

定量差异蛋白质组学数据分析处理方法

Data analysis of quantitative differential proteomics

组员：王舒心 白雪 李泽洋 鲍锦涛

汇报人：鲍锦涛

时间：2012. 06. 11

主要内容

- 蛋白质组学简介
- 定量差异蛋白质组学工作流程
- 定量差异蛋白质组学数据分析

组学（Omics）

组学指的是一类个体的集合。

包括：基因组学（Genomics），转录组学(transcriptomics)
蛋白组学（Proteomics），代谢组学（Metabolomics），转
录组学脂类组学（lipidomics），免疫组学（Immunomics），
糖组学（glycomics）和 RNA组学(RNomics)等。

蛋白质组与蛋白质组学

- **蛋白质组**指的是一个细胞，一个组织或一种生物的基因组所表达的全部蛋白质总和。
- 蛋白质组的研究特点：

全面性

+

动态性

+

整体性

- 蛋白质组学的研究领域：

表达蛋白质组学

定量蛋白质组学

蛋白质翻译后修饰

功能蛋白质组学

表达蛋白质组学

蛋白质组学简介

常用的定量蛋白质组学技术

SILAC技术

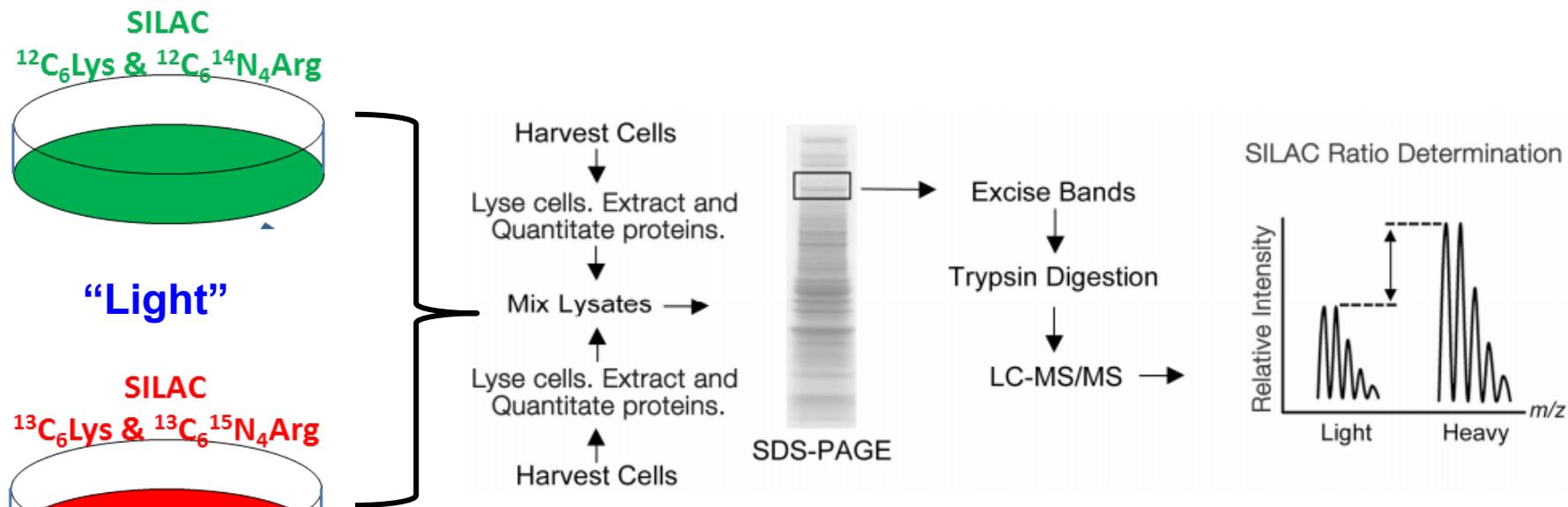
荧光差异电泳技术

SRM技术

iTRAQ标记技术

Label-free 技术

SILAC定量蛋白质组学工作流程



定量精确

高效性

高灵敏度

高通量

数据分析

原始数据初筛

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	id	PeptidOxidat	Oxidation	Protein	Peptide C	Peptide	Peptide	Only ide	Protein Names	Gene Names	Protein Uniprot	Uniprot Name	ENSEMBL	VEGA	MGI		
2	1487	457;7917;10118;10198		IPI003238	4;4;4;1	4;4;4;1	4;4;4;1	FALSE	Putative unchar	Rpl24	Putative	E9QNJ0;Q3UW40;Q8BP67Q3UW40_MOUSE;RL24_MOUSE;Q	ENSMUSTO	(OTTMUSPO	(MGI:1		
3	893	3217;5837;9061;9459;1008		IPI002245	6;6;6;5;5	6;6;6;5;5	6;6;6;5;5	FALSE	60S ribosomal p	Rpl13	60S ribo:	P47963;D3YX54;Q5RKP3RL13_MOUSE;Q5RKP3_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1		
4	806	2759;5094;8844;11482		IPI001324	4;4;3;3;2;4;4;3;3;4;4;3;3;	4;4;3;3;2;4;4;3;3;	4;4;3;3;2;4;4;3;3;	FALSE	60S ribosomal p	Rpl26;RP23-39	60S ribo:	P61255;Q3U7N1;Q4FZH2RL26_MOUSE;B1ARA3_MOUSE;B	ENSMUSTO	(OTTMUSPO	(MGI:1		
5	699	2760;9477		IPI001297	2	2	2	FALSE	WD repeat and S	Wdsof1;Gm83	DDB1- an	Q6PAC3	SOF1_MOUSE	ENSMUSTO	0000002290	MGI:2	
6	1211	168;10251		IPI002297	2;2;2	2;2;2	2;2;2	FALSE	Polymerase delt	Poldip3;RP23-	Polymera:	Q8BG81;Q3TJ78;Q3U8V1PDIP3_MOUSE;Q3UDD3_MOUSE;	ENSMUSTO	(OTTMUSPO	(MGI:1		
7	1694	3115;5086;7153;8254;1112		IPI004202	5;5;4;4;4;5;5;4;4;4;3;3;2;3;	5;5;4;4;4;5;5;4;4;4;3;3;2;3;	5;5;4;4;4;5;5;4;4;4;3;3;2;3;	FALSE	High mobility g	Hmgbl1;Hmg-1;H	High mob:	P63158;Q3UBK2;Q3UJKCHMGB1_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:9		
