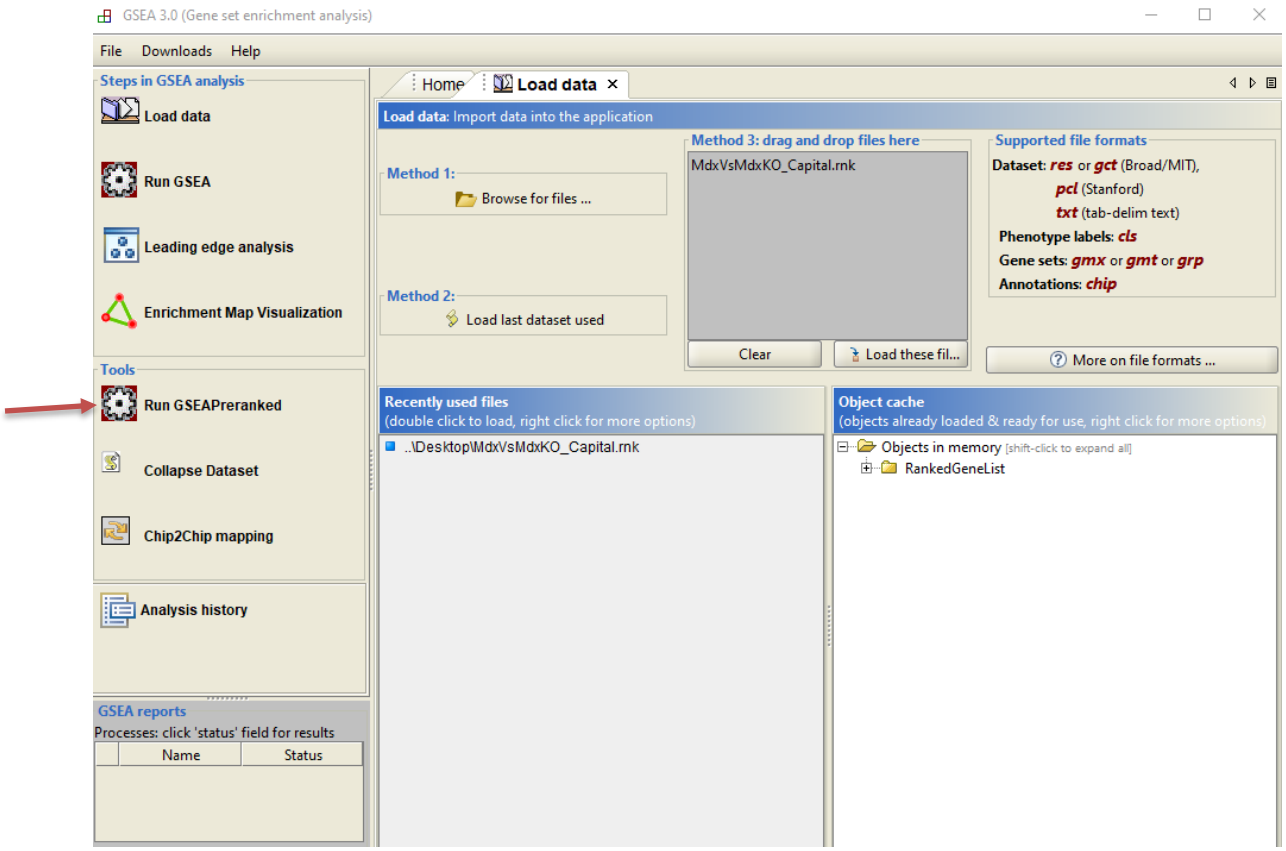
3. GSEA will open. Click on Load data (green arrow).



