GSEA tutorial

You can choose to do these exercises using the questions as your only guide - or see the following pages for the step-by-step checklist to finding these answers.

The data set used for this practical lab contains transcriptomics data obtained from MCF7 cells, a human breast cancer line, treated or non treated with estradiol. The cells were treated with estradiol for 12, 24 or 48 hours. Total RNA extracted from the cells was amplified, labeled and hybridized to Affymetrix GeneChip U133 Plus 2.0 microarrays. The data are available in the Gene Expression Omnibus (GEO) repository under the accession number GSE11352 (PMID: <u>17542648</u>). The practical lab contains two exercises. Exercise 1 uses GSEA (http://www.broadinstitute.org/gsea/index.jsp) to perform gene-set enrichment analysis and exercise 2 uses g:Profiler (http://bit.cs.ut.ee/gprofiler/).

For this exercise, our goal is to upload the 3 required files into GSEA, set up the parameters, run GSEA, open and explore the gene-set enrichment results. We use as input file for GSEA the normalized data for all samples included in the GSE11352 dataset and formatted as a '.gct' file. GSEA will assess the amplitude of differential gene expression levels between the two groups of interest, in this case the treated samples and non treated samples at 12 hours using a t-test for each gene. The '.cls' file tells GSEA which samples correspond to our groups of interest. GSEA ranks the genes based on t values from the t-test and performs the gene-set enrichment analysis using a modified Kolmogorov-Smirnov statistics. The output result folder contains several files, and two of them are the summary tables displaying enrichment statistics for each gene-set (pathway) that has been tested and contained in the provided '.gmt' file. The '.gmt' file (gene-set file) provided for this exercise contains gene-sets obtained from KEGG, MsigDB-c2, NCI, Biocarta, IOB, Netpath, HumanCyc, Reactome and the Gene Ontology (GO) databases. (http://baderlab.org/GeneSets).

Before starting this exercise, launch GSEA using the instructions provided on the wiki and download the 3 required files:

- MCF7_Expression_matrix.gct
- MCF7_groups.cls
- Human_GO_AllPathways.gmt

STEP BY STEP:

Step	Action	Check
1	Launch GSEA.	
2	Locate the 'Load data' icon at the upper left corner of the window and click on it.	
3	In the central panel, select ' Method 1' and ' Browse for files' . A new window pops up. Browse your computer to locate the 3 files : Import the MCF7_Expression_matrix.gct, MCF7_groups.cls and Human_GO_AllPathways.gmt. Click on ' Choose' . A message pops us when the	

	files are loaded successfully. Click on 'OK'.	
	Alternatively, you can choose ' Method 3' to ' drag and drop files here' ; you need to click on the ' Load these files!' button in this case.	
4	Locate the ' Run GSEA ' icon at the left side of the main window located below the 'Load data button' and click on it.	
5	In the central window called 'GSEA: Set parameters and run enrichment tests', fill the first field called 'Expression dataset' by clicking on the up and down arrows. Choose MCF7_Expression_matrix. Tip: Mousing over the parameters fields will highlight a short description.	
6	Click on the 3 dots [] of the radio button corresponding to the 'Gene sets database' field. A new window will pop up after approximately 10 seconds. Using the right arrow in the menu bar of this window, locate the 'Gene Matrix (local gmx/gmt)' tab and select the file Human_GO_AllPathways.gmt. Click on 'OK'.	
7	In the field ' Number of permutations' enter the number 100. Note: for this exercise and purpose of demonstration, please use 100. For real life data analysis, 2000 permutations is recommended and it will require about 1 hour to run using a complete set of gene-sets.	
8	In the ' Phenotype labels ' field, click on the 3 dots [] of the radio button. A window " Select a phenotype " will pop up. Make sure that the file MCF7_groups.cls appears as selected source file; locate and select the comparison ES12_versus_NT12. Click on 'OK'.	
9	Set the 'Collapse dataset to gene symbols' field to 'false'.	
10	Set 'Permutation' type to 'gene-set'.	
11	Leave the 'Chip platform(s)' empty.	
12	In 'Basic Fields', choose an informative name for your analysis in the 'Analysis name field'. Tip: name of the comparison that you are making and date (e.g ES12_versus_NT12_date).	
13	Set the 'Metric for ranking genes' to 'tTest'.	
14	'In the Save results in this folder', click on the 3 dots [] and browse your computer to select a folder.	
15	Locate and click the ' Run' button at the bottom right corner of the window. Tip: you may need to expand the window to see the Run button. Note: it takes 5 min to run using a maximum of 1.4Gb of memory. GSEA has finished to run when a message ' Success 5 ' appears in the Status field of the GSEA reports box.	
16	In the ' GSEA reports' box, click on ' Success 5' to see the results. Open the links and explore the results.	