8	485	10905;10967;11435		IPI001224	3;3;2;2;1;1;3;3;2;2;3;3;2;2;	3;3;2;2;1;1;3;3;2;2;3;3;2;2;	3;3;2;2;1;1;3;3;2;2;3;3;2;2;	FALSE	60S ribosomal p	Rpl27	60S ribo:	P61358;Q5BLJ9;Q9CSM4RL27_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:9		
9	1638	113;290;434;1861;2374;28		IPI008577	24;24;24;6;24;24;24	24;24;24	24;24;24	FALSE	MKIAA4075 prote	Chd4;mKIAA407	Uncharac:	E9QAS5;Q5DTP7;Q6PDQ2Q5DTP7_MOUSE;CHD4_MOUSE;Q	ENSMUSTO	(OTTMUSPO	(MGI:1		
10	1927	8790		IPI006785	1	1	1	FALSE	Regulator of mi	Fam82a2;Fam82	Regulato:	Q3UJU9	RMD3_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1	
11	1705	3945;4217;4385;5463;5550		IPI004207	9;7;7;7;1;1;9;7;7;7;9;7;7;7;	9;7;7;7;1;1;9;7;7;7;9;7;7;7;	9;7;7;7;1;1;9;7;7;7;9;7;7;7;	FALSE	40S ribosomal p	Rps9	40S ribo:	Q6ZWN5;Q96ECO;Q9CXW7RS9_MOUSE;Q96ECO_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1		
12	597	3399;3487;3493;5839;8689		IPI001259	5;5;5;4;4	5;5;5;4;4	5;5;5;4;4	FALSE	40S ribosomal p	Rps13	40S ribo:	P62301;Q5BLJ7;Q921R2RS13_MOUSE;Q921R2_MOUSE	ENSMUSTO	0000008504	MGI:1		
13	506	800;3594;6142;6143		IPI001230	4;4;2;2	4;4;2;2	4;4;2;2	FALSE	Glutamate-rich	Grwd1;A301	Glutamat:	Q810D6;Q5XJZ3;F7CPC6GRWD1_MOUSE;Q8BIF7_MOUSE;	ENSMUSTO	(OTTMUSPO	(MGI:2		
14	1471	1640;2843;5602;6836;7188		IPI003227	7	7	7	FALSE	Transcriptional	Atrx;Hplbp2;X	Transcri:	Q61687;Q3TP53;Q3TSV4ATRX_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1		
15	1346	345;844;1798;2538;2539;4		IPI003112	11;9;7;5	11;9;7;5	10;8;7;4	FALSE	60S ribosomal p	Rpl17	60S ribo:	P14148;Q3TK73;Q3UBI6RL7_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:9		
16	801	323;6042;6449;6519;8826		IPI001324	5	5	5	FALSE	Probable dimeth	Dimt1;Dimt1	Probable	Q9D0D4;Q3UK38	DIMT1_MOUSE	ENSMUSTO	0000002220	MGI:1	
17	1244	2620;3619;4006;9166		IPI002306	4;4;1;1	4;4;1;1	4;4;1;1	FALSE	60S ribosomal p	Rpl32	60S ribo:	P62911;Q5PR09	RL32_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:9	
