



Metascope

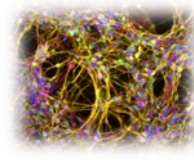
A Gene Annotation & Analysis Resource

<http://metascope.org>

Gene Prioritization is a Routine Task

High-throughput target discovery platforms produce hundreds of candidate genes, but only a handful can be followed up in downstream analysis.

HT-Target
Discovery
Platforms



Gene ID, Activities

ADCY1,10.5

AFG3L2,8.5

AKT1,6.6

ALG2,3.5

...

**Primary
Gene
List**

This is too
much for me
to follow up





What Data to Gather

We need to know their descriptions, biological processes involved, are they GPCRs, are they secreted proteins, are they expressed in the tissue of interest, are there chemical probes readily available? ...

Gathering such data for a gene list is a non-trivial task for biomedical researchers.



GENEONTOLOGY
Unifying Biology

DRUGBANK
Open Data Drug & Drug Target Database



THE HUMAN PROTEIN ATLAS

I can browse these sites for a few genes, but not for hundreds of genes!

http://metascape.org

- Fresh Data sources
- Free for everyone
- Easy to use

DAVID

Re: Update of database
by Guest » Fri Mar 20, 2015 11:57

Dear DAVID team,

Where could I find information on data

Many thanks,

Bertrand

Release & Version Information
DAVID 6.7 (current release) Jan. 2010

Re: Update of database
by Guest » Tue Apr 28, 2015 4:21 am

David died

<https://david.ncifcrf.gov/>

Alternative desperately needed!

Genecards.org

ACADEMIC PRICING

LIMITED TIME OFFER!

SINGLE USER
SUBSCRIPTION

\$349 /yr

BUY NOW >

OR

\$79/month

BUY NOW >

LIMITED TIME OFFER!

RESEARCH
GROUP

\$899 / yr

Up to 10 users

BUY NOW >

<http://geneanalytics.genecards.org/pricing/>

Express Analysis – Answer is One Click Away

Step 1

ADRBK1,AEBP2,AES,AFG3L2,AHCTF1,AKAP13,ALKBH3,AMDHD2,ANAPC2,AP1G2,AQR,ARID1A,ARID5A,ATMIN,ATP6V0C,ATXN2,B4GALNT4,BCL1

Submit

Cancel

Just Paste in a Gene List* and Click “Express Analysis”!

Step 2

Express Analysis

All I need for Decision Making

Homologene	Gene Symbol	Description	Gene Summ	Biological Pr	Kinase Class	Protein Func	Subcellular L	Drug (DrugB
156	ADRBK1	adrenergic, k	The product	GO:0003108	AGC Ser/Thr	Enzymes/{EN	Vesicles	Adenosine tr
121536	AEBP2	AE binding protein 2		GO:0000122	negative regu	Transcription factors/Zinc-coordinating		
166	AES	amino-termi	The protein	GO:2000210	positive regulation of anoi	Nucleus but not nucleoli		
10939	AFG3L2	AFG3-like A	This gene en	GO:0042407	crisaa forma	Enzymes/Pe	Mitochondri	Adenosine tr
25909	AHCTF1							
11214	AKAP13							
221120	ALKBH3							
	AMDHD							

GroupID	Category	Term	Description	LogP	InTerm_InLis	Genes
1_Summary	GO Biological	GO:0016032	viral process	-18.851	49/771	156,527,2033
1_Member	GO Biological	GO:0016032	viral process	-18.851	49/771	156,527,2033
1_Member	GO Biological	GO:0044764	multi-organi	-18.549	49/784	156,527,2033
1_Member	GO Biological	GO:0044419	interspecies	-18.086	50/838	156,527,1536
1_Member	GO Biological	GO:0044403	symbiosis, e	-18.086	50/838	156,527,1536

Custom Analysis for More Controls

C-A-M-E: a generic gene analysis workflow in four simple steps: ID Conversion, Annotation, Membership & Enrichment.

