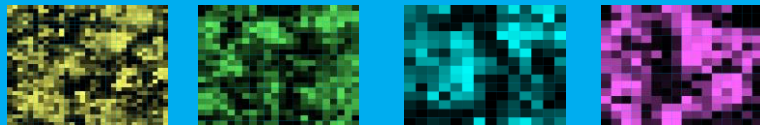


# Enrichment Analysis and Enrichr



## Network Analysis in Systems Biology

Avi Ma'ayan, PhD

Associate Professor

Department of Pharmacology and Systems Therapeutics

Icahn School of Medicine at Mount Sinai, New York, NY 10029



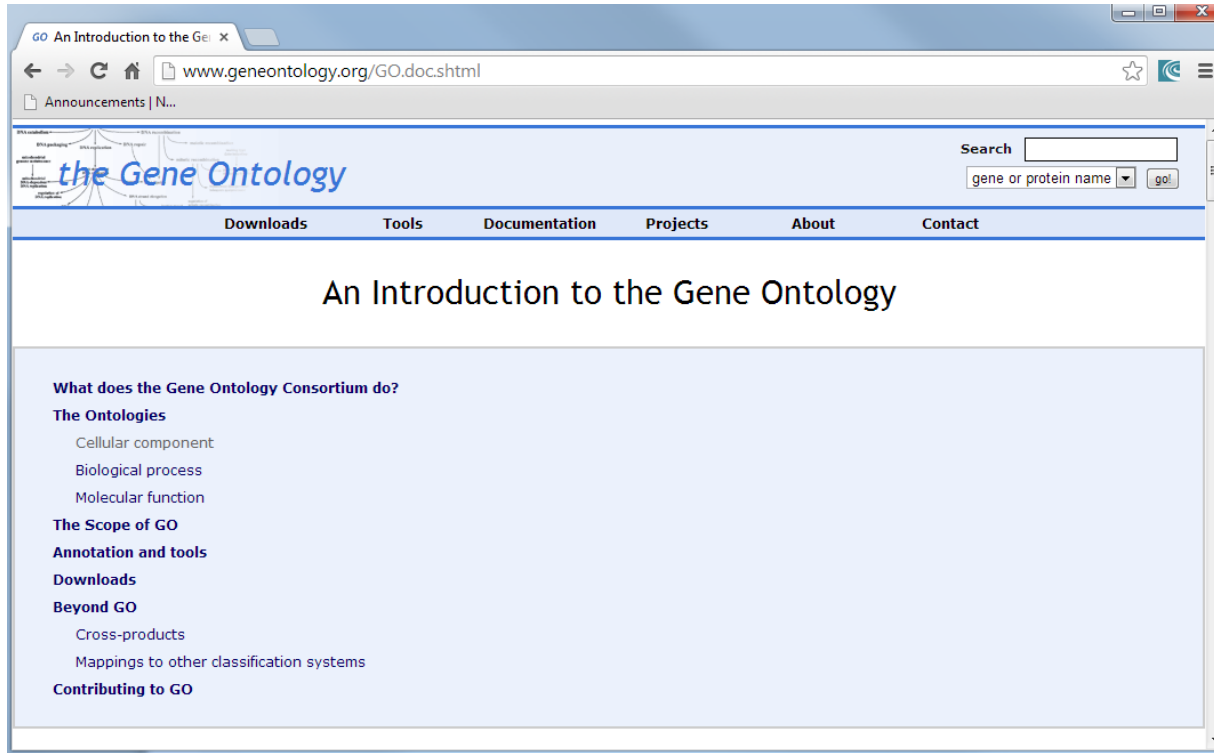
**Mount  
Sinai**

# Gene-set Libraries Stored in Gene Matrix Transpose (GMT) files- GO BP Example

GeneOntology\_BP - Microsoft Excel

R25 RPL10									
	A	B	C	D	E	F	G	H	I
1	protein secretion (GO:0009306)	na	PPY	ARFGAP3	NOX5	PLEK	LTBP2	PDIA4	ABCA1
2	rRNA transcription (GO:0009303)	na	GTF3A	BRF1	RNASEK	POLR1E	MKI67IP	CD3EAP	ANG
3	positive regulation of DNA replication (GO:0045740)	na	CSF2	IL3	PDGFB	PDGFA	TGFB3	IGF1	KITLG
4	respiratory burst (GO:0045730)	na	CYBA	CD55	CYBB	NCF2	NCF1	NOX1	PGAM1
5	positive regulation of protein catabolic process (GO:0045732)	na	TAF1	F12	TNF	AURKAIP1	SMAD7	NRD1	EGLN2
6	positive regulation of DNA repair (GO:0045739)	na	BRCC3	FAM175A	CEBPG	MERIT40	PRKCG	BRCA1	UIMC1
7	negative regulation of adenylate cyclase activity (GO:0007194)	na	MCHR1	GNAI3	DRD3	DRD2	NPY2R	OPRK1	DRD4
8	inhibition of adenylate cyclase activity by G-protein signaling (GO:0007193)	na	MCHR1	DRD3	DRD2	OPRL1	NPY2R	OPRK1	DRD4
9	regulation of transcription factor activity (GO:0051090)	na	STK36	PRDX3	IL10	RPS3	MEN1	NOD2	NLRC3
10	activation of adenylate cyclase activity (GO:0007190)	na	CALCR	ACR	AVPR2	ADORA3	CAP2	DRD3	ADORA2A
11	positive regulation of transcription factor activity (GO:0051091)	na	TNF	MTDH	UBE2V1	PRDX3	IL10	IRAK3	NOD2
12	positive regulation of NF-kappaB transcription factor activity (GO:0051092)	na	TNF	MTDH	UBE2V1	PRDX3	IRAK3	NOD2	NLRC3
13	response to radiation (GO:0009314)	na	RP1	XRCC4	FAM175A	COPS3	NR2E3	MEN1	AKT1
14	oligosaccharide metabolic process (GO:0009311)	na	LALBA	B4GALT1	GAL3ST3	GAL3ST4	ST6GAL2	FUT8	MAN1B1
15	positive regulation of glycogen biosynthetic process (GO:0045725)	na	AKT1	IRS2	SORBS1	GCK	INS	DYRK2	IRS1
16	positive regulation of tyrosine phosphorylation of Stat3 protein (GO:0042517)	na	LIF	GH1	IL6	CNTF	IL6ST	CLCF1	IL6R
17	positive regulation of binding (GO:0051099)	na	MTDH	TNF	UBE2V1	FKBP1A	DPH3	PRDX3	IL10
18	positive regulation of translation (GO:0045727)	na	DAZ1	IL6	WIBG	EIF5A	BCL3	DAZL	RPS4X

