

# Differential expression and downstream analysis

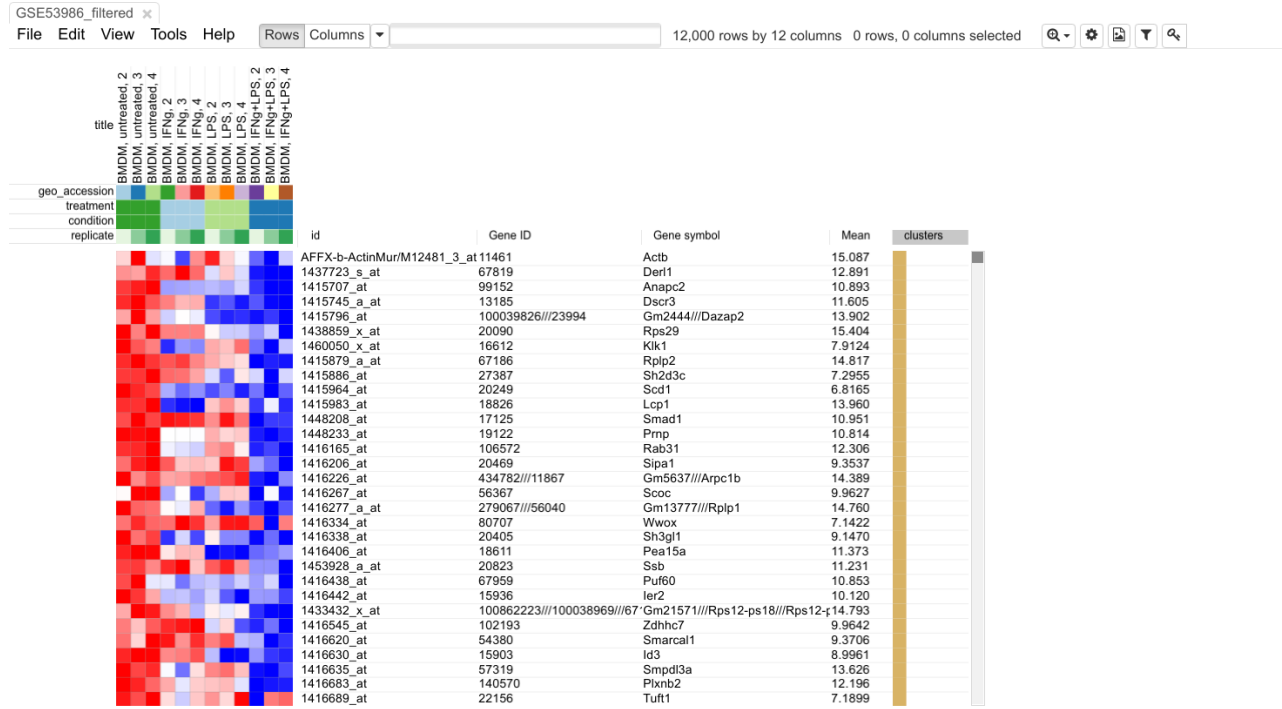
Alexey Sergushichev

Sep 23, Nice

# Outline

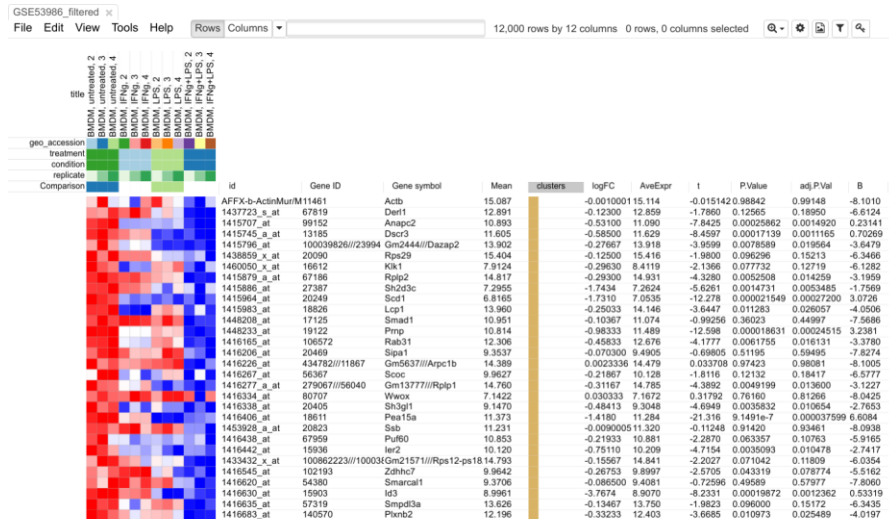
- ✓ Exploring gene expression datasets
- ✓ **Simple analysis methods**
- ✓ Working with public datasets

# Open the filtered dataset from gct (or continue working in the open tab)



# Differential expression

## Tools/Differential expression/limma



limma

Field

strain  
tissue  
title  
treatment

Field <-  
treatment

Class A

Search

1/4

IFNg

IFNg\*LPS

LPS

Untreated

untreated vs LPS

Class B

Search

1/4

IFNg

IFNg\*LPS

LPS

Untreated

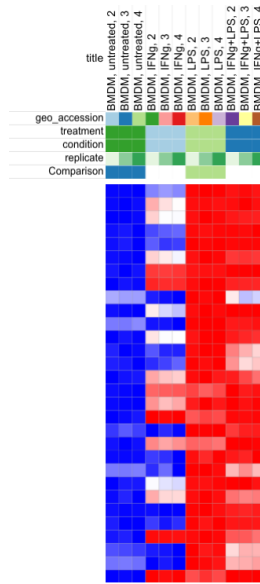
OK Cancel

# Differential expression: results

- ✔ Sort decreasing by “t”
- ✔ Save the results as GSE53986\_Ctrl\_vs\_LPS.gct

GSE53986\_filtered x

File Edit View Tools Help Rows Columns 12,000 rows by 12 columns 0 rows, 0 columns selected