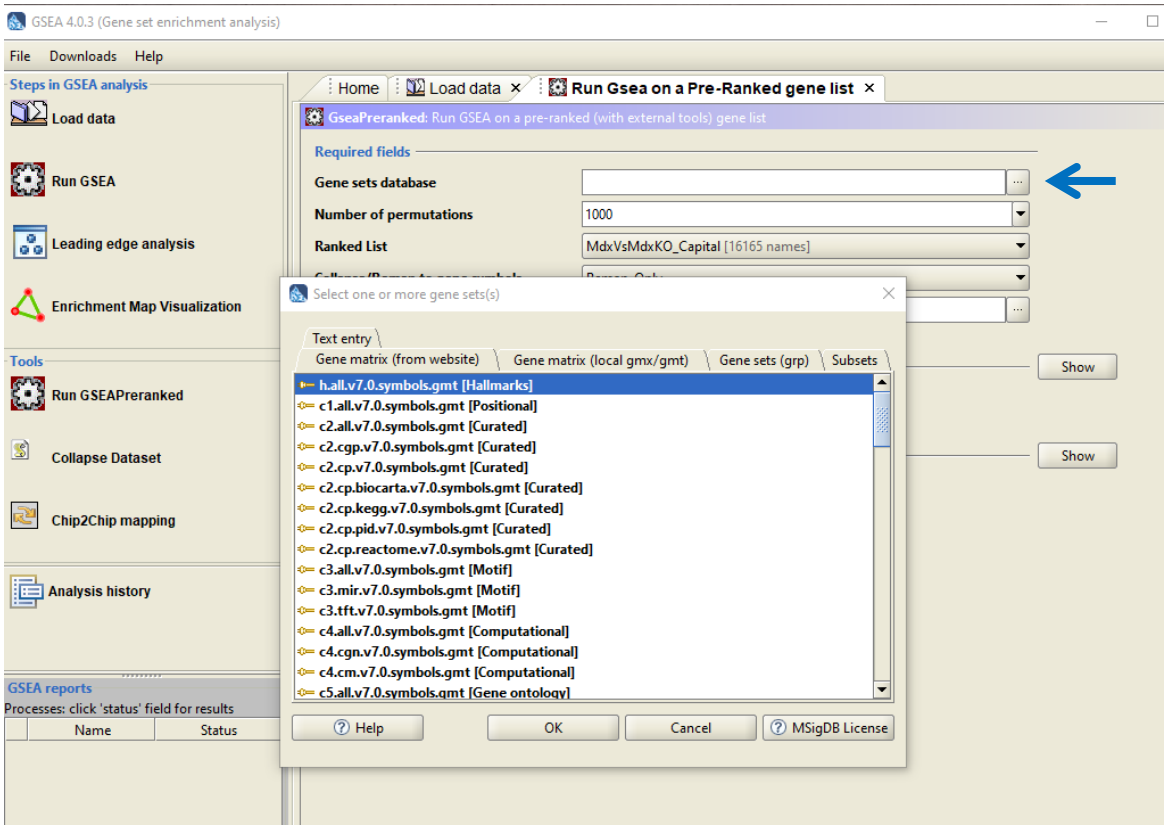
4. The easiest way to load the file is to drag it into the gray box, then click on Load these files! A message indicating that the file was successfully loaded should appear. Click OK.