# Few Words about the Gene Ontology



# Gene-set Libraries Stored in Gene Matrix Transpose (GMT) files- KEGG Example

KEGG_pathways.gmt - Microsoft Excel											
Home Insert Page Layout Formulas Data Review View Acrobat											
Clipboard Font Alignment Number Styles											
R42											
	A	B	C	D	E	F	G	H	I	J	K
1	HSA00010_GLYCOLYSIS_AND_GLUcONEOGENESIS	na	LDHC	LDHB	LDHA	ADH1C	PGAM1	ADH1B	PGAM2	ADH1A	ACSS2
2	HSA00020_CITRATE_CYCLE	na	OGDHL	OGDH	CLYBL	IDH3G	LOC28339	IDH2	IDH1	SUCLA2	FH
3	HSA00030_PENTOSE_PHOSPHATE_PATHWAY	na	ALDOA	TALDO1	ALDOC	ALDOB	PGD	TKTL2	TKTL1	DERA	RPIA
4	HSA00031_INOSITOL_METABOLISM	na	ALDH6A1	TPI1							
5	HSA00040_PENTOSE_AND_GLUcURONATE_INTERCONVERSIONS	na	UGDH	UGT1A7	UGT1A6	UGT1A9	UGT1A8	UGT1A3	UGT1A5	UGT1A4	UGT2A1
6	HSA00051_FRUCTOSE_AND_MANNOSE_METABOLISM	na	ALDOA	SORD	PFKFB4	HSD3B7	PFKFB3	ALDOC	PFKFB2	ALDOB	PFKFB1
7	HSA00052_GALACTOSE_METABOLISM	na	LALBA	HSD3B7	HK2	HK1	G6PC2	GLB1	GALK2	GALK1	HK3
8	HSA00053_ASCORBATE_AND_ALDARATE_METABOLISM	na	ALDH7A1	ALDH1B1	ALDH1A3	MIOX	UGDH	ALDH2	ALDH3A2	ALDH9A1	ALDH3A1
9	HSA00061_FATTY_ACID_BIOSYNTHESIS	na	OLAH	MCAT	ACACA	FASN	ACACB	OXSM			
10	HSA00062_FATTY_ACID_ELONGATION_IN_MITOCHONDRIA	na	HSD17B10	ACAA2	PPT2	ECHS1	PPT1	HSD17B4	HADH	MECR	HADHA
11	HSA00071_FATTY_ACID_METABOLISM	na	ACOX1	HSD17B10	ACADSB	CPT2	ADHFE1	EHHADH	ADH5	ADH1C	ADH6
12	HSA00072_SYNTHESIS_AND_DEGRADATION_OF_KETONE_BODIES	na	HMGCS2	OXCT1	HMGCS1	OXCT2	BDH2	ACAT2	BDH1	ACAT1	HMGCL
13	HSA00100_BIOSYNTHESIS_OF_STEROIDS	na	TM7SF2	GGCX	EBP	MVD	CYP51A1	HMGCR	FDP5	LSS	PMVK
14	HSA00120_BILE_ACID_BIOSYNTHESIS	na	ADHFE1	HSD3B7	ADH5	ADH1C	ADH6	ADH1B	ADH7	ADH1A	ALDH3A2
15	HSA00130_UBIQUINONE_BIOSYNTHESIS	na	ND1	NDUF81	ND4	ND5	ND2	ND3	NDUFA13	COQ7	COQ5
16	HSA00140_C21_STEROID_HORMONE_METABOLISM	na	HSD3B2	CYP17A1	HSD3B1	AKR1C4	CYP11A1	CYP21A2	CYP11B1	CYP11B2	HSD11B1
17	HSA00150_ANDROGEN_AND_ESTROGEN_METABOLISM	na	ARSD	ARSE	CYP11B1	CYP11B2	SULT2B1	PRMT3	AKR1C4	PRMT2	PRMT7
