# **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" (<u>http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005457#abstract1</u>). 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<sup>-</sup>/Runx1<sup>f/f</sup>*) to mice lacking dystrophin and having muscle Runx1 (*mdx<sup>-</sup>/Runx1<sup>L/L</sup>*). 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 <u>http://dors.weizmann.ac.il/course/GSEA/</u>. 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:

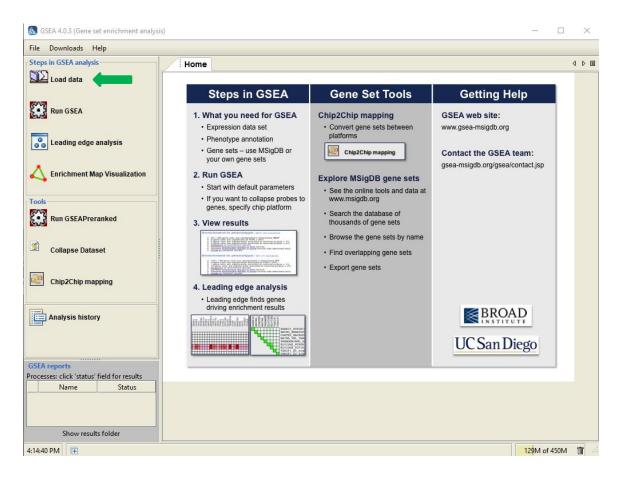




2. When the following dialog box appears, click on Run.

x Do you want to run this application?					
	4	Name:	GSEA		
	<u>ڪ</u>	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.					
Do not show this again for apps from the publisher and location above					
	More Informa	ation		Run Cancel	

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



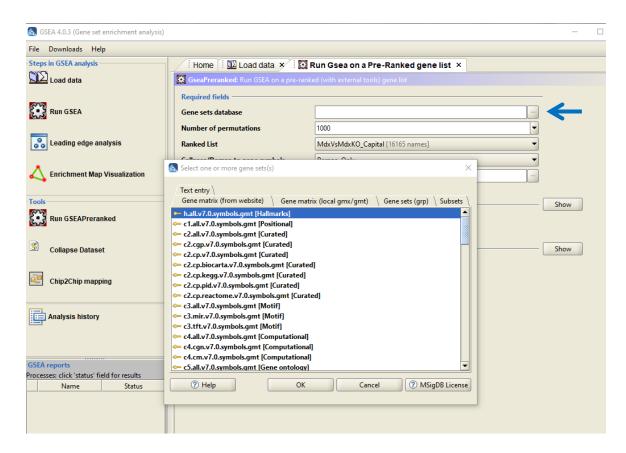
4. The easiest way to load the file is to drag it into the gray box, then click on Load these files! A message indicating that the file was successfully loaded should appear. Click OK.

Home 🗄 🕎 Load data 🗙 🗍	🔛 Run Gse	ea on a Pre-Ranked gene lis	it ×
Load data: Import data into the application	on		
Method 1:		Method 3: drag and drop fi	les here
📂 Browse for files		MdxVsMdxKO_Capital.rnk	
Method 2:			
		Clear	Load these files!
Recently used files (double click to lo	Message		🗙 ache (objects a
Recently used files (double click to lo	Message		ojects in memory
	Message	Loading 1 files	ojects in memory I RankedGeneLis
\exercise gsea\MdxVsMdxKO.rnk	m i		ojects in memory I RankedGeneLis
\exercise gsea\MdxVsMdxKO.rnk	m 1	Loading 1 files MdxVsMdxKO_Capital.rnk	ojects in memory I RankedGeneLis
\exercise gsea\MdxVsMdxKO.rnk	m	Loading 1 files	ojects in memory

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

Steps in GSEA analysis	Home Doad data ×			4 Þ
Load data	Load data: Import data into the application			
Run GSEA	Method 1:	Method 3: drag and MdxVsMdxKO_Capita		Supported file formats Dataset: <i>res</i> or <i>gct</i> (Broad/MIT), <i>pcl</i> (Stanford) <i>txt</i> (tab-delim text)
Leading edge analysis				Phenotype labels: <i>cls</i> Gene sets: <i>gmx</i> or <i>gmt</i> or <i>grp</i> Annotations: <i>chip</i>
A Enrichment Map Visualization	Method 2: S Load last dataset used			· · ·
Tools		Clear	Load these fil	⑦ More on file formats
Run GSEAPreranked	Recently used files (double click to load, right click for more opt		<b>Object cache</b> (objects already loade	ed & ready for use, right click for more option
S Collapse Dataset	.\Desktop\MdxVsMdxKO_Capital.mk		⊡ 🗁 Objects in mer ⊡ 🔁 RankedGer	nory [shift-click to expand all] neList
Chip2Chip mapping				
Analysis history				
GSEA reports Processes: click 'status' field for results				
Name Status				

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.