Input Gene Identifier	Conversion	Annotation			Membership	Other Input Columns
RefSeq Protein	Human Gene	Symbol	Description	Protein Function	Secreted	Membership: invasion Activity
NP_031938	1956	EGFR	epidermal growth	Enzymes/{ENZYM	Yes	0.39568
NP_001076021	6737	TRIM21	tripartite motif c			1 0.19248

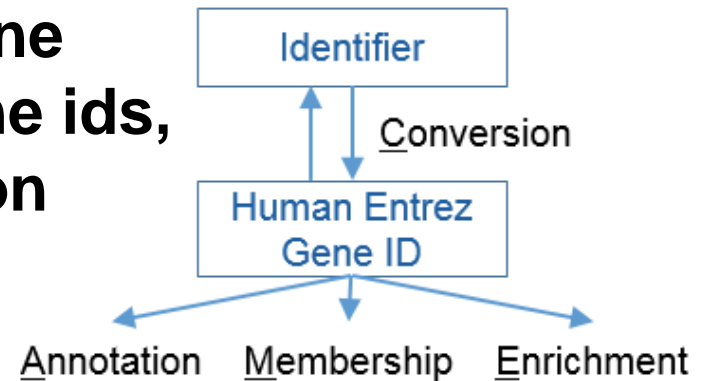
First convert input identifiers into human Entrez gene IDs, then extract annotation fields, flag memberships (optional) and apply function enrichment analysis.

Enrichment

Category	Term	Description	LogP	Symbols
Canonical Pathways	M113	PID NFAT 3PATHWAY	-3.746	KPNB1,PRKACA,XPO1,NUP214

ID Conversion

We first need to convert input gene identifiers into Entrez human gene ids, as all subsequent analyses rely on human gene IDs.



Apply Download Excel Reset Default Setting

ID Conversion Annotation Membership Enrichment

From current ID type : Best guess

To ID type : Human Homolog Gene ID

Highlight Input ID column, click "Apply" to convert IDs into Human Entrez Gene IDs

Input ID	Gene ID	Type	Tax ID	Homologene
ADRBK1	156	symbol	Human	156
AEBP2	121536	symbol	Human	121536
AES	166	symbol	Human	166
AFG3L2	10939	symbol	Human	10939
AHCTF1	25909	symbol	Human	25909
AKAP13	11214	symbol	Human	11214
ALKBH3	221120	symbol	Human	221120
AMDHD2	51005	symbol	Human	51005
ANAPC2	29882	symbol	Human	29882

Annotation

Extract annotation columns for the gene list, including gene descriptions, functions, protein classes, and whether the protein is secreted or transmembrane, etc.

Apply Download Excel Reset Default Setting

ID Conversion Annotation Membership

Check All Terms Uncheck All Terms

Description

- Gene Symbol
- Synonyms
- Type of Gene
- Description
- Gene Summary

Function/Location

- Biological Process (GO)
- Kinase Class (UniProt)
- Protein Function (Protein Atlas)
- Subcellular Location (Protein Atlas)

Check annotations of interest, click "Apply" to extract contents.

Homologene	Gene Symbol	Description	Gene Summary	Biological Process (GO)	Kinase Class (UniProt)	Protein Function (Protein Atlas)
156	ADRBK1	adrenergic, beta-1	The product of this gene is a member of the beta-1 adrenergic receptor family.	GO:0003108	AGC Ser/Thr Kinase	Enzymes/Enzymes
121536	AEBP2	AE binding protein 2	AE binding protein 2	GO:0000122	negative regulation of transcription	Transcription factors
166	AES	amino-terminal phosphatase	The protein encoded by this gene is a member of the phosphatase family.	GO:2000210	positive regulation of angiogenesis	Enzymes/Enzymes
10939	AFG3L2	AFG3-like AAF3 domain containing protein 2	This gene encodes a protein that is a member of the AFG3-like AAF3 domain containing protein family.	GO:0042407	crystal formation	Enzymes/Proteins
25909	AHCTF1	AT hook containing transcription factor 1	AT hook containing transcription factor 1	GO:0051292	nuclear pore complex assembly	Proteins
11214	AKAP13	A kinase (PRK) domain containing protein 13	The A-kinase anchoring protein 13	GO:1900169	regulation of glucocorticoid receptor activity	Proteins
221120	ALKBH3	alkB homolog 3	The Escherichia coli alkB homolog 3	GO:0035552	oxidative single-stranded DNA repair	Proteins
51005	AMDHD2	amidohydrolase domain containing protein 2	Amidohydrolase domain containing protein 2	GO:0006046	N-acetylglucosaminidase activity	Enzymes/Enzymes
29882	ANAPC2	anaphase promoting complex subunit 2	A large protein complex that is involved in the regulation of the cell cycle.	GO:0007094	mitotic spindle assembly checkpoint	Proteins
8906	AP1G2	adaptor-related protein 1 gamma 2	Adaptins are a family of proteins that are involved in the regulation of the cell cycle.	GO:0006892	post-Golgi vesicle-mediated transport	Proteins

Membership

Flag genes of interest, e.g., flag those, which fall under GO biological process terms that include "invasion" as a keyword.