18	1050	4746;5160;5537;6441;8639		IPI001872	6;6;1	6;6;1	6;6;1	FALSE	Probable ATP-de	Ddx47	Probable	Q9CWX9;Q3TEK1;Q4VBG1DDX47_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1		
19	1306	835;1230;1247;2205;5503;		IPI003423	10;10;10;1	10;10;10	10;10;10	FALSE	Exosome compone	Exosc10;Pmscl	Exosome	(P56960;B1AVQ3;Q3UNK6EXOSX_MOUSE;B2KFPQ9_MOUSE;	ENSMUSTO	(OTTMUSPO	(MGI:1		
20	1101	207;1060;7206;11263		IPI002232	4;3;3;2;1;1;4;3;3;2;4;3;3;2;	4;3;3;2;1;1;4;3;3;2;4;3;3;2;	4;3;3;2;1;1;4;3;3;2;4;3;3;2;	FALSE	60S ribosomal p	Rpl13a;P198;T	60S ribo:	P19253;Q5M9M0;E9Q5ACRL13A_MOUSE;Q3TDS9_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1		
21	77	7726;10350		IPI002298	2;2;2;2;2	2;2;2;2;2	2;2;2;2;2	FALSE	Transformation	Trp53bp1;RP23	tumor su:	A2AU91;Q3V1P7;Q8BZ87A2AU91_MOUSE;TP53B_MOUSE;	ENSMUSTO	(OTTMUSPO	(MGI:1		
22	1261	7941;10538		IPI002638	2;1	2;1	2;1	FALSE	60S ribosomal p	Rpl35	60S ribo:	Q6ZWW7	RL35_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1	
23	1553	341;910;3056;3983;3984		IPI005550	5;5;4;4;3	3;3;3;3;3	3;3;3;3;3	FALSE	Histone H2A.V;	H2afv;H2av;H2	Histone	IQ3THW5;B2RVP5;POCOS6H2AV_MOUSE;H2AZ_MOUSE;Q5N	ENSMUSTO	(OTTMUSPO	(MGI:1		
24	1782	2286;5187;11378		IPI004744	3;3;3;3;3	3;3;3;3;3	3;3;3;3;3	FALSE	60S ribosomal p	Rpl36;mcG_203	60S ribo:	D3YW41;Q5M9L1;Q6ZWZ4Q5M9L1_MOUSE;Q6ZWZ4_MOUSE	ENSMUSTO	0000008049	MGI:1		
25	1774	4237;4506;7153;11125		IPI004622	4	2	2	FALSE	High mobility g	Hmg2;Hmg2	High mob:	P30681;Q3U566;Q3UAZ7HMG2_MOUSE	ENSMUSTO	0000006792	MGI:9		
26	1777	1786;2084;2828;2924;3052		IPI004625	12;12	12;12	12;12	FALSE	LAS1-like (S. c	Las1;RP23-47	Isoform	A2BE28-1;A2BE28;A9X4A2BE28_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:1		
27	1767	5149;5335;5422;10624		IPI008749	4;4;4;4;4	4;4;4;4;4	4;4;4;4;4	FALSE	60S ribosomal p	Rpl23a	Uncharac:	D3YWP3;D3Z1D6;P62751RL23A_MOUSE	ENSMUSTO	(OTTMUSPO	(MGI:3		
28	1393	9663;10966		IPI009897	2;2;2;2;2	2;2;2;2;2	2;2;2;2;2	FALSE	60S ribosomal p	Rpl21;mcG_120	60S ribo:	D3YWP4;D3YX83;D3Z4H5RL21_MOUSE;Q9CQM8_MOUSE;Q	ENSMUSTO	(OTTMUSPO	(MGI:1		