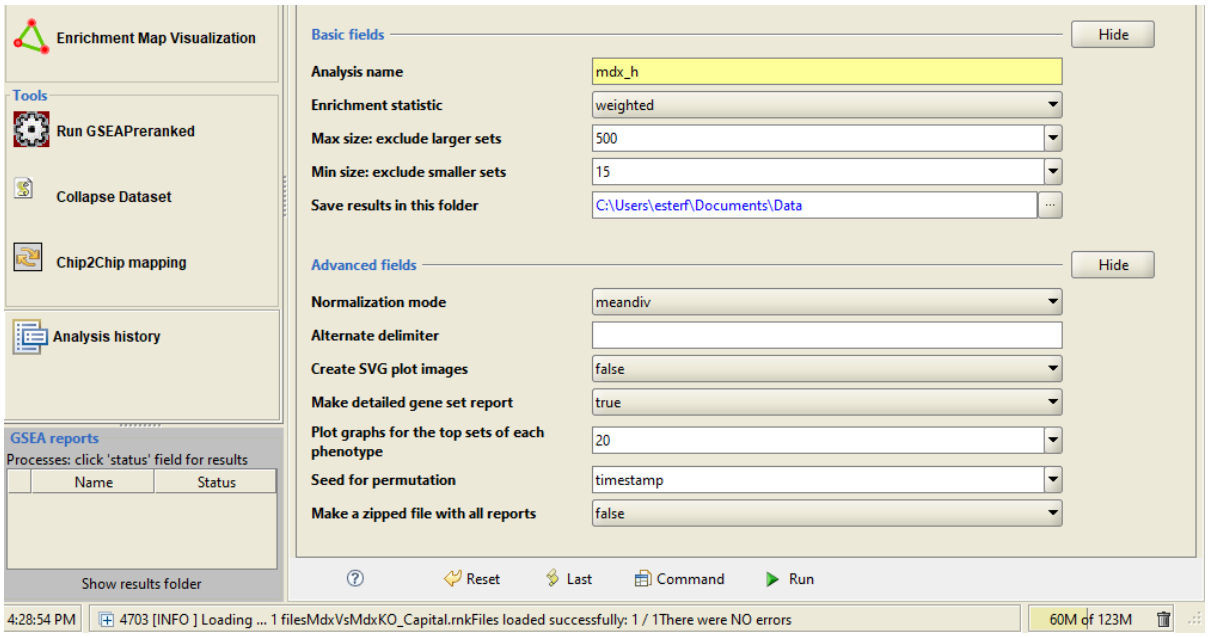
5. Go to Tools menu on the left bar and choose Run GSEAPreranked (red arrow).



6. We will start filling the form with the Gene sets database. Click on the three points (blue arrow), a dialog box will appear, select the first one (Hallmarks) and click OK.

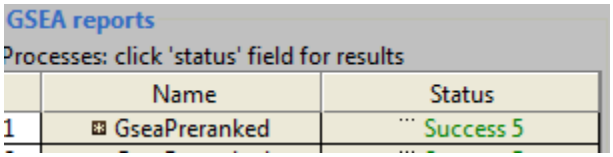


7. Click on the two Show buttons. Be sure that all the parameters are defined as shown below. Change the analysis name and folder for saving the results as you wish.



8. On the lower part of the window click on Run.

9. Look at the lower right panel.



10. When the analysis finishes, click on the word Success. A web page will open showing the results. An additional way to get to the results (after you close the application) is by double clicking on the file called index.html inside the gsea results folder.

11. In the newly opened web page of the results there are 2 parts with the title **Enrichment in phenotype : na.** The first one is for the results of the upregulated genes and the second one for the down regulated genes. First take a look at the Guide to interpret results.