18	HSA00190_OXIDATIVE_PHOSPHORYLATION	na	ATP6AP1	NDUFAB1	COX5A	COX5B	ATP8	ATP7	UQCRC	COX6C	ATP5C1
19	HSA00220_UREA_CYCLE_AND_METABOLISM_OF_AMINO_GROUPS	na	SAT1	ALDH18A1	SRM	NAG5	ASS1	SAT2	AGMAT	ASL	ALDH3A2
20	HSA00230_PURINE_METABOLISM	na	ADCY3	FHIT	ADCY4	GDA	ADCY1	ADCY2	GMPR2	ADCY7	ADCY8
21	HSA00232_CAFFEINE_METABOLISM	na	XDH	CYP2A13	NAT1	NAT2	CYP2A6	CYP2A7	CYP1A2		
22	HSA00240_PYRIMIDINE_METABOLISM	na	CTPS	DTYMK	CAD	CANT1	PRIM1	NTSM	NT5C3	PRIM2	NT5C2
23	HSA00251_GLUTAMATE_METABOLISM	na	GCLC	GLUD2	GLUD1	GPNPAT1	CAD	QARS	NAGK	GCLM	PPAT
24	HSA00252_ALANINE_AND_ASPARTATE_METABOLISM	na	ADSS	ASS1	NARS	CAD	ASNS	ASL	AGXT	PDHB	GOT2
25	HSA00260_GLYCINE_SERINE_AND_THREONINE_METABOLISM	na	CHKA	TARS2	HSD3B7	CHKB	AMT	GCAT	PSPH	AGXT	PIPOX
26	HSA00271_METHIONINE_METABOLISM	na	KIAA0828	DNMT3A	AHCY	SRM	TAT	CTH	MAT1A	BHMT	MTR
27	HSA00272_CYSTEINE_METABOLISM	na	LDHC	LDHB	CARS	LDHA	LDHAL6B	LDHAL6A	CDO1	GOT2	SULT4A1
28	HSA00280_VALINE_LEUCINE_AND_IsoleucINE_DEGRADATION	na	BCAT1	HSD17B10	BCAT2	EHHADH	HMGCS1	ECHS1	ACAT2	ACAT1	ALDH3A2
29	HSA00290_VALINE_LEUCINE_AND_IsoleucINE_BIOSYNTHESIS	na	BCAT1	IARS	BCAT2	VARS2	IARS	ILVBL	PDHA2	IARS2	PDHA1
30	HSA00300_LYSINE_BIOSYNTHESIS	na	AADAT	AASDHP	AASS	KARS					
31	HSA00310_LYSINE_DEGRADATION	na	HSD17B10	HSD3B7	EHHADH	SETD1A	OGDHL	ECHS1	AASS	ACAT2	OGDH
32	HSA00330_ARGININE_AND_PROLINE_METABOLISM	na	PYCR1	CKMT1B	ASS1	GLUD2	GLUD1	ASL	CKB	GOT2	NOS2A
33	HSA00340_HISTIDINE_METABOLISM	na	CNDP1	WBSCR22	ALDH3A2	ALDH3A1	PRMT3	ASPA	PRMT2	PRMT7	HNMT
34	HSA00350_TYROSINE_METABOLISM	na	PNMT	ADH1C	ADH1B	ADH1A	FAH	PRMT3	GOT2	TYR	GOT1
35	HSA00360_PHENYLALANINE_METABOLISM	na	LYCAT	ARD1A	ALDH3A1	MIF	GOT2	GOT1	SH3GLB1	EPX	ALDH1A3
36	HSA00361_GAMMA_HEXACHLOROCYCLOHEXANE_DEGRADATION	na	CYP3A4	ALPL	CYP3A5	ALPPL2	CYP3A7	ALPI	ACP6	ACP5	ACPT
37	HSA00363_BISPHENOL_A_DEGRADATION	na	HSD3B7	DHRSX	DHRS7	DHRS1	RDH12	DHRS2	RDH11	DHRS3	AKR1B10