◆ 筛选出关键数据，清除Contaminant等分析障碍

数据分析

➤ 原始数据初筛

	A	B	C	D	E	F
1	Uniprot	Uniprot Name	Gene Names	Ratio H/L Normalized Run 1	Ratio H/L Normalized Run 2	Ratio H/L Normalized Significance F
2	E9QNJ0	Q3UW40_MOUSE	Rpl24	11.227	10.514	3.488900567
3	P47963	RL13_MOUSE	Rpl13	9.784	8.4863	3.290424404
4	P61255	RL26_MOUSE	Rpl26	9.284	6.0764	3.214746523
5	Q6PAC3	SOF1_MOUSE	Wdsof1	8.533	3.575	3.093053048
6	Q8BG81	PDIP3_MOUSE	Poldip3	8.4302	4.3213	3.075566858
7	P63158	HMGB1_MOUSE	Hmgb1	8.2379	8.0404	3.042276613
8	P61358	RI17_MOUSE	Rn17	8.9116	13.903	3.037683353

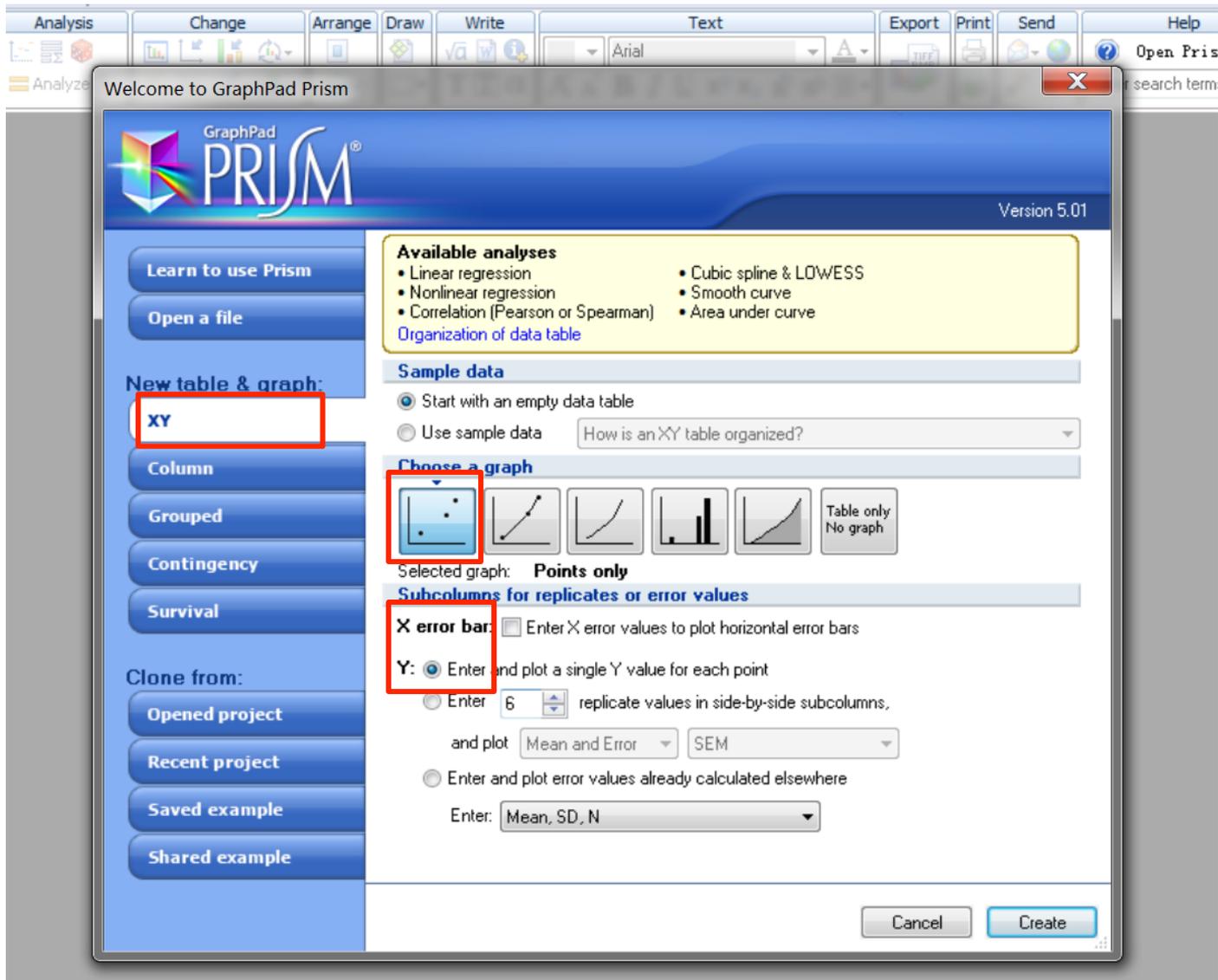
◆ 保留关键条目，如Uniprot AC 等

◆ 找出两次重复都出现的蛋白（利用Excel VLOOKUP & IF函数）

◆ 根据两次重复的平均Ratio值排序删去(0.5,2) OR (0.67,1.5) 对应结果

数据分析

➤ 散点图重复性分析



数据分析

➤ 散点图重复性分析

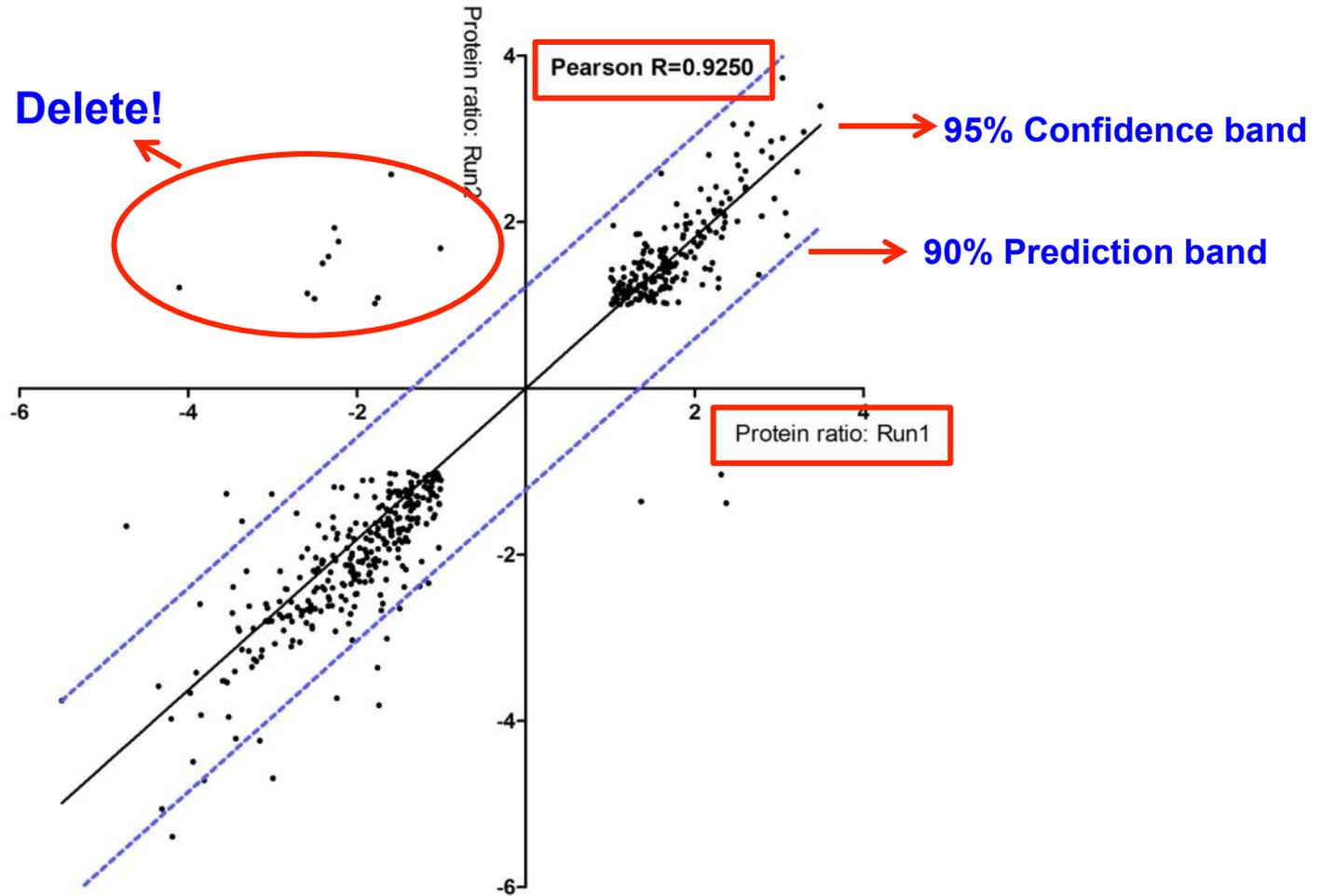
The screenshot displays the GraphPad Prism interface. The top menu bar includes 'Prism', 'File', 'Sheet', 'Undo', 'Clipboard', 'Analysis', 'Change', 'Import', 'Draw', 'Write', 'Text', 'Export', 'Print', 'Send', and 'Help'. The 'Analysis' menu is open, and the 'Analyze' option is highlighted with a red box. Below the menu is a data table with columns X, A, B, C, D, E, F, G, H. The 'X' and 'A' columns are labeled 'LOG2 Ratio H/L Normalized'. The 'Y' column is labeled 'Y'. The data table contains 26 rows of data. The 'Analyze Data' dialog box is open, showing a list of analysis options. The 'Correlation' option under 'XY analyses' is highlighted with a red box and labeled '2.'. The 'Analyze which data sets?' section shows 'A:LOG2 Ratio H/L Normalized' selected. The dialog box also includes 'Select All', 'Deselect All', 'Help', 'Cancel', and 'OK' buttons.