Steps 1-2



File Options Downloads Tools Help Supported files markets Iterations Iterations Supported files formats Load data: Method 3: Supported files formats Dataset: res vet tools Supported files formats Iterations Dataset: Supported files formats Dataset: Part (Stando) Ter over Stando Iterations Iterations Supported file formats Dataset: Part (Stando) Supported file formats Ter over Stando Iterations Iterations Supported file formats Ter over Stando Ter over Stando Ter over Stando Ter over Stando Sta	•••	GSEA v2.0.14 (Gene set enrichme	nt analysis Broad Institut	te)	
Steps in GSEA analysis 4 b # Ibone QL Load data in the application 4 b # Ibone QL Load data in the application Supported file formats Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Method 3: drag and drop files here Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application Ibone QL Load data in the application in the application <td< td=""><td>File Options Downloads Tools Help</td><td>-</td><td></td><td></td><td></td></td<>	File Options Downloads Tools Help	-			
Load data Image: Call cast Image: Cast cast Image:	Steps in GSEA analysis	Home Do Load data ×			
Wethod 1: Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Cene set tools Stanford Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 2: Stanford Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 2: Stanford Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 2: Stanford Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 2: Stanford Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 2: Stanford Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 2: Method 2: Method 2: Dubase: Feer or get (Broad/MIT), pel (Stanford) Wethod 3: Torone MisigoB Clear Itad these files! Othor on file formus For core captions) Recently used files (double click to load, right click for Purge) Dibjects in memory jubit-click to expand all Name Stans Stans Stans Objects in memory jubit-click to expand all Method 2: Stans Stans Stans Stans Method 1: Stans Method 1: Stans Stans Stans<	Load data	Load data: Import data into the application		6 1 1	
Image: Status	Run GSEA	Method 1:	Method 3: drag and drop	files here	Supported file formats Dataset: <i>res</i> or <i>gct</i> (Broad/MIT), <i>pcl</i> (Stanford) <i>txt</i> (tab-delim text)
Cene set tools Chp2Chip mapping Towse MsigDB Clear Load these files! Object cache Object saleady loaded & ready for use, right click for purge Processes: click 'status' field for results Status Status Show results folder 9:05:36 AN Tet	eading edge analysis	Method 2: S Load last dataset used			Phenotype labels: <i>CIS</i> Gene sets: <i>gmx</i> or <i>gmt</i>
Stock Status	Gene set tools				
Provese MSigDB Clear Clear Clear Clear Object cache (b) More on file formats Porcesses: (c) More on file formats Porcesses: (c) More on file formats Porcesses: (c) More on file formats (c) More on file formats <th></th> <th></th> <th></th> <th></th> <th></th>					
Clear Load these files! Object cache (objects already loaded & ready for use, right click for more options) CSEA reports Processes: click 'status' field for results Objects in memory jubif-click to expand all) Name Istatus Objects in memory jubif-click to expand all) Show results folder 9:05:36 AM Clear Image: Clear	W Browse MSigDB				
SEA reports Processes: click 'status' field for results Name Status Status Status Show results folder Object solution	Analysis history		Clear	Load these files!	⑦ More on file formats
CSEA reports CSEA reports CSEA reports Name Status Status Show results folder 9:05:36 AM		Recently used files (double click to load, right more options)	t click for Purge	Object cache (object cache) for more options)	ects already loaded & ready for use, right click
Processes: click 'status' field for results Name Status Show results folder 9:05:36 AM	CSEA reports			Dbjects in memo	ory [shift-click to expand all]
Show results folder 9:05:36 AM 健	Processes: click 'status' field for results				
Show results folder 9:05:36 AM	Hame Status				
Show results folder Image: Control of the second sec					
Show results folder Image: Control of the second seco					
Show results folder 9:05:36 AM 🕀 👔 30M of 45M : ::					
Show results folder Image: Control of the second sec					
Show results folder Image: The second					
Show results folder Image: Control of the					
Show results folder 9:05:36 AM 🕀 👔 30M of 45M : ::					
Show results folder Image: Control of the second sec					
Show results folder Image: Control of the second sec					
Show results folder Image: Control of the state of the					
9:05:36 AM 🗈 👔	Show results folder				
	9:05:36 AM				☐ 30M of 45M .::