KEGG Pathway: MAPK

www.genome.jp/kegg-bin/show\_pathway?map=hsa04010&show\_description=show

Announcements | N...

KEGG MAPK signaling pathway - Homo sapiens (human)

[ Pathway menu | Organism menu | Pathway entry | Download KGM | Hide description | User data mapping ]

The mitogen-activated protein kinase (MAPK) cascade is a highly conserved module that is involved in various cellular functions, including cell proliferation, differentiation and migration. Mammals express at least four distinctly regulated groups of MAPKs, extracellular signal-related kinases (ERK-1/2, Jun amino-terminal kinases (JNK1/2/3), p38 proteins (p38alpha/beta/gamma/delta) and MEK5, that are activated by specific MAPKKs: MEK1/2 for ERK1/2, MKK3/6 for the p38, MKK4/7 (JNK1/2) for the JNKs, and MEK5 for ERK5. Each MAPKK, however, can be activated by more than one MAPKKK, increasing the complexity and diversity of MAPK signalling. Presumably each MAPKKK confers responsiveness to distinct stimuli. For example, activation of ERK1/2 by growth factors depends on the MAPKKK c-Raf, but other MAPKKKs may activate ERK1/2 in response to pro-inflammatory stimuli.

Homo sapiens (human) Go 100% Go

MAPK SIGNALING PATHWAY

Phosphatidylinositol signaling system

Chetral MAP kinase pathway

JNK and p38 MAP kinase pathway

Serum, cytotoxic drugs, irradiation, heatshock, reactive oxygen species, lipopolysaccharide, and other stress

MAPK signaling pathway diagram showing the activation of ERK, JNK, and p38 pathways by various stimuli and the resulting downstream effects on cell proliferation, differentiation, and migration.