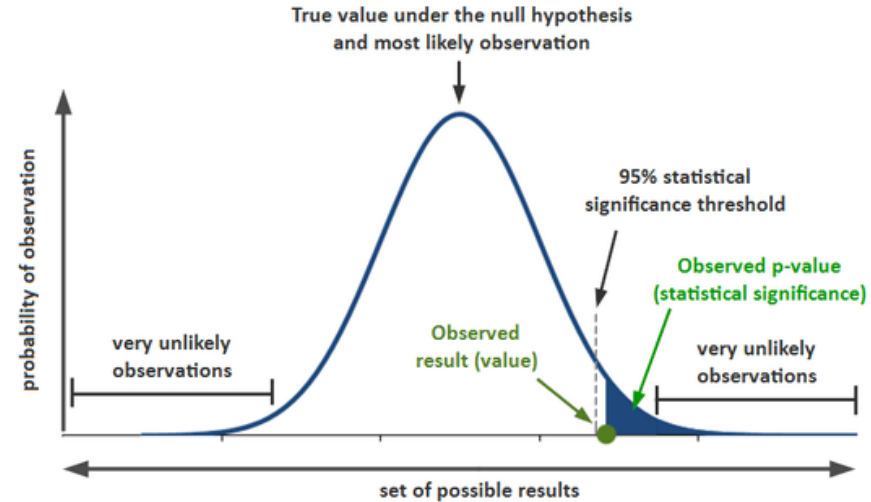


id	Gene ID	Gene symbol	Mean	clusters	logFC	AveExpr	t	PValue	adj.PVal	B
1433741_at	12494	Cd38	10.683	7.8141	10.100	114.93	4.9996e-11	5.9995e-7	14.528	
1431008_at	110557//f68395//i15-H2-Q6//LOC68395//H2.12.612		12.860	6.3787	11.413	87.173	2.5125e-10	0.00001507	13.785	
1431591_s_at	100038882//677161sg15//Gm9706		12.448	5.4521	11.996	73.099	7.0251e-10	0.00002008	13.178	
1450826_a_at	20210	Saa3	12.345	7.3527	11.337	72.131	7.5933e-10	0.00002008	13.128	
1422953_at	14289	Fpr2	11.448	7.6465	10.681	66.749	1.1941e-9	0.00002008	12.824	
1418191_at	24110	Usp18	12.462	5.2880	11.628	66.467	1.2239e-9	0.00002008	12.807	
1434372_at	107350	AW112010	13.511	5.5861	12.114	63.393	1.6137e-9	0.00002008	12.611	
1429947_a_at	58203	Zbp1	12.057	5.2546	10.727	62.988	1.6752e-9	0.00002008	12.584	
1419082_at	18788	Serpinb2	8.8510	6.5854	10.980	62.740	1.7142e-9	0.00002008	12.567	
1436058_at	58185	Rsad2	13.219	4.6917	12.566	62.580	1.7399e-9	0.00002008	12.556	
1420330_at	56119	Clec4e	11.967	5.6553	12.191	61.975	1.8414e-9	0.00002008	12.514	
1460783_at	15957	Ifi1	12.453	5.6055	11.616	56.206	3.2566e-9	0.00003256	12.081	
1427102_at	20558	Sfn4	10.142	8.1172	10.353	52.814	4.6823e-9	0.00004322	11.789	
1445639_at	215772	Adbg	6.8997	5.6457	7.2081	50.076	6.3874e-9	0.00004933	11.531	
1437636_at	623121	Pydc4	10.163	5.5406	9.2754	49.497	6.8353e-9	0.00004933	11.474	
1449453_at	12182	Bst1	11.326	5.9150	9.6841	49.308	6.9895e-9	0.00004933	11.455	
1438037_at	67138	Herc6	12.034	3.9940	11.248	48.600	7.6045e-9	0.00005069	11.383	
1418240_at	14469	Gbp2	13.506	5.0928	11.807	47.493	8.6985e-9	0.00005167	11.267	
1449399_a_at	16176	Il1b	10.929	7.3050	11.155	46.634	9.6751e-9	0.00005167	11.174	
1419714_at	60533	Cd274	12.956	3.8129	11.852	45.819	1.0723e-8	0.00005167	11.084	
1436555_at	11988	Slc7a2	10.437	5.4902	10.918	45.816	1.0727e-8	0.00005167	11.083	
1418946_at	20442	Sl3gal1	8.5020	2.9640	8.5960	45.775	1.0782e-8	0.00005167	11.079	
1423853_at	14081	Acsf1	13.547	3.7097	13.171	45.514	1.1149e-8	0.00005167	11.049	
1449025_at	15959	Ifi3	12.703	5.2835	11.977	44.782	1.2253e-8	0.00005167	11.065	
1448610_a_at	20656	Sod2	13.702	2.8793	13.670	44.668	1.2437e-8	0.00005167	11.052	
1423289_a_at	66282	Tma16	10.838	4.8042	10.828	44.582	1.2577e-8	0.00005167	11.042	
1453304_s_at	17069	Ly6e	13.675	3.9013	12.676	44.582	1.2577e-8	0.00005167	11.042	
1434786_at	329251	Ppp1r12b	8.6192	4.2090	9.4095	44.377	1.2920e-8	0.00005167	11.018	
1435040_at	73914	Irak3	10.703	3.3803	10.958	44.121	1.3363e-8	0.00005172	10.897	
1417292_at	15953	Ifi47	13.380	3.7227	12.109	43.556	1.4406e-8	0.00005337	10.820	

Click to sort

# P-value

- ✓ P-value - the probability that, when the **null hypothesis** is true, the statistical summary (such as the sample mean difference between two groups) would be equal to, or **more extreme** than, the actual observed results.



# Differential expression p-value

- ✔ Null hypothesis: average gene expression is distributed in two conditions **and** the distributions follows a certain model
- ✔ Summary: t-statistic like
- ✔ When p-value is low, then we **reject** null hypothesis and assume genes are differentially expressed

# Multiple hypothesis testing

- ✔ P-values are uniformly distributed when null hypothesis is true
- ✔ If there are no real differential expression at all, for  $10^4$  genes there will be P-values as low as  $10^{-4}$
- ✔ Using Benjamini-Hochberg adjustment procedure which limits False Discovery Rate (FDR)



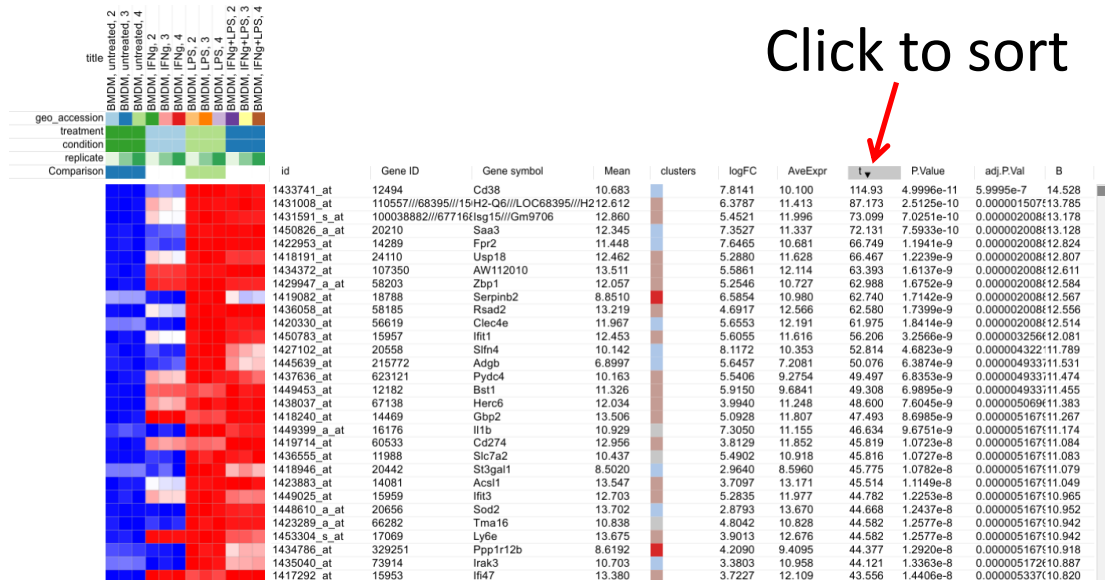


# Differential expression: results

- ✔ Sort decreasing by “t”
- ✔ Save the results as GSE53986\_Ctrl\_vs\_LPS.gct

GSE53986\_filtered x

File Edit View Tools Help Rows Columns 12,000 rows by 12 columns 0 rows, 0 columns selected