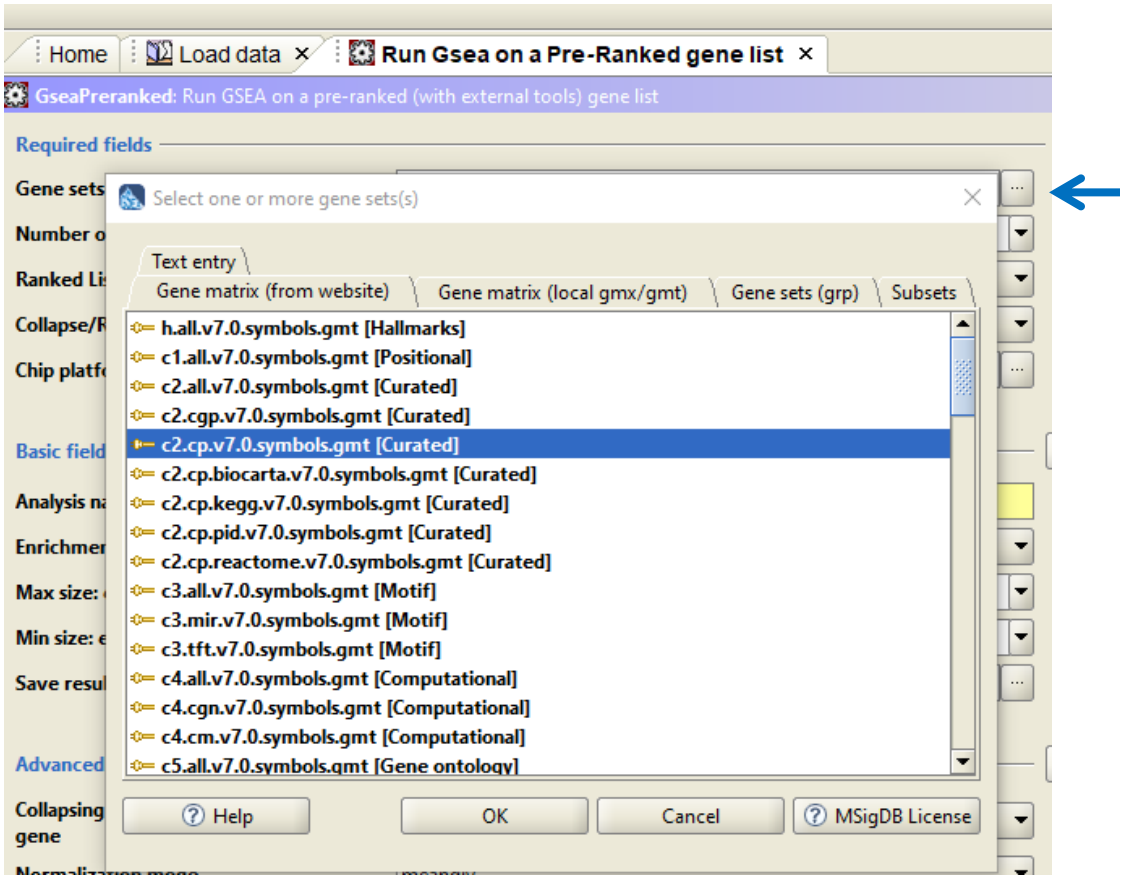
12. Open the Snapshot link (preferably in a new tab) to see the enrichment score graphs of the most significant gene sets. Some of them look very convincing and some of them are noisy. Not all of them are significant.

    Question 1: Looking at the snapshot for the positive phenotype, which hallmarks are clearly enriched with upregulated genes?

13. To see more details, return to the results page and click on “Detailed enrichment results in html format” link (preferably in a new tab). You will see a table with the results. Compare the FDRs to the snapshots that you saw before. To view the details of the analysis and the genes that contribute to the score go to Details… (or press on the relevant snapshot). To

view the gene set definition, click on its name. You may be requested to register. Please do it. Let's look at the genes included in the gene set HALLMARK_MYOGENESIS.

14. We will perform an additional analysis in GSEA to get more detailed information (using the same data). This time we will choose to run our genes against canonical pathways from several databases (the Gene set that starts with c2.cp.v7…). Remember to change the name of your analysis.



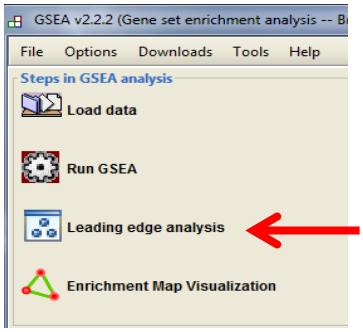
15. This analysis will take longer than before because there are more gene sets in the canonical pathways. Look at the table of results when the analysis finishes.

Question 2: Are the obtained significant gene sets (pathways) in agreement with the previously obtained HALLMARK_MYOGENESIS? Why?