Step 3 (continued)

	iport data into the ap	prication	Method 3	drag and drop f	iles here	Supported file fo
Method 1:-	Browse for files			drag and drop n		Dataset: <i>res</i> or <u>i</u> <i>pcl</i> (S <i>txt</i> (t Phenotype lab Gene sets: <i>gm</i>
		Oper	1			
	GS	EATutorial		*		
Recent folders (de	ouble click to list cor	itent)		Name A Human_GO MCF7_Expr MCF7_grou	Date Modified Monday, April Monday, April Monday, April	() M
						ects already loaded ory (shift-click to expand abases
	File Format:	All Files		*		
				Cancel	Choose	

Step 3 (continued)



	GSEA v2.0.14 (Gene set enrichme	nt analysis Broad Institut	te)	
File Options Downloads Tools Help				
Steps in GSEA analysis	Home 🛛 💭 Load data 🗙			4 ▷ ■
Load data	Load data: Import data into the application			
	Method 1:	Method 3: drag and drop	files here	Supported file formats
Run GSEA	► Browse for files			Dataset: res or gct (Broad/MIT), pcl (Stanford) txt (tab-delim text)
Leading edge analysis	Method 2: S Load last dataset used			Phenotype labels: <i>cls</i> Gene sets: <i>gmx</i> or <i>gmt</i>
Gene set tools				
Chip2Chip mapping				
Browse MSigDB				
Analysis history		Clear	Load these files!	(?) More on file formats
	Recently used files (double click to load, right more options)	t click for Purge	Object cache (object cache) for more options)	ects already loaded & ready for use, right click
GSEA reports	./GSEATutorial/MCF7_groups.cls ./GSEATutorial/MCF7_Expression_matrix.gct ./GSEATutorial/MCF7_Expression_matrix.gct		▼ 📄 Objects in memo ▶ 📄 Gene set data	ry [shift-click to expand all] abases
Processes: click 'status' field for results Name Status	 ./HALT03_jan23_2014/bayseqNR_uRANK.rnk ./HALT03_jan13_2014/Bayseq_R_jan17_u_RANK.rr 	ik	 Datasets Phenotypes 	
	 ./HALT03_jan23_2014/treated_uRANK.rnk ./AS02 Dec30 2013/LSC rank symbol.rnk 			
	./rank_files/CD8_NICplus_vs_CD8_NICmin_u_rankU	IPDATED.mk		
	/rank_files/CD4_NIC+_vs_CD4_NICmin_u_rankUPD /rank_files/DP_NICplus_vs_DP_NICminus_u_rankU	ATED.rnk		
	 /GSEA/Treated9days_Con9days.rnk 	DATEDANK		
	./GSEA/Treated48hrs_Con0hr.rnk			
	./GSEA/Treated3weeks_Con3weeks.mk			
	./KH01/OE_rank.rnk			
	JCG04_RBPJ_bindingsites_Dec2_2014/CD4_NIC+	_vs_CD4_NICmin_u_rank.		
	AHR targets april27/OF rank rnk			
	 /AHR_targets_april27/KD_rank.rnk 			
Show results folder				
9:28:50 AM 🔄 4730 [INFO] Loading 3 files Hur	man_GO_AllPathways.gmt MCF7_Expression_matrix.e	ct MCF7_groups.cls Files	loaded successfully: 3	/ 3 There were NO err 1 122M of 217M .:

•••	GSEA v2.0.14 (Gene set enric	chment analysis Broad Institute)	
File Options Downloads Tools Help			
Steps in GSEA analysis	Home Load data × Run Gsea	1 X	
Load data	Gsea: Set parameters and run enrichme	nt tests	
8003	Required fields		
Constant SEA	Expression dataset	MCF7_Expression_matrix [20326x18 (ann: 20326,18,chip na)]	\$
Leading edge analysis	Gene sets database		
	Number of permutations	1000	•
Gene set tools	Phenotype labels		
Chip2Chip mapping	Collapse dataset to gene symbols	true	\$
Browse MSigDB	Permutation type	phenotype	\$
	Chip platform(s)		
Analysis history	Basic fields		Show
GSEA reports Processes: click 'status' field for results Name Status	Advanced fields		Show