	X	A	B	C	D	E	F	G	H
	LOG2 Ratio H/L Normalized	LOG2 Ratio H/L Normalized	Title	Title	Title	Title	Title	Title	Title
	X	Y	Y	Y	Y	Y	Y	Y	Y
1	E9QNJ0	3.488901	3.394240						
2	P47963	3.290424	3.085136						
3	P61255	3.214746	2.603217						
4	Q6PAC3	3.093053	1.837943						
5	Q8BG81	3.075567	2.111465						
6	P63158	3.042277	3.007267						
7	P61358	3.037663	3.732595						
8	Q3UJU9	2.942683	2.281134						
9	Q6ZWN5	2.909178	2.772583						
10	P62301	2.904715	2.971295						
11	Q810D6	2.795081	2.852578						
12	Q61687	2.793563	2.070904						
13	P14148	2.759007	1.365245						
14	Q9D0D4	2.674867	3.181087						
15	P62911	2.621642	3.059442						
16	Q9CWX9	2.601744	2.613767						
17	P56960	2.601197	2.421883						
18	P19253	2.597412	2.391465						
19	Q3THW5	2.548856	2.514400						
20	D3YW41	2.512707	2.684684						
21	A2BE28-1	2.506424	2.009813						
22	D3YWP3	2.498430	2.810752						
23	D3YWP4	2.454886	3.174630						
24	Q80X41-1	2.414352	2.278490						
25	Q7TNV0	2.374372	2.358115						
26	P62242	2.358059	1.994109						