<http://www.genome.jp/kegg/>

# Why Destroy Ontologies and Pathways and Use this Simple Test?

Because it is easy and potentially useful...

## Applying the Fisher Exact Test to Compute Overlap Between Differentially Expressed Genes and Gene Sets from Gene Set Libraries

Overlapping Genes Between My List and a List from a Gene Set Library	Genes in the Gene Set Library
My Differentially Expressed Genes	All Genes

# The Fisher Exact Test

$$P_{\text{cutoff}} = \frac{(R_1! R_2! \cdots R_m!) (C_1! C_2! \cdots C_n!)}{N! \prod_{i,j} a_{ij}!}$$

	Math.	Mag.	Science	
math	5	0		$R_1 = 5$
biology	1	4		$R_2 = 5$
	$C_1 = 6$	$C_2 = 4$		$N = 10.$

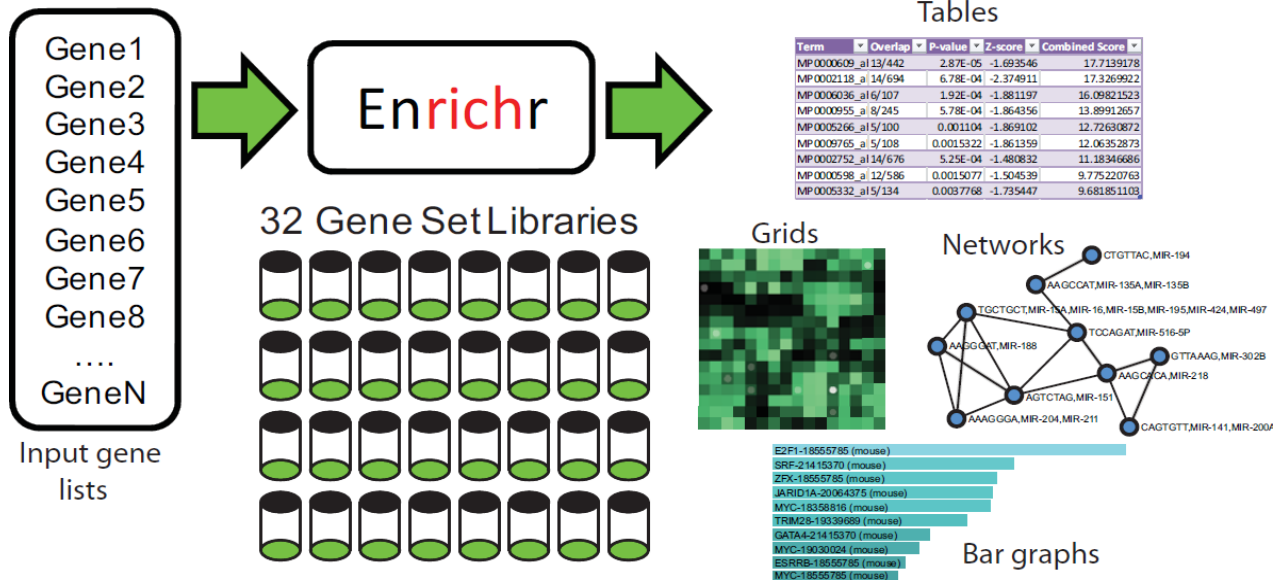
Computing  $P_{\text{cutoff}}$  gives

$$P_{\text{cutoff}} = \frac{5!^2 6! 4!}{10! (5! 0! 1! 4!)} = 0.0238$$

From <http://mathworld.wolfram.com/FishersExactTest.html>



# Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool



Edward Y. Chen et al. BMC Bioinformatics (2013) In press

<http://amp.pharm.mssm.edu/Enrichr/index.html>

# Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool



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

32563 lists analyzed!

**Analyze** What's New? Dataset Statistics Find A Gene About ?

## Input data

Choose an input file to upload. Separate each gene symbol with a new line. For a quantitative set, add a comma and the level of membership of that gene between 0 and 1 after each gene symbol.