Step 6 (continued)

File Ontions Downloads Tools Help	GOEA VZ.U	7. 14 (Gene set enrichment ana	iysis bioau institute)			
Steps in GSEA analysis	Home No I Home	x Bun Gsea x				4 10 10
	Gsea: Set parameters	and run enrichment tests				
Run GSEA	Required fields Expression dataset	MCF7_	Expression_matrix [2032]	6x18 (ann: 20326,18,cl	nip na)] 🗘	
Cading edge analysis	Gene sets database Number of permutat	ions 1000				
Gene set tools	Phenotype labels					
Chip2Chip mapping		Select one or more ge	ne sets(s)		O	
💓 Browse MSigDB	Pe Gene m	atrix (local gmx/gmt)	Subsets lext	entry	\$	
17	Cł Human_GO_AllPat	thways.gmt [17259 gene sets]				
Analysis history					Chau	
	Ва				Snow	
CSEA reports Processes: click 'status' field for results Name Status	Ac				Show	
	(?) Help		Cancel	ОК		
Show results folder	0	💛 Reset	🔞 Last	🛱 Command	Low (cou usage)	► F
9:45:58 AM 🕀 4730 [INFO] Loading 3 files Hur	man_GO_AllPathways.gmt MC	F7 Expression matrix.gct MC	F7 groups.cls Files load	led successfully: 3 / 3	There were NO err 125	M of 217M

	GSEA v2.0.14 (Gene set enri	chment analysis Broad Institute)		
File Options Downloads Tools Help				
Steps in GSEA analysis	Home 🗓 Load data 🛛 🗙 🛛 🔛 Run Gse	a ×		⊲ Þ ⊞
Load data	Gsea: Set parameters and run enrichme	ent tests		
	Paguirad fields			
Run GSEA	Required netus			
(and	Expression dataset	MCF7_Expression_matrix [20326x18 (ann: 20326,18,	:hip na)] 🗘	
Leading adge analysis	Gene sets database	DRAIab/GSEA_tutorial/GSEATutorial/Human_GO_AllPat	hways.gmt	
o o Leaung euge analysis	Number of permutations	100	T	
Cane set tools				
	Phenotype labels			
	Collapse dataset to gene symbols	true	\$	
20.	Permutation type	phenotype	\$	
Browse MSigDB				
	Chip platform(s)			
Analysis history				
	Basic fields		Show	
GSEA reports				
Processes: click 'status' field for results	Advanced fields		Show	
Name Status	Advanced fields		511011	
Show results folder	(?) 💛 Reset	≫ Last 🖻 Command	Low (cpu usage) 💠	► R
9:50:10 AM 🛛 🕀 4730 [INFO] Loading 3 files Hum	nan_GO_AllPathways.gmt MCF7_Expression_m	atrix.gct MCF7_groups.cls Files loaded successfully: 3 / 3	There were NO err 134M of	217M .::

•••	Screen	Shot 2015-05-17 at 9.54.14 AM		
				Q Search
🛄 ~ 🦄 🥒 🛛 🕞 ~ T 🖉 ~ 🖆 eps in GSEA analysis	N □ Ξ ~ □ ~ □ ~ A ~ Home ΩΣ Load data ×	n Gsea 🗙		4 1
ڬ Load data	Gsea: Set parameters and run enr	richment tests		
Run GSEA	Required fields Expression dataset	MCF7_Expression_matrix [20326x	18 (ann: 20326,18,chip na)] 🗘	
Generation and the second seco	Gene sets database Number of permutations	3RAlab/GSEA_tutorial/GSEATutorial	/Human_GO_AllPathways.gmt	
ine set tools	Phenotype labels			
Chip2Chip mapping	G • •	Select a phenotype		
Browse MSigDB	Pe Select source file MCF2_groups.cls [18 sample MCF2_groups.cls [18 sample REST_versus_NT48 ES12_versus_NT48 ES12_versus_ES24 ES12_versus_NT24 ES12_versus_S148 ES12_versus_ES48 ES12_versus_ES48 ES12_versus_ES24 Ac Options Ac Options Create	les=>6 classes] henotypes from all source files e an on-the-fly phenotype		Show
	() Help	a gene as the phenotype		
Show results folder	? Contract Contrac	😚 Last	Command Low (cou us	sage)
4:04 AM 🕀 4730 [INFO] Loading 3 files 🕒	luman_GO_AllPathways.gmt MCF7_Expressi	on_matrix.gct MCF7_groups.cls Files loaded	successfully: 3 / 3 There were NO e	err 📋 144M of 217M .::