◆ 取Ratio值的对数进行相关性分析

数据分析

➤ 散点图重复性分析



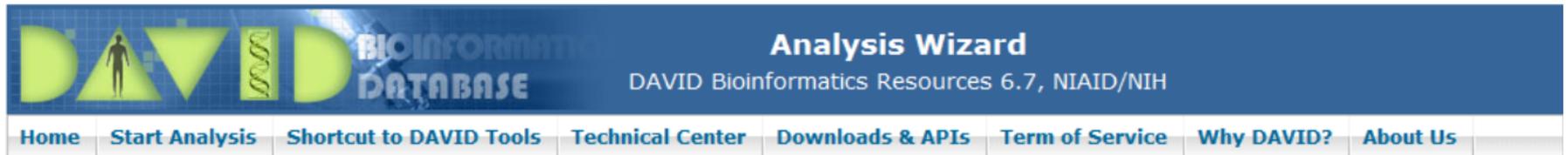
◆ 重复组相关性分析后得到最终“可信”的数据

数据分析

➤ 高通量组学数据库分析——KOBAS 2.0

KOBAS 2.0

<http://kobas.cbi.pku.edu.cn/home.do>



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



<http://genecodis2.dacya.ucm.es/analysis/>

数据分析

➤ 高通量组学数据库分析——KOBAS 2.0

KOBAS 2.0

Run KOBAS 2.0

Annotate

Identify

Advanced KOBAS

User Space

Analysis history

Download

Download

Help

Tutorial

Feedback

Contact

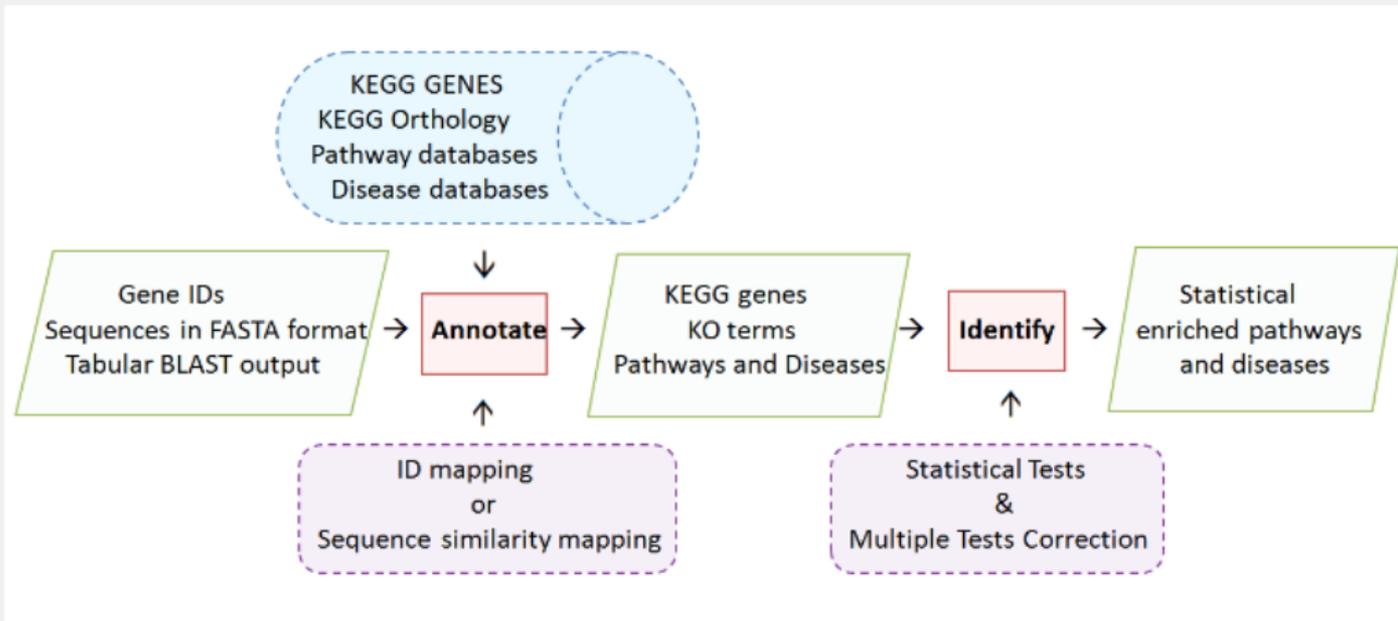
Users

My Account

Logout

Welcome to KOBAS 2.0

✦ KOBAS 2.0 is an update of KOBAS (KEGG Orthology Based Annotation System). Its purpose is to identify statistically enriched related pathways and diseases for a set of genes or proteins, using pathway and disease knowledge from multiple commonly used databases.



Updates of KOBAS 2.0

- ✦ KOBAS 2.0 includes [KEGG PATHWAY](#), [PID Curated](#), [PID BioCarta](#), [PID Reactome](#), [BioCyc](#), [Reactome](#) and [Panther](#) pathway databases.
- ✦ KOBAS 2.0 includes [KEGG DISEASE](#), [GAD](#), [FunDO](#), [OMIM](#) and [NHGRI](#) human disease databases.
- ✦ KOBAS 2.0 can annotate queries to both [KEGG GENES](#) and [KEGG Orthology](#) (KO) terms.

➤ 高通量组学数据库分析——KOBAS 2.0

- Annotate
- Identify
- Advanced KOBAS
- User Space
- Analysis history
- Download
- Download
- Help
- Tutorial
- Feedback
- Contact
- Users
- My Account
- Logout

Function

The types of queries can be a set of protein or nucleotide sequences in FASTA format, the tabular format output of blast program, or a list of IDs (can be Entrez Gene ID, UniProtKB AC, or GI). Given queries inputted by users, Annotate assigns KEGG GENES or KO terms based on sequence similarity search, parsing blast output or ID mapping. If the type of input file is sequence, the maximum number of sequences is 500. If you want to annotate more sequences, you need to download sequences file of the desired species and run BLAST locally. Then do annotation on KOBAS with the tabular output file of BLAST. And you also can download standalone version to run locally.

After clicking the "Example" hyperlink, all the form will be filled automatically, and you can execute the program by simply clicking the 'Run' button.

[FASTA example](#)

[Blasttab example](#)

[Gene id example](#)

Input

Use which file source?