id	Gene ID	Gene symbol	Mean	clusters	logFC	AveExpr	t	PValue	adj.PVal	B
1433741_at	12494	Cd38	10.683	7.8141	10.100	114.93	4.9996e-11	5.9995e-7	14.528	
1431008_at	110557//f68395//i15-H2-Q6//LOC68395//H2.12.612		12.860	6.3787	11.413	87.173	2.5125e-10	0.00001507	13.785	
1431591_s_at	100038882//677161sg15//Gm9706		12.860	5.4521	11.996	73.099	7.0251e-10	0.00002008	13.178	
1450826_a_at	20210	Saa3	12.345	7.3527	11.337	72.131	7.5933e-10	0.00002008	13.128	
1422953_at	14289	Fpr2	11.448	7.6465	10.681	66.749	1.1941e-9	0.00002008	12.824	
1418191_at	24110	Usp18	12.462	5.2880	11.628	66.467	1.2239e-9	0.00002008	12.807	
1434372_at	107350	AW112010	13.511	5.5861	12.114	63.393	1.6137e-9	0.00002008	12.611	
1429947_a_at	58203	Zbp1	12.057	5.2546	10.727	62.988	1.6752e-9	0.00002008	12.584	
1419082_at	18788	Serpinb2	8.8510	6.5854	10.980	62.740	1.7142e-9	0.00002008	12.567	
1436058_at	58185	Rsad2	13.219	4.6917	12.566	62.580	1.7399e-9	0.00002008	12.556	
1420330_at	56119	Clec4e	11.967	5.6553	12.191	61.975	1.8414e-9	0.00002008	12.514	
1460783_at	15957	Ifi1	12.453	5.6055	11.616	56.206	3.2566e-9	0.00003256	12.081	
1427102_at	20558	Sfnf4	10.142	8.1172	10.353	52.814	4.6823e-9	0.00004322	11.789	
1445639_at	215772	Adbg	6.8997	5.6457	7.2081	50.076	6.3874e-9	0.00004933	11.531	
1437636_at	623121	Pydc4	10.163	5.5406	9.2754	49.497	6.8353e-9	0.00004933	11.474	
1449453_at	12182	Bst1	11.326	5.9150	9.6841	49.308	6.9895e-9	0.00004933	11.455	
1438037_at	67138	Herc6	12.034	3.9940	11.248	48.600	7.6045e-9	0.00005069	11.383	
1418240_at	14469	Gbp2	13.506	5.0928	11.807	47.493	8.6985e-9	0.00005167	11.267	
1449399_a_at	16176	Il1b	10.929	7.3050	11.155	46.634	9.6751e-9	0.00005167	11.174	
1419714_at	60533	Cd274	12.956	3.8129	11.852	45.819	1.0723e-8	0.00005167	11.084	
1436555_at	11988	Slc7a2	10.437	5.4902	10.918	45.816	1.0727e-8	0.00005167	11.083	
1418946_at	20442	Sl3gal1	8.5020	2.9640	8.5960	45.775	1.0782e-8	0.00005167	11.079	
1423853_at	14081	Acsf1	13.547	3.7097	13.171	45.514	1.1149e-8	0.00005167	11.049	
1449025_at	15959	Ifi3	12.703	5.2835	11.977	44.782	1.2253e-8	0.00005167	11.065	
1448610_a_at	20656	Sod2	13.702	2.8793	13.670	44.668	1.2437e-8	0.00005167	11.052	
1423289_a_at	66282	Tma16	10.838	4.8042	10.828	44.582	1.2577e-8	0.00005167	11.042	
1453304_s_at	17069	Ly6e	13.675	3.9013	12.676	44.582	1.2577e-8	0.00005167	11.042	
1434786_at	329251	Ppp1r12b	8.6192	4.2090	9.4095	44.377	1.2920e-8	0.00005167	11.018	
1435040_at	73914	Irk3	10.703	3.3803	10.958	44.121	1.3363e-8	0.00005172	10.897	
1417292_at	15953	Ifi47	13.380	3.7227	12.109	43.556	1.4406e-8	0.00005337	10.820	

Click to sort

# Sharing the results

- ✔ File/Get link to a dataset
- ✔ Open the link in another browser tab

Get link to a dataset ×

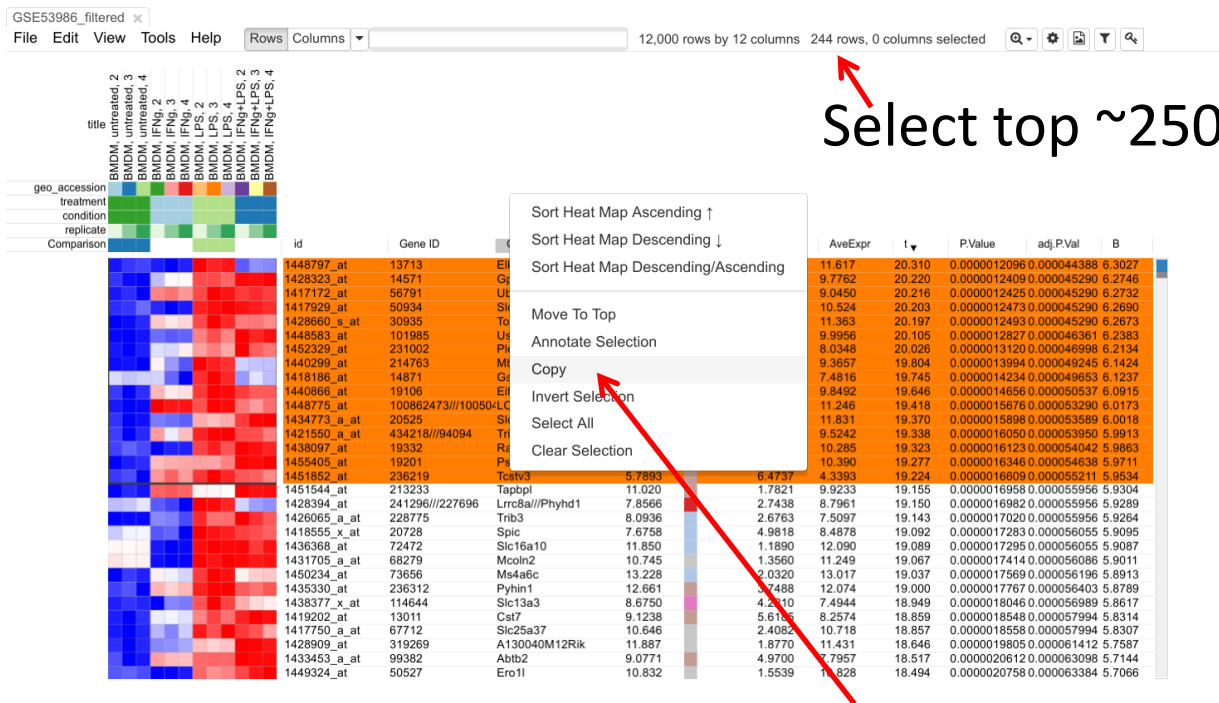
---

Link `https://genome.ifmo.ru/phantasus/?session=x08f4e3e79`

Please note that link will be valid for 30 days.

---

# Pathway enrichment: selecting signature genes



Select top ~250 genes

Right click on "Gene symbol" column and click "Copy" (not Ctrl-C)

# Pathway enrichment: using MSigDB tool

- ✓ Go to <http://software.broadinstitute.org/gsea/msigdb/annotate.jsp>

**GSEA**  
Gene Set Enrichment Analysis

[GSEA Home](#) [Downloads](#) [Molecular Signatures Database](#) [Documentation](#) [Contact](#)

## Login to GSEA/MSigDB

**Login**

Click [here](#) to register to view the MSigDB gene sets and/or download the GSEA software. This helps us track and better serve our user community.

If you have already registered for GSEA or MSigDB please enter your registration email address below.