Apply Download Excel Reset Default Setting

ID Conversion Annotation **Membership** Enrichment

Select Membership Category: **GO Biological Processes**

Search invasion

In Term Name In Term Name or In Description GO IDs

Check All Terms Uncheck All Terms

GO:0030260, entry into host cell
 GO:0051806, entry into cell of other organism involved in symbiotic interaction

Search ontology terms by keywords, click "Apply" to flag genes of interest.

Homologene	Membership	Membership	Gene Symbc	Description
156	1	GO:0051806, ADRBK1		adrenergic,
9972	1	GO:0051806, NUP153		nucleoporin
8021	0		NUP214	nucleoporin
10762	0		NUP50	nucleoporin

Enrichment

Identify pathways (or complexes, published hit lists, etc.) that has statistically significant p-values; automatically cluster them to reduce redundancy.

Enrichment Analysis

Min Overlap: 3 P Value Cutoff: 0.01 Min Enrichment: 1.5

- Functional Set
 - GO Molecular Functions (706)
 - KEGG Functional Set (1)
- Pathway
 - Reactome Gene Sets (318)
 - Canonical Pathways (128)

Pathway	GeneGo ID	p-value	Gene Symbols
Transport_Intracellular cholesterol transport	gMAP2687	-7.701 7/56	ARCN1 COPA COPB1 RAB5A COPB2
Transport_Intracellular cholesterol transport	gMAP2687	-7.701 7/56	ARCN1 COPG1
Transport_RAB1A regulation pathway	gMAP405	-6.896 5/23	ARCN1
Cholesterol and Sphingolipid transport / Transport from Golgi and ER to the apical membrane (normal)	gMAP2934	-6.797 5/24	ARCN1
wtCFTR and deltaF508-CFTR traffic / Generic schema (normal and CF)	gMAP2668	-6.358 6/54	ARCN1 COPA COPB1 RAB5A COPB2
wtCFTR traffic / ER-to-Golgi (normal)	gMAP2669	-6.283 5/30	ARCN1 COPA COPB1 COPB2 COPG1

Click for Term Report

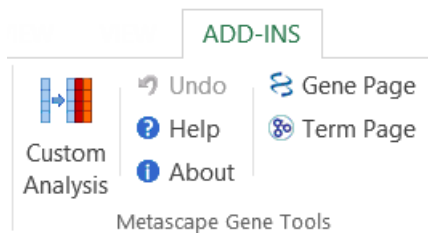
Mouse over gene symbol for quick annotation

Click to red highlight a gene

Gene Information

Synonyms	RAB5
Description	RAB5A, member RAS oncogene family
GeneGo Functional Class	Generic protein
GeneGo Brief Class	
GeneGo Location	cytoplasm
Gene Summary	

Metascope for Excel (M4X)



With Metascope tools installed on my Excel ribbon. No upload and download is needed.

Apply Reset Default Setting

	A	B
1	HumanGeneID	Synonyms
2	3630	IDDM IDDM1
3	7857	CHGC EM66 S
4	3309	BIP GRP78 HE
5	2495	FHC FTH FTH1
6	2512	LFTD NBIA3
7	6747	TRAPG
8	2512	LFTD NBIA3

ID Conversion Annotation M

Check All Terms Uncheck All T

Description

- Gene Symbol
- Synonyms
- Type of Gene
- Description

	A	B	C
	HumanGeneID	Description	Synonyms
	3630	insulin	IDDM IDDM1
	7857	secretogranin I	CHGC EM66
	3309	heat shock 70k	BIP GRP78 HE
	2495	ferritin, heavy	FHC FTH FTH1
	2512	ferritin, light p	LFTD NBIA3
	6747	signal sequenc	TRAPG
	2512	ferritin, light p	LFTD NBIA3

Step 1. Highlight input column

Step 2. Specify output in "Custom Analysis", click "Apply"

Step 3. New column(s) are automatically filled



Summary

- **Gene lists can be prioritized with a generic workflow: ID **C**onversion > **A**nnotation > **M**embership > **E**nrichment.**
- **Metascape.org** is a free, easy-to-use and well-maintained gene prioritization resource.



About Us

Give it a try for your next gene list and provide your comments to

[metascape.team at gmail.com](mailto:metascape.team@gmail.com)



<http://metascope.org>

GNF

Bin Zhou

Alireza Hadj Khodabakhshi

Olga Tanaseichuk

Yingyao Zhou

Salk

Max Chang

Chris Benner

Sanford Burnham Prebys

Andre Gatorano

Lars Pache

Sumit Chanda