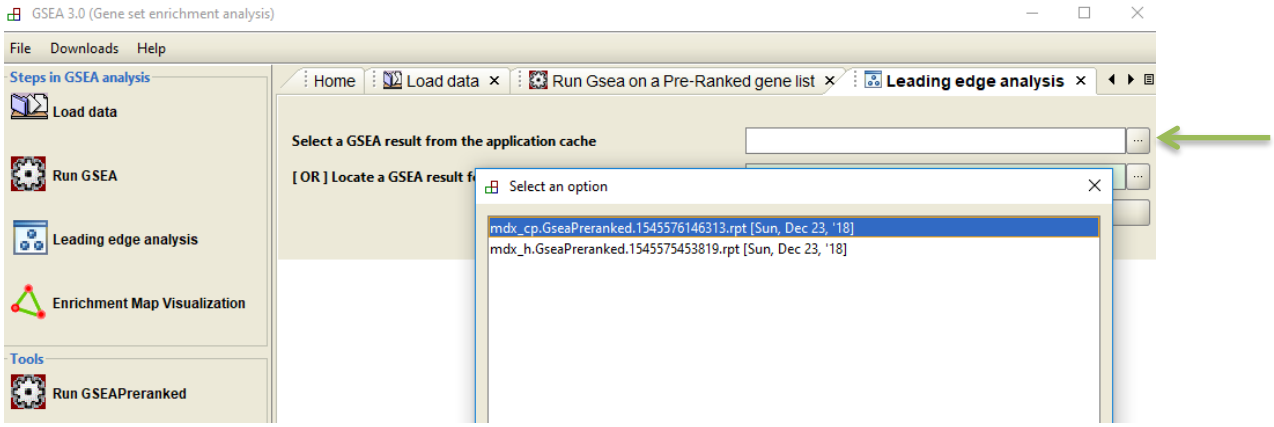
Question 3: Are the obtained significant gene sets (pathways) in agreement with the previously obtained HALLMARK_OXIDATIVE_PHOSPHORYLATION?

16. Since there are much more significant gene sets in this analysis, it is difficult to look at long tables and there is overlap between the pathways, we will look at the graphic summaries of the results using a tool.

17. Return to the GSEA software and choose the Leading edge analysis (red arrow).

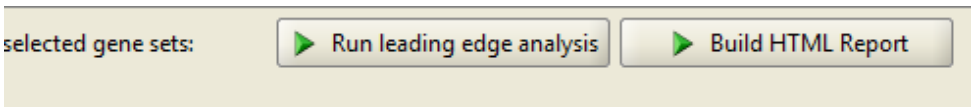


18. Select a GSEA result from the application cache (green arrow), a window will open, select the canonical pathways result and click OK. Then click on Load GSEA Results.



19. Select the rows of the Gene Sets in the table with FDR of 0 and positive NES and click on Run leading edge analysis.

    We select a number of gene sets (43) that enables visualization. You could choose a different cutoff or use other criteria.



    To understand the results go to: http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html?_Interpreting_GSEA_Results. Look for Interpreting Leading Edge Analysis Results in the left panel.

Question 4: Back to the results, look at top right graph. Find two sets of pathways (with more than 3 pathways) that their genes are overlapping. Which pathways are in each set and what function do you think is associated with each of them?