Items marked with \* are required.

**Email: \***

Put [alsergbox@gmail.com](mailto:alsergbox@gmail.com) here for now  
Register later, it's free!

# Pathway enrichment: using MSigDB tool

Paste genes here

**Gene Identifiers**

Cd38  
H2-  
Q6///LOC68395///H  
2-Q8  
Isg15///Gm9706  
Saa3  
Fpr2  
Usp18  
AW112010  
Zbp1  
Serpib2  
Rsad2  
Clec4e  
Ifit1  
Slfn4  
Adgb  
Pydc4  
Bst1  
Herc6  
Gbp2  
Il1b  
Cd274  
Slc7a2  
St3gal1  
Acs11  
Ifit3  
Sod2

**Compute Overlaps**

- H: hallmark gene sets ?
- C1: positional gene sets ?
- C2: curated gene sets ?
- CGP: chemical and genetic perturbations ?
- CP: Canonical pathways ?
- CP:BIOCARTA: BioCarta gene sets ?
- CP:KEGG: KEGG gene sets ?
- CP:REACTOME: Reactome gene sets ?
- C3: motif gene sets ?
- MIR: microRNA targets ?
- TFT: transcription factor targets ?
- C4: computational gene sets ?
- CGN: cancer gene neighborhoods ?
- CM: cancer modules ?
- C5: GO gene sets ?
- BP: GO biological process ?
- CC: GO cellular component ?
- MF: GO molecular function ?
- C6: oncogenic signatures ?
- C7: immunologic signatures ?

show top 10 genesets

with FDR q-value below

**Compendia expression profiles**













- Human tissue compendium (Novartis)
- NCI-60 cell lines (National Cancer Institute)

**Gene families**

Check "hallmark gene sets" and "Canonical pathways"

















Click "compute overlaps"

# Pathway enrichment results

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value 	FDR q-value 
<a href="#">HALLMARK_INTERFERON_GAMMA_RESPONSE</a> [200]	Genes up-regulated in response to IFNG [GeneID=3458].	67		5.51 e <sup>-112</sup>	7.6 e <sup>-109</sup>
<a href="#">HALLMARK_INTERFERON_ALPHA_RESPONSE</a> [97]	Genes up-regulated in response to alpha interferon proteins.	34		6.43 e <sup>-57</sup>	4.43 e <sup>-54</sup>
<a href="#">REACTOME_IMMUNE_SYSTEM</a> [933]	Genes involved in Immune System	49		2.51 e <sup>-39</sup>	1.15 e <sup>-36</sup>
<a href="#">HALLMARK_INFLAMMATORY_RESPONSE</a> [200]	Genes defining inflammatory response.	30		7.61 e <sup>-38</sup>	2.63 e <sup>-35</sup>
<a href="#">HALLMARK_TNFA_SIGNALING_VIA_NFKB</a> [200]	Genes regulated by NF-kB in response to TNF [GeneID=7124].	29		3.76 e <sup>-36</sup>	1.04 e <sup>-33</sup>
<a href="#">REACTOME_INTERFERON_SIGNALING</a> [159]	Genes involved in Interferon Signaling	26		6.39 e <sup>-34</sup>	1.47 e <sup>-31</sup>
<a href="#">REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM</a> [270]	Genes involved in Cytokine Signaling in Immune system	30		9.04 e <sup>-34</sup>	1.78 e <sup>-31</sup>
<a href="#">REACTOME_INTERFERON_ALPHA_BETA_SIGNALING</a> [64]	Genes involved in Interferon alpha/beta signaling	17		1.85 e <sup>-26</sup>	3.19 e <sup>-24</sup>
<a href="#">HALLMARK_ALLOGRAFT_REJECTION</a> [200]	Genes up-regulated during transplant rejection.	18		5.83 e <sup>-19</sup>	8.78 e <sup>-17</sup>
<a href="#">REACTOME_INTERFERON_GAMMA_SIGNALING</a> [63]	Genes involved in Interferon gamma signaling	13		6.37 e <sup>-19</sup>	8.78 e <sup>-17</sup>

# gene/pathway matrix

Gene/geneset overlap matrix

Entrez Gene Id	Gene Symbol	HALLMARK_INTERFERON_GAMMA_RESPONSE	HALLMARK_INTERFERON_ALPHA_RESPONSE	REACTOME_IMMUNE_SYSTEM	HALLMARK_INFLAMMATORY_RESPONSE	HALLMARK_TNFA_SIGNALING_VIA_NFKB	REACTOME_INTERFERON_SIGNALING	REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM	REACTOME_INTERFERON_ALPHA_BETA_SIGNALING	HALLMARK_ALLOGRAFT_REJECTION	REACTOME_INTERFERON_GAMMA_SIGNALING	Entrez	Source	Gene Description
3665	IRF7												S	interferon regulatory factor 7
5610	EIF2AK2												S	eukaryotic translation initiation factor 2-alpha kinase 2
3433	IFIT2												S	interferon-induced protein with tetratricopeptide repeats 2
6890	TAP1												S	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
64135	IFIH1												S	interferon induced with helicase C domain 1
8638	OASL												S	2'-5'-oligoadenylate synthetase-like
6773	STAT2												S	signal transducer and activator of transcription 2, 113kDa
3669	ISG20												S	interferon stimulated exonuclease gene 20kDa
5696	PSMB8												S	proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)
11274	USP18												S	ubiquitin specific peptidase 18
4599	MX1												S	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)
3430	IFI35												S	interferon-induced protein 35
3437	IFIT3												S	interferon-induced protein with tetratricopeptide repeats 3
9246	UBE2L6												S	ubiquitin-conjugating enzyme E2L 6
5698	PSMB9												S	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)
79132	DHX58												S	DEXH (Asp-Glu-X-His) box polypeptide 58

# How p-value is calculated

- ▶ Investigate Gene Sets
- ▶ View Gene Families
- ▶ Help






Converted 250 submitted identifiers into 196 entrez genes. [click here for details.](#)

Collections	# Overlaps Shown	# Gene Sets in Collections	# Genes in Comparison (n)	# Genes in Universe (N)
CP	10	1329	196	45956

Click the gene set name to see the gene set page. Click the number of genes [in brackets] to download the list of genes.

Color bar shading from light green to black, where lighter colors indicate more significant FDR q-values (< 0.05) and black indicates less significant FDR q-values (>= 0.05).

Save to: [Excel](#) | [GenomeSpace](#)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
<a href="#">REACTOME_IMMUNE_SYSTEM [933]</a>	Genes involved in Immune System	50		5.65 e-40	7.51 e-37
<a href="#">REACTOME_INTERFERON_SIGNALING [159]</a>	Genes involved in Interferon Signaling	26		1.29 e-33	8.6 e-31
<a href="#">REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM [270]</a>	Genes involved in Cytokine Signaling in Immune system	30		2.04 e-33	9.04 e-31
<a href="#">REACTOME_INTERFERON_ALPHA_BETA_SIGNALING [64]</a>	Genes involved in Interferon alpha/beta signaling	17		2.91 e-26	9.67 e-24
<a href="#">REACTOME_INTERFERON_GAMMA_SIGNALING [63]</a>	Genes involved in Interferon gamma signaling	13		1.0 e-26	1.0 e-24

Fisher test table

Expected overlap  
if independent

	query	non-query
pathway	17	196-17
non-pathway	64-17	45956-64-196+17

$$196 * 64 / 45956 = 0.27$$



# Notice: MSigDB p-values are biased

- ▶ Investigate Gene Sets
- ▶ View Gene Families
- ▶ Help






Converted 250 submitted identifiers into 196 entrez genes. [click here for details.](#)

Collections	# Overlaps Shown	# Gene Sets in Collections	# Genes in Comparison (n)	# Genes in Universe (N)
CP	10	1329	196	45956

Click the gene set name to see the gene set page. Click the number of genes [in brackets] to download the list of genes.