No file chosen

Or paste in a list of gene symbols optionally followed by a comma and levels of membership between 0 and 1 with each gene separated by a new line. Try [a regular example](#) or [an example of a quantitative set](#).

0 gene(s) entered

Enter a brief description for the list in case you want to share it. (Optional)

Please acknowledge Enrichr in your publications by citing the following reference:  
Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;128(14).



# Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool



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

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

**Description** Sample gene list (413 genes)

KEGG

**Bar Graph** Table Grid Network ⚙

Click the bars to sort. Now sorted by **combined score**.

[SVG](#) [PNG](#) [JPG](#)

HSA00630 GLYOXYLATE AND DICARBOXYLATE METABOLISM

HSA00640 PROPANOATE METABOLISM

HSA00350 TYROSINE METABOLISM

HSA00120 BILE ACID BIOSYNTHESIS

HSA00641 3 CHLOROACRYLIC ACID DEGRADATION

HSA00071 FATTY ACID METABOLISM

HSA00280 VALINE LEUCINE AND ISOLEUCINE DEGRADATION

HSA00380 TRYPTOPHAN METABOLISM

HSA00624 1 AND 2 METHYLNAPHTHALENE DEGRADATION

HSA00643 STYRENE DEGRADATION

WikiPathways

Reactome

BioCarta

PPI Hub Proteins



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

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

**Description** Sample gene list (413 genes)

KEGG

Bar Graph **Table** Grid Network ⚙

Hover each row to see the overlapping genes.

Show 10 entries

Search:

Index	Name	P-value	Z-score	Combined Score
1	HSA00630 GLYOXYLATE AND DICARBOXYLATE METABOLISM	0.03420	-1.54	5.21
2	HSA00640 PROPANOATE METABOLISM	0.1546	-1.77	3.30
3	HSA00350 TYROSINE METABOLISM	0.1571	-1.66	3.08
4	HSA00120 BILE ACID BIOSYNTHESIS	0.1476	-1.45	2.78
5	HSA00641 3 CHLOROACRYLIC ACID DEGRADATION	0.1656	-1.47	2.64
6	HSA00071 FATTY ACID METABOLISM	0.1739	-1.42	2.48
7	HSA00280 VALINE LEUCINE AND ISOLEUCINE DEGRADATION	0.1766	-1.36	2.36
8	HSA00380 TRYPTOPHAN METABOLISM	0.2451	-0.92	1.30
9	HSA00624 1 AND 2 METHYLNAPHTHALENE DEGRADATION	0.1582	-0.58	1.08
10	HSA00643 STYRENE DEGRADATION	0.1943	-0.61	1.00

Showing 1 to 10 of 113 entries

◀ Previous Next ▶

[Export to Table](#)

WikiPathways

Reactome

# Further Reading and Resources

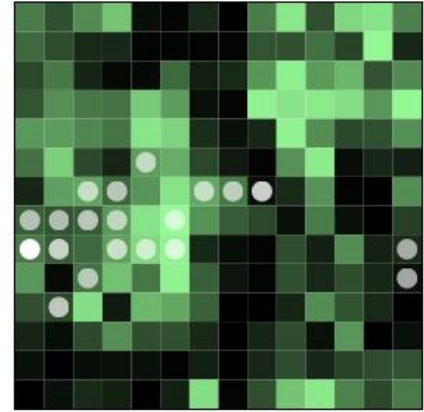
<http://david.abcc.ncifcrf.gov/>

<http://www.broadinstitute.org/gsea>

<http://www.lists2networks.org>

<http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html>

[http://bioinformatics.mdanderson.org/MicroarrayCourse/Lectures09/gsea1\\_bw.pdf](http://bioinformatics.mdanderson.org/MicroarrayCourse/Lectures09/gsea1_bw.pdf)



Aravind Subramanian, et al. **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.** PNAS 2005 102 (43) 15545-15550

Sherman, Brad T., and Richard A. Lempicki. **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic acids research* 37.1 (2009): 1-13.