A Enrichment Map Visualization	Basic fields		Hide
	Analysis name	mdx_h	
Tools	Enrichment statistic	weighted	
Run GSEAPreranked	Max size: exclude larger sets	500	
	Min size: exclude smaller sets	15	
S Collapse Dataset	Save results in this folder	C:\Users\esterf\Documents\Data	
Chip2Chip mapping	Advanced fields		Hide
	Normalization mode	meandiv	
Analysis history	Alternate delimiter		
	Create SVG plot images	false	
	Make detailed gene set report	true	
GSEA reports Processes: click 'status' field for results	Plot graphs for the top sets of each phenotype	20 💌	
Name Status	Seed for permutation	timestamp 💌	
	Make a zipped file with all reports	false	
Show results folder	⑦	st 💼 Command 🕨 Run	
4:28:54 PM	ilesMdxVsMdxKO_Capital.rnkFiles loaded succe	essfully: 1 / 1There were NO errors 60	M of 123M

- 8. On the lower part of the window click on Run.
- 9. Look at the lower right panel.

GS	GSEA reports					
Pro	cesses: click 'status' field fo	r results				
	Name	Status				
1	GseaPreranked	··· Success 5				
-						

- 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.

Home 🗄 🕎 Load data 🗴 🗄 🔛 Run Gsea on a Pre-Ranked gene list 🗴			
GseaPreranked: Run GSEA on a pre-ranked (with external tools) gene list			
Required fields			
Gene sets 🔬 Select one or more gene sets(s) X			
Number o			
Text entry			
Ranked Lie Gene matrix (from website) Gene matrix (local gmx/gmt) Gene sets (grp) Subsets			
Collapse/F 🗠 h.all.v7.0.symbols.gmt [Hallmarks]			
Chip platfe *= c1.all.v7.0.symbols.gmt [Positional]			
← c2.all.v7.0.symbols.gmt [Curated]			
e= c2.cgp.v7.0.symbols.gmt [Curated]			
+= c2.cp.v7.0.symbols.gmt [Curated]			
Image: the state of the sta			
Analysis n: 0= c2.cp.kegg.v7.0.symbols.gmt [Curated] 0= c2.cp.pid.v7.0.symbols.gmt [Curated]			
Enrichmer $0 = c2.cp.pid.v7.0.symbols.gmt [Curated]$			
Max size: 0= c3.all.v7.0.symbols.gmt [Motif]			
the c3.mir.v7.0.symbols.amt [Motif]			
Min size: e 🐗 c3.tft.v7.0.symbols.gmt [Motif]			
Save resul 🕫 c4.all.v7.0.symbols.gmt [Computational]			
the c4.cgn.v7.0.symbols.gmt [Computational]			
e c4.cm.v7.0.symbols.gmt [Computational]			
Advanced 🗠 c5.all.v7.0.symbols.qmt [Gene ontology]			
Collapsing ⑦ Help OK Cancel ⑦ MSigDB License			
gene			
Normalizzuan mana			

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.

ⓓ GSEA 3.0 (Gene set enrichment analysis) – □ ×					
File Downloads Help					
Steps in GSEA analysis	🔢 Home 🕴 🏧 Load data	a 🗴 📋 🔛 Run Gsea on a Pre-Ranke	ed gene list 🗴 🗄 🐱 Leading edge anal	lysis × 🚺 🕨 🗉	
Load data					
	Select a GSEA result from th	e application cache			
Run GSEA	[ OR ] Locate a GSEA result f	delect an option		×	
Leading edge analysis		mdx_cp.GseaPreranked.1545576146313.rp mdx_h.GseaPreranked.1545575453819.rpt			
Lnrichment Map Visualization					
Tools Run GSEAPreranked					

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.

selected gene sets:	Run leading edge analysis	Build HTML Report	

#### To understand the results go to:

<u>http://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html?</u> Interpreting <u>GSEA\_Results</u>. Look for <u>Interpreting Leading Edge Analysis Results</u> 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.)

🛃 Java Control Panel 🛛 🗕		$\times$	
General Java Security Advanced			
Enable Java content in the browser			
Security level for applications not on the Exception Site list			
○ <u>V</u> ery High			
Only Java applications identified by a certificate from a trusted authority are a and only if the certificate can be verified as not revoked.	allowed to ru	n,	
() High		2	🔊 Exception Site List 🛛 🕹
Java applications identified by a certificate from a trusted authority are allower the revocation status of the certificate cannot be verified.	d to run, ev		Applications launched from the sites listed below will be allowed to run after the appropriate security prompts.
			Location
Exception Site List			http://software.broadinstitute.org/gsea/downloads.jsp
Applications launched from the sites listed below will be allowed to run after the a prompts. http://software.broadinstitute.org/gsea/downloads.jsp	appropriate : dit <u>S</u> ite List		
Restore Security Prompts Manag	ge Certificate	2S	
OK Car	ncel	Apply	FILE and HTTP protocols are considered a security risk. We recommend using HTTPS sites where available.
			OK Cancel

- To install the GSEA software in your computer, go to <u>https://www.gsea-</u> <u>msigdb.org/gsea/downloads.jsp</u>, 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 W	arning		×	
Do you want to run this application?				
		Name:	xapps.gsea.Main	
	<u></u>	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.				
1	More Inform	ation	Run Cancel	