Color bar shading from light green to black, where lighter colors indicate more significant FDR q-values ( $< 0.05$ ) and black indicates less significant FDR q-values ( $\geq 0.05$ ).

Save to: [Excel](#) | [GenomeSpace](#)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
<a href="#">REACTOME_IMMUNE_SYSTEM [933]</a>	Genes involved in Immune System	50		$5.65 \times 10^{-40}$	$7.51 \times 10^{-37}$
<a href="#">REACTOME_INTERFERON_SIGNALING [159]</a>	Genes involved in Interferon Signaling	26		$1.29 \times 10^{-33}$	$8.6 \times 10^{-31}$
<a href="#">REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM [270]</a>	Genes involved in Cytokine Signaling in Immune system	30		$2.04 \times 10^{-33}$	$9.04 \times 10^{-31}$
<a href="#">REACTOME_INTERFERON_ALPHA_BETA_SIGNALING [64]</a>	Genes involved in Interferon alpha/beta signaling	17		$2.91 \times 10^{-26}$	$9.67 \times 10^{-24}$
<a href="#">REACTOME_INTERFERON_GAMMA_SIGNALING [63]</a>	Genes involved in Interferon gamma signaling	13		$1.0 \times 10^{-20}$	$1.0 \times 10^{-18}$

our N=12000

Fisher test table

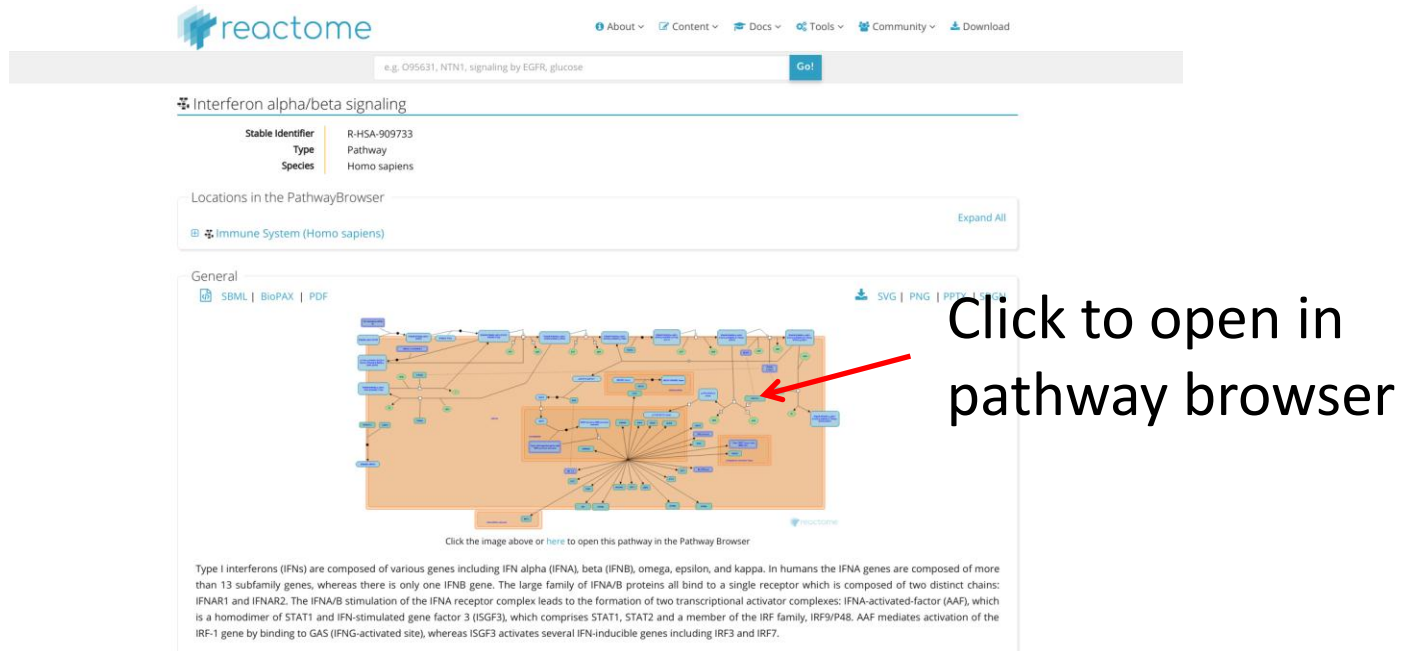
Expected overlap if independent

	query	non-query
pathway	17	196-17
non-pathway	64-17	45956-64-196+17

$$196 * 64 / 45956 = 0.27 \longrightarrow = 1.05$$

# Looking into reactome pathways

Google: reactome interferon alpha beta signalling













The screenshot shows the Reactome website interface. At the top, there is a search bar with the text "e.g. O195631, NTN1, signaling by EGFR, glucose" and a "Go!" button. Below the search bar, the pathway "Interferon alpha/beta signaling" is displayed. A table provides details: Stable Identifier (R-HSA-909733), Type (Pathway), and Species (Homo sapiens). Below this, there are sections for "Locations in the PathwayBrowser" (showing "Immune System (Homo sapiens)") and "General" (with links for SBML, BioPAX, and PDF). A large diagram of the pathway is shown, with a red arrow pointing to a button that says "Click to open in pathway browser".

Click to open in pathway browser

Type I interferons (IFNs) are composed of various genes including IFN alpha (IFNA), beta (IFNB), omega, epsilon, and kappa. In humans the IFNA genes are composed of more than 13 subfamily genes, whereas there is only one IFNB gene. The large family of IFNA/B proteins all bind to a single receptor which is composed of two distinct chains: IFNAR1 and IFNAR2. The IFNA/B stimulation of the IFNA receptor complex leads to the formation of two transcriptional activator complexes: IFNA-activated-factor (AAF), which is a homodimer of STAT1 and IFN-stimulated gene factor 3 (ISGF3), which comprises STAT1, STAT2 and a member of the IRF family, IRF9/P48. AAF mediates activation of the IRF-1 gene by binding to GAS (IFNG-activated site), whereas ISGF3 activates several IFN-inducible genes including IRF3 and IRF7.