Steps 9-15

	GSEA v2.0.14	(Gene set enrichment analysis Broad Institute)	
File Options Downloads Tools	Help		
Steps in GSEA analysis	Home 💯 Load data 🗙 🔣 Run Gse	a ×	4 Þ 🗉
Load data	Gsea: Set parameters and run enrichme	ent tests	
	Required fields		
C • J Run GSEA	Required fields		
Eug	Expression dataset	MCF7_Expression_matrix [20326x18 (ann: 20326,18,chip na)]	
	Gene sets database	DRAIab/GSEA_tutorial/GSEATutorial/Human_GO_AllPathways.gmt	
Leading edge analysis	Number of commencesting	100	
	Number of permutations	100	
Gene set tools	Phenotype labels	SEA_tutorial/GSEATutorial/MCF7_groups.cls#ES12_versus_NT12	
Chip2Chip mapping	Collapse dataset to gene symbols	false \$	
	Permutation type	gene_set \$	
	Chip platform(s)		
			1114
	Basic fields		Hide
	Analysis name	ES12_vs_NT12_CBW	
GSEA reports	Enrichment statistic	weighted \$	
Processes: click 'status' field for results	Metric for ranking genes	tTest \$	
Name Status	Gene list sorting mode	real \$	
2 🛛 Gsea Running	Gene list ordering mode	descending \$	
	Max size: exclude larger sets	۶00 💌	
	Min size: exclude smaller sets	15	
	Save results in this folder	/Users/veroniquevoisin/Downloads	
	Advanced fields		Show Click on Run
			+
Show results folder	⑦	🖇 Last 🖹 Command 🛛 Low (cpu usage	a) 🗘 🕨 Run
10:43:23 AM 🛛 9690 [INFO] Done pre	proc for smaller than: 15		627M o f 1281M .::

	GSEA v2.0.14	(Gene set enrichment analysis Broad Institute)	
File Options Downloads Tools	Help		
Steps in GSEA analysis	Home 🔛 Load data 🗙 💽 Run Gse	a ×	4 Þ E
Load data	Gsea: Set parameters and run enrichme	ent tests	
	Required fields		
Run GSEA	Expression dataset	MCE7 Expression matrix [20226x18 (app: 20226.18 chip.pa)]	•
	Expression dataset	Mer /_expression_matrix [20520x18 (ann. 20520,18,emp ha)]	•
Leading edge analysis	Gene sets database	DRAIab/GSEA_tutorial/GSEATutorial/Human_GO_AllPathways.gmt	
00	Number of permutations	100	-
Gene set tools	Dhamatana labata	TERA ANALYSI (CERATING IN MORE ANALYSI ANA	5
Chin2Chin manning	Phenotype labels	SEA_tutonai/GSEATutonai/MCF7_groups.cls#ES12_versus_NT12	
	Collapse dataset to gene symbols	false	÷
	Permutation type	gene_set	•
Browse MSigDB	Chin platform(s)		- II
	cinp platorin(s)		
Analysis history			
	Basic fields		Hide
	Analysis name	ES12 ve NT12 CRW	
	Analysis name		
GSEA reports	Enrichment statistic	weighted	÷
Processes: click 'status' field for	Metric for ranking genes	tTest	\$
Name Status	Gene list sorting mode	real	\$
1 Gsea Success 5			
	Gene list ordering mode	descending	•
	Max size: exclude larger sets	500	•
	Min size: exclude smaller sets	15	-
		(III) and the second state (Denoralized state)	
	Save results in this folder	/Users/veroniquevoisin/Downloads	
	Advanced fields		Show
Show results folder	⑦	🖇 Last 📄 Command 🛛 Low (cp	u usage) 🗘 🕨 Run
10:24:12 AM	m unigene / gene symbol: 38870		570M of 1281M .::