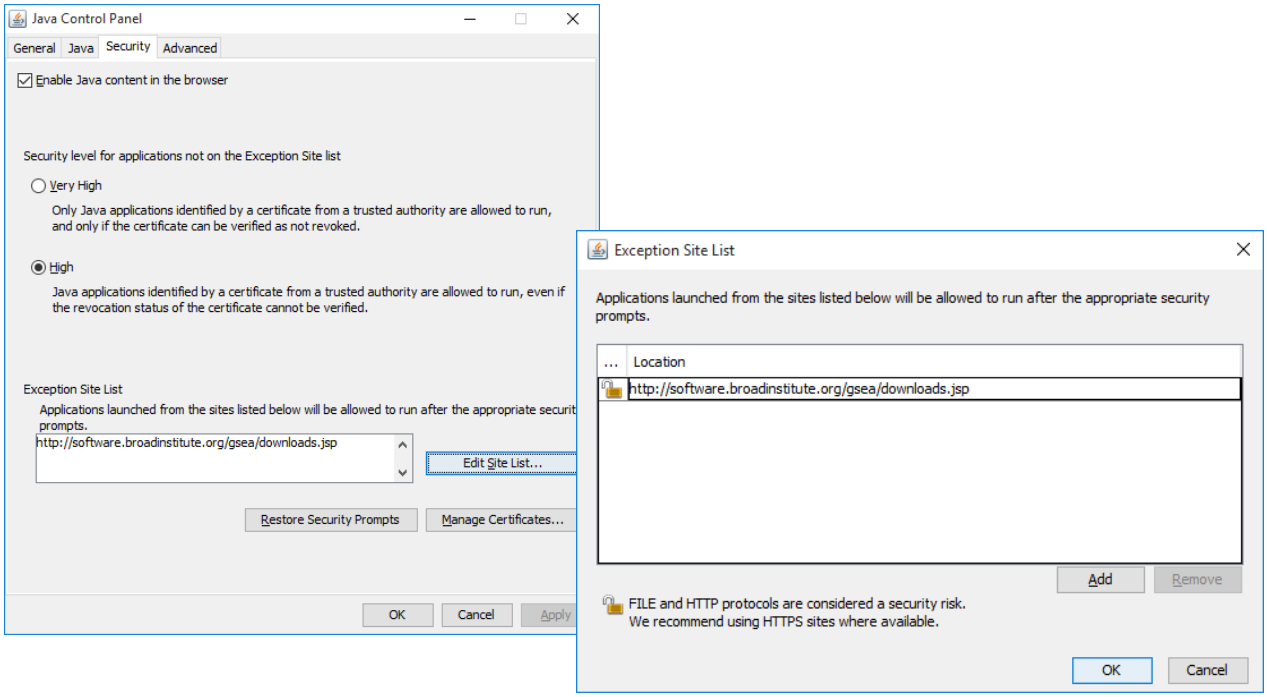
Tips: 1. Neurodegenerative diseases are related to mitochondrial processes (oxidative phosphorylation), 2. Names of pathways can be selected and copied (Ctrl +C) from the plot in the GSEA Leading edge analysis.

Appendix: Installation of GSEA

In case you want to install the software in your computer for further use.

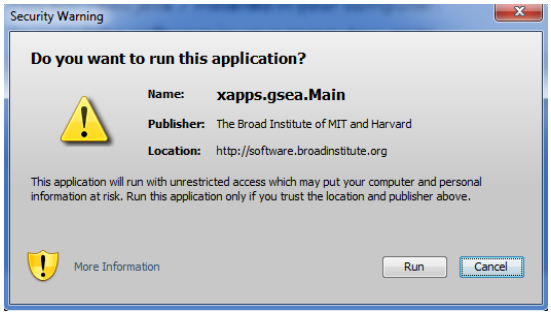
1. You will need java 8 installed on your computer.

2. Update the Java security: From Control Panel of your computer, open Java. In the Java control panel, select the Security tab, press the Edit Site List button, then press Add in the Exception Site List pop-up window and add the URL bellow (in 3.)



3. To install the GSEA software in your computer, go to https://www.gsea-msigdb.org/gsea/downloads.jsp, click on the orange button Launch to install the software. Save the gsea.jnlp file and double click on it. You may be requested to register during the process.

4. Click on Run when prompted.


Security Warning
Do you want to run this application?
Name: xapps.gsea.Main
Publisher: The Broad Institute of MIT and Harvard
Location: http://software.broadinstitute.org
This application will run with unrestricted access which may put your computer and personal information at risk. Run this application only if you trust the location and publisher above.
More Information
Run
Cancel