# Let's look how the whole pathway looks in our dataset

Click on IFNa/b signalling

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K
REACTOME_IMMUNE_SYSTEM [933]	Genes involved in Immune System	50	
REACTOME_INTERFERON_SIGNALING [159]	Genes involved in Interferon Signaling	26	
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM [270]	Genes involved in Cytokine Signaling in Immune system	30	
<a href="#">REACTOME_INTERFERON_ALPHA_BETA_SIGNALING [64]</a>	Genes involved in Interferon alpha/beta signaling	17	
REACTOME_INTERFERON_GAMMA_SIGNALING [63]	Genes involved in Interferon gamma signaling	13	
PID_IL23_PATHWAY [37]	IL23-mediated signaling events	8	
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION [267]	Cytokine-cytokine receptor interaction	14	
NABA_MATRISOME_ASSOCIATED [753]	Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors	20	
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES [57]	Genes involved in Chemokine receptors bind chemokines	8	
REACTOME_ADAPTIVE_IMMUNE_SYSTEM [539]	Genes involved in Adaptive Immune System	17	

Gene Set: REACTOME\_INTERFERON\_ALPHA\_BETA\_SIGNALING

<b>Standard name</b>	REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
<b>Systematic name</b>	M973
<b>Brief description</b>	Genes involved in Interferon alpha/beta signaling
<b>Full description or abstract</b>	
<b>Collection</b>	C2: curated gene sets CP:REACTOME: Reactome gene sets
<b>Source publication</b>	
<b>Exact source</b>	
<b>Related gene sets</b>	
<b>External links</b>	<a href="http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_25162">http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_25162</a>
<b>Organism</b>	Homo sapiens
<b>Contributed by</b>	Reactome
<b>Source platform</b>	EntrezGeneIds
<b>Dataset references</b>	
<b>Download gene set</b>	format: grp   <a href="#">text</a>   gmt   gmx   xml
<b>Compute overlaps</b> ?	(show collections to investigate for overlap with this gene set)
<b>Compendia expression profiles</b> ?	Human tissue compendium (Novartis) NCI-60 cell lines (National Cancer Institute)
<b>Advanced query</b>	Further investigate these 64 genes
<b>Gene families</b> ?	Categorize these 64 genes by gene family
<b>Show members</b>	(show 65 members mapped to 64 genes)
<b>Version history</b>	3.1: First introduced

click "text"

# Exercise

- ✓ Get list of genes in IFNa/b reponse pathway from QuickGO:
  - <https://www.ebi.ac.uk/QuickGO>

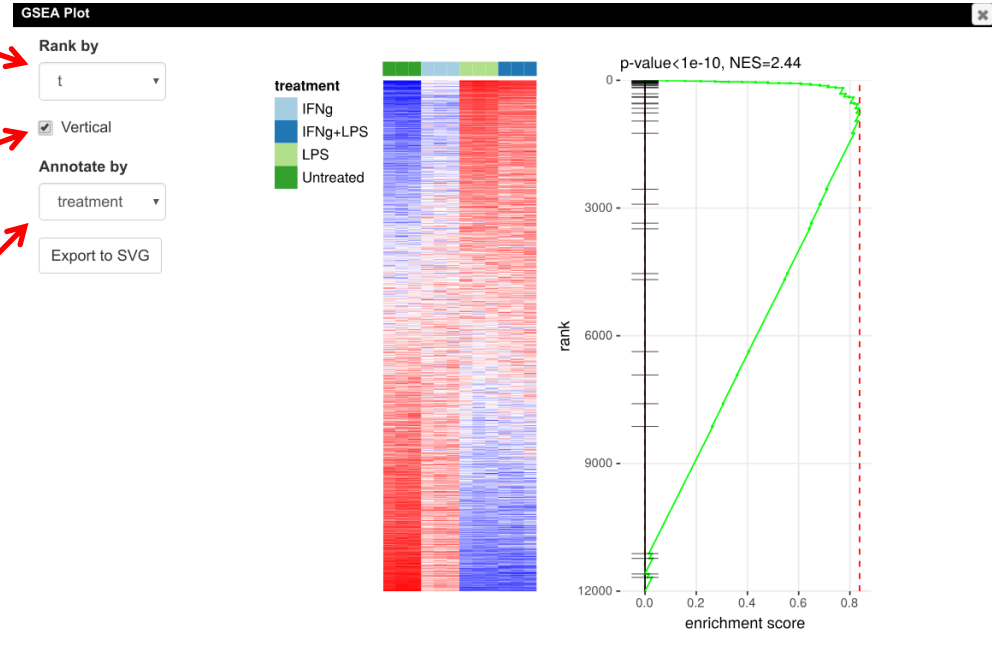


# Tools/Plots/GSEA plot

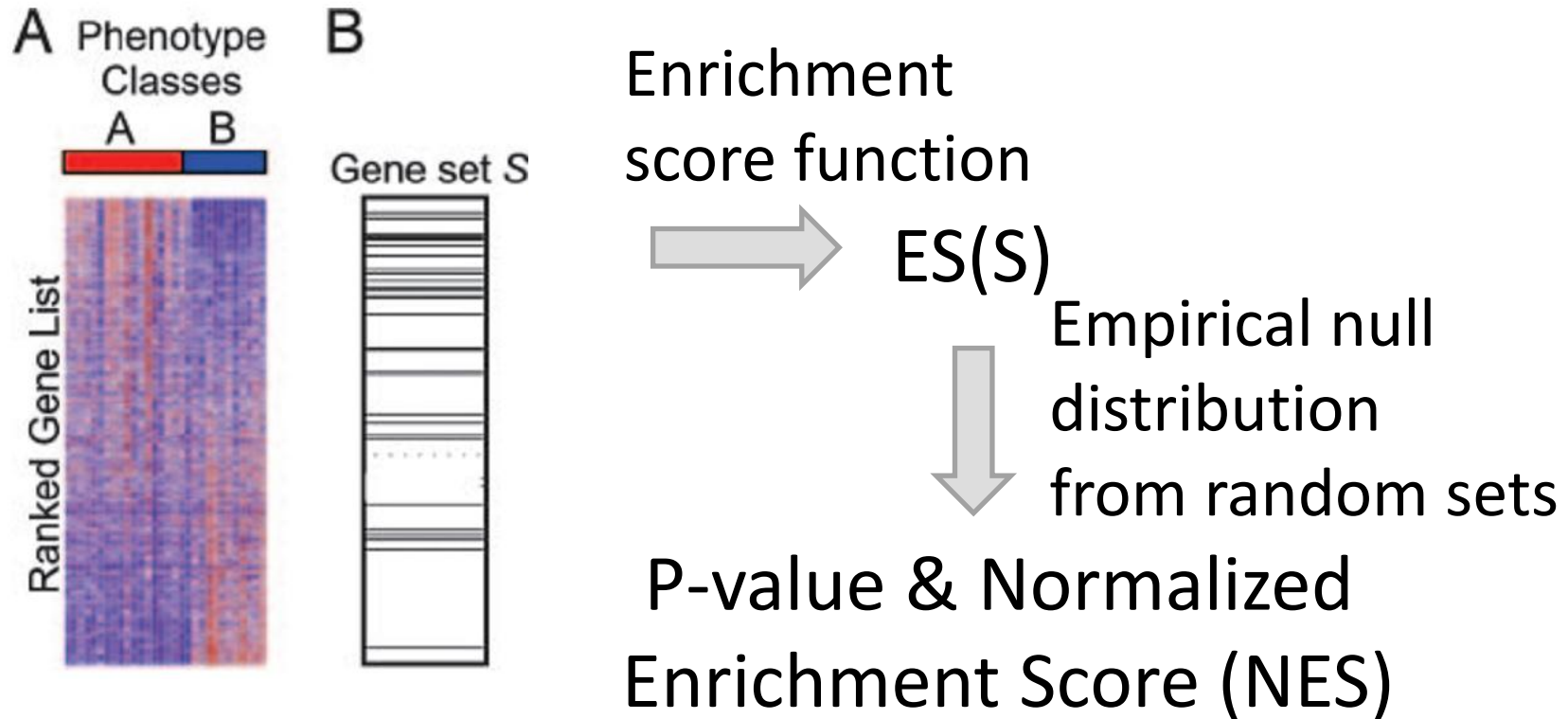
Select "t"

Check "vertical"  
(or don't)

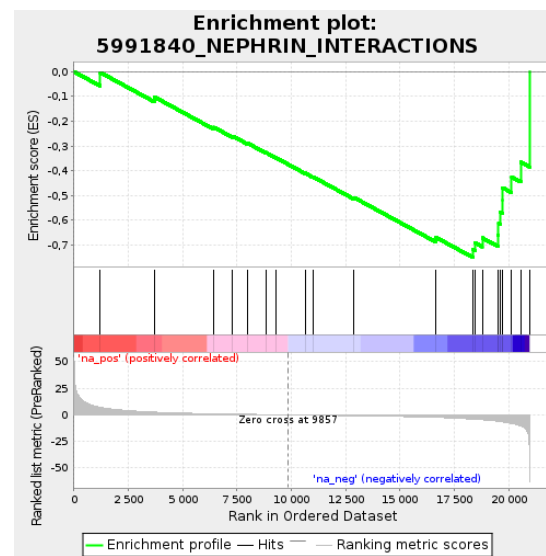
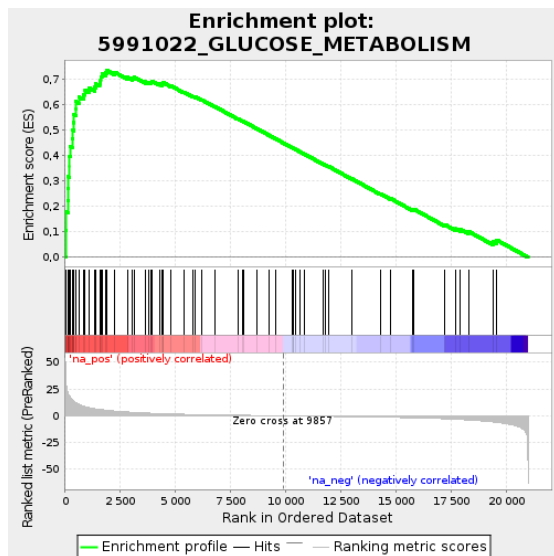
Annotate by  
"treatment"



# Pre-ranked gene set enrichment analysis



# GSEA statistic calculation



- ✓  $ES[i] = ES[i-1] + |stat[i]| / NS$ , if  $i \in$  gene set
- ✓  $ES[i] = ES[i-1] - 1 / (N - k)$ , if not
- ✓  $NS = \sum |stat[i]|$  for  $i \in$  gene set



# Fast gene set enrichment analysis (FGSEA)

- ✓ Fast estimation of arbitrarily low p-values
- ✓ <https://github.com/ctlab/fgsea>
- ✓ The error can be estimated



Gennady  
Korotkevich

size	ES	Z	log2P (exact)	log2P (approx)	mean error	sd error	expected error
100	0.9	1001	-159.53812	-159.53769	0.00043054	0.61	0.48
100	0.7	1001	-79.54245	-79.5298	0.01265205	0.41	0.34
100	0.6	1001	-31.54746	-31.54093	0.00652617	0.26	0.21
100	0.5	1001	-13.56805	-13.5644	0.00364966	0.17	0.14



Vladimir  
Sukhov

# Running FGSEA

- ✓ Tools/Pathway analysis/Perform FGSEA
- ✓ Pathway database: MSigDB Hallmarks (Mouse)
- ✓ Rank by: t
- ✓ Column with gene ID: Gene ID

Actions:

Save as TSV

FGSEA:

pathway	pval	padj	log2err	ES	NES	size	leadingEdge
Interferon gamma response	1.00e-10	8.17e-10	NA	0.836	3.14	183	17329 14962 20293 14293 16193 ...
Interferon alpha response	1.00e-10	8.17e-10	NA	0.814	2.82	94	15945 56066 99899 231655 57444 ...
Inflammatory response	1.00e-10	8.17e-10	NA	0.771	2.81	150	17329 17167 16175 20293 14293 ...
IL6-JAK-STAT3 signaling	1.00e-10	8.17e-10	NA	0.787	2.59	72	17329 330122 16193 12494 15945 ...
Allograft rejection	1.00e-10	8.17e-10	NA	0.681	2.47	142	17329 20293 16193 16176 18126 ...

Pathway detail:

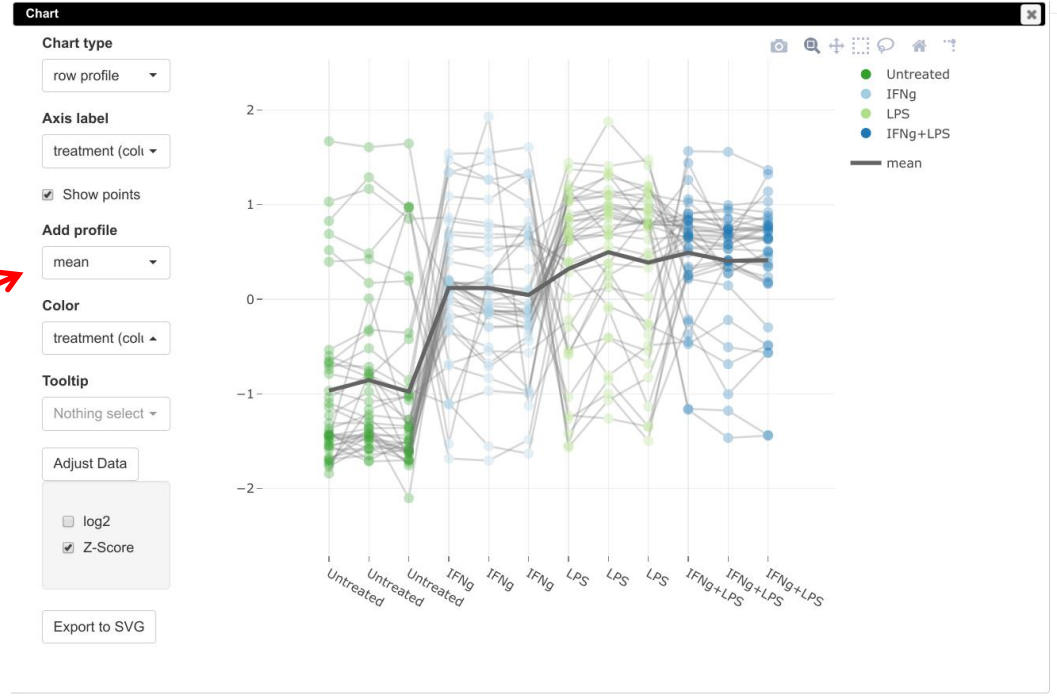
**Pathway name:** Interferon gamma response  
**Pathway genes (ID):** 56417 71939 71371 11861 3199 74 12010 242248 74481 12183 69550 12226 667277 317677 12362 12367 12363 12369 12370 20293 2030 4 20306 60533 12494 21939 12515 16149 12524 125 75 14962 12628 12265 14747 22169 74157 12984 15 945 56066 17329 230073 234311 80861 19106 66892 108670 14102 14129 14190 14293 55932 236573 145 28 110168 14938 229003 67138 15251 14964 15006 15007 15013 15015 15018 110557 15040 630294 667 977 14990 14991 14998 14960 14969 15894 15930 5

# Pathway profile plot

- ✔ Tools/Chart
- ✔ Select genes from IFNa/b pathway

Add profile: mean

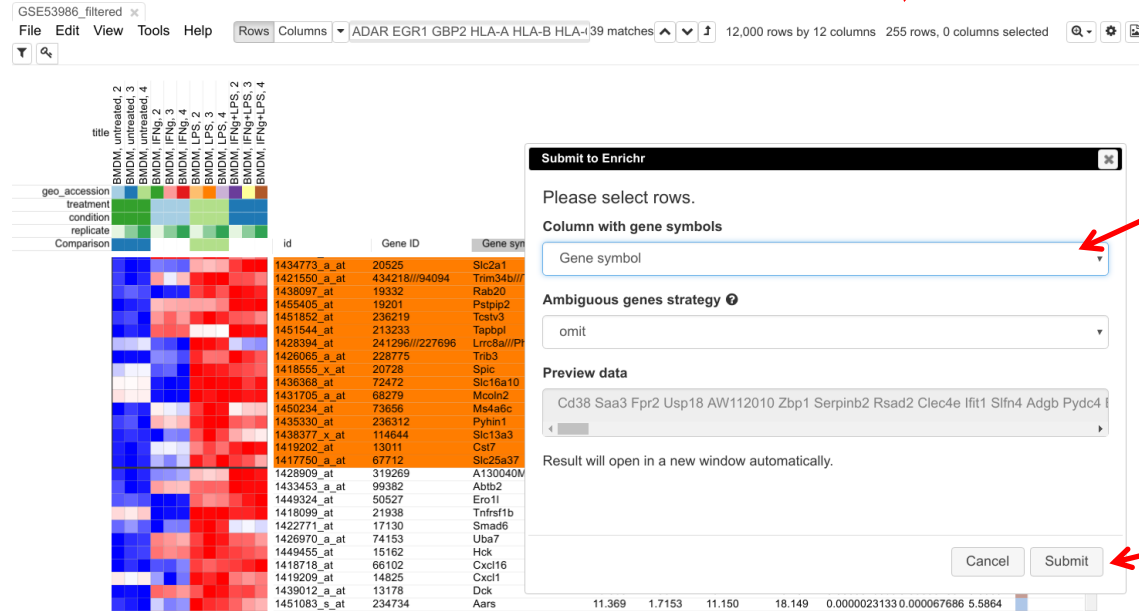
Adjust: Z-score



# Using Enrichr

Select ~250 genes

✓ Tools/Pathway analysis/Submit to Enrichr



The screenshot shows the Enrichr web interface. On the left, there is a heatmap with columns labeled 'BMDM, untreated, 2' through 'BMDM, IFNg+LPS, 4'. The main area displays a table with columns for 'id', 'Gene ID', and 'Gene sym'. A dialog box titled 'Submit to Enrichr' is open, showing options for 'Column with gene symbols' (set to 'Gene symbol'), 'Ambiguous genes strategy' (set to 'omit'), and a 'Preview data' section. A red arrow points to the 'Gene symbol' dropdown, and another red arrow points to the 'Submit' button.

Choose gene symbol column

Submit

# Enrichr results

Check out different tabs



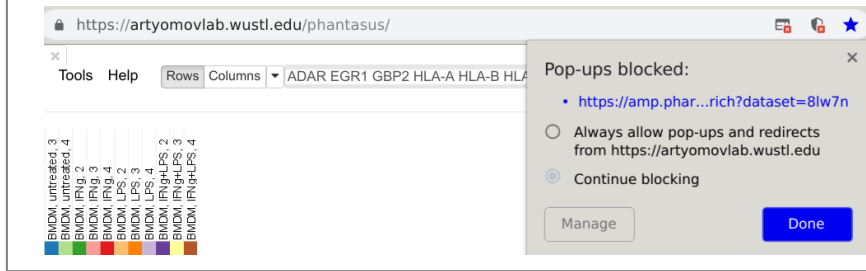
[Login](#) | [Register](#)

[Transcription](#)
[Pathways](#)
[Ontologies](#)
[Disease/Drugs](#)
[Cell Types](#)
[Misc](#)
[Legacy](#)
[Crowd](#)

Description GSE53986\_filtered (237 genes)

<p><b>ChEA 2016</b></p> <ul style="list-style-type: none"> <li>IRF8_22096565_ChIP-ChIP_GC-B_Mouse</li> <li>IRF1_21803131_ChIP-Seq_MONOCYTES_Hun</li> <li>IRF8_21731497_ChIP-ChIP_J774_Mouse</li> <li>IRF8_22096565_ChIP-ChIP_GC-B_Human</li> <li>IRF8_27001742_ChIP-Seq_BMDM_Mouse</li> </ul>	<p><b>ENCODE and ChEA Consensus TFs from</b></p> <ul style="list-style-type: none"> <li>IRF8_CHEA</li> <li>IRF1_ENCODE</li> <li>RELA_ENCODE</li> <li>STAT3_ENCODE</li> <li>VDR_CHEA</li> </ul>	<p><b>ARCHS4 TFs Coexp</b></p> <ul style="list-style-type: none"> <li>ETV7_human_tf_ARCHS4_coexpression</li> <li>BATF2_human_tf_ARCHS4_coexpression</li> <li>ZNF1_human_tf_ARCHS4_coexpression</li> <li>STAT2_human_tf_ARCHS4_coexpression</li> <li>IRF7_human_tf_ARCHS4_coexpression</li> </ul>
<p><b>TF Perturbations Followed by Expression</b></p> <ul style="list-style-type: none"> <li>STAT1_KO_MOUSE_GSE48970_CREEDSID_GE</li> <li>DUX4_DE_HUMAN_GSE33799_CREEDSID_GE</li> <li>IRF9_OE_HUMAN_GSE50002_CREEDSID_GEN</li> <li>NCOA1_KO_HUMAN_GSE28645_CREEDSID_G</li> <li>TWIST1_OE_MOUSE_GSE50002_CREEDSID_G</li> </ul>	<p><b>Enrichr Submissions TF-Gene Cooccurrence</b></p> <ul style="list-style-type: none"> <li>ZBP1</li> <li>BATF2</li> <li>IRF1</li> <li>SP140</li> <li>TRAFD1</li> </ul>	<p><b>TRANSFAC and JASPAR PWMs</b></p> <ul style="list-style-type: none"> <li>STAT1 (mouse)</li> <li>PRDM1 (mouse)</li> <li>IRF8 (mouse)</li> <li>REL (human)</li> <li>MYO2-3 (human)</li> </ul>
<p><b>Epigenomics Roadmap HM ChIP-seq</b></p> <ul style="list-style-type: none"> <li>H3K27me3 H1 BMP4 Derived Mesendoderm</li> <li>H2BK120ac IMR90</li> </ul>	<p><b>TargetScan microRNA 2017</b></p> <ul style="list-style-type: none"> <li>hsa-miR-4317</li> <li>hsa-miR-219-1-3p</li> </ul>	<p><b>miRTarBase 2017</b></p> <ul style="list-style-type: none"> <li>hsa-miR-26b-5p</li> <li>hsa-miR-124-3p</li> </ul>

NB: if the new tab is not opening, probably pop-up windows are blocked



# Other tools

## ✓ Gorilla

- <http://cbl-gorilla.cs.technion.ac.il/>

## ✓ DAVID

- <https://david.ncifcrf.gov/home.jsp>

## ✓ GSEA desktop tool

- <http://software.broadinstitute.org/gsea/downloads.jsp>

## ✓ GenePattern

- <https://genepattern.broadinstitute.org/>

# Summary

- ✔ Differential expression analysis and pathway enrichment are one of the most useful tools for data interpretation
- ✔ There are many online tools and databases for downstream analysis
- ✔ Next: working with public gene expression databases