A file on the server

Paste from clipboard

Upload a file from local disk

File type:

UniProtKB AC

File type

Paste the sequence/ID lists into this box:

P10126
Q02788
Q8BH47
Q8BFZ3
P99024
Q922Q8
Q04857
A8DUV1
F6TJZ4
P68369
P60843
P15864

To which species do you want to map these genes? (If you want to map the genes to KO terms, please select 'KO'):

H.sapiens

Species

Available database(s) for this species:

Ontology database

Pathway databases

Disease databases

KEGG PATHWAY PID Curated PID BioCarta PID Reactome
BioCyc Reactome Panther

KEGG DISEASE GAD FunDO OMIM NHGRI

Databases

Output

Save result in directory:

Work Directory

Output file:

Untitled.annotate

Save Results

Run

➤ 高通量组学数据库分析——KOBAS 2.0

Result of file: ABCçî³îË¾Äý.annotate

DOWNLOAD... HELP **USE THIS FILE AS IDENTIFY'S SAMPLE INPUT**

Direct “identify” sample input

RAW CONTENT **TABLE VIEW**

339 succeed, 150 fail

Query	Gene ID	Gene Name
E9QNJ0 (details)	None	None
P47963 (details)	mmu:100039656	Gm12918, OTTMUSG00000009186
	mmu:270106	Rpl13, L13
	mmu:100040605	Gm15710, OTTMUSG000000025606
P61255 (details)	mmu:19941	Rpl26, SIG-20
Q6PAC3 (details)	mmu:223499	Dcaf13, Gm83, MGC65348, Wdsof1
Q8BG81 (details)	mmu:73826	Poldip3, 1110008P04Rik, AA408269, AL022852, C77954, KIAA1649, PDIP46, SKAR, mKIAA1649
P63158 (details)	mmu:15289	Hmgb1, DEF, HMG-1, Hmg1, MGC103168, MGC103169, MGC117896, MGC117897, SBP-1, amphoterin, p30
P61358 (details)	mmu:19942	Rpl27, MGC103134
Q3UJU9 (details)	mmu:67809	Fam82a2, 1200015F23Rik, A1131757, Ptpip51, RMD-3, Rmd3
Q6ZWN5 (details)	mmu:76846	Rps9, 3010033P07Rik, AL022771, AL022885, MGC4138
P62301 (details)	mmu:668239	Gm12270, OTTMUSG00000005828
	mmu:68052	Rps13, 2700063M04Rik, MGC102403, MGC118043
Q810D6 (details)	None	None

KEGG database

Displaying 1 ~ 20 of 514 Page 1 of 26

◆ 通过“Annotate”链接到KO获取条目注释信息，并将结果保存到“User space”

➤ 高通量组学数据库分析——KOBAS 2.0

Run KOBAS 2.0

Annotate

Identify

Advanced KOBAS

User Space

Analysis history

Download

Download

Help

Tutorial

Feedback

Contact

Users

My Account

Logout

Identify - Identifies enriched pathway or human disease terms

Frequently occurring or statistically significantly enriched pathway or human disease terms are identified by frequencies of terms or statistical significance of terms. The sample file is the output of Annotate. The background can be either whole gene set of a species or Annotate result of another gene set (e.g. all probe sets on a microarray).

After clicking the "Example" hyperlink, all the fields will be filled automatically, and you can execute the program simply by clicking the "Run" button.

[Example](#) (This input in text field is the output of Annotate Gene id example, except we truncated part not needed for computation)

Sample file (result of Annotate):

|-ABC?????????.annotate

Show available databases according to the species used in Sample Input

Species and Databases

The species' name is M.musculus

Please select the database(s) as your search range of pathways/diseases.

Note: The Corrected PValue in the result will be affected by the number of databases selected.

KEGG PATHWAY PID Curated PID BioCarta PID Reactome

BioCyc Reactome Panther

KEGG DISEASE GAD FunDO OMIM NHGRI

Check all

Clear all

Background (defined by user or default)

Annotate result of another gene set (If no annotate file provided, KOBAS will use genes from whole genome as default background):

Options for statistics

Output

Save result in directory:

Work Directory

Output file:

ABC?????????.identify

Run

➤ 高通量组学数据库分析——KOBAS 2.0

Result of file: ABC?????????.identify

DOWNLOAD... HELP

Pathway **Targets involved** **P-value**

RAW CONTENT TABLE VIEW (FOR PATHWAY IDENTIFICATION RESULT) TABLE VIEW (FOR DISEASE IDENTIFICATION RESULT)

Term	Database	ID	Sample Number (click to sort; click again to toggle sorting direction)	Background Numb...	PValue (click to sort; click again to toggle sorting direction)	Corrected PValue (click to sort; click again to toggle sorting direction. Cannot sort when the value is 'Non...
Metabolism of proteins	Reactome	None	46	258	3.99680288865e-15	1.08713038571e-12
3' -UTR-mediated translational regulation	Reactome	None	29	122	3.74922315416e-13	5.09894348966e-11
Ribosome	KEGG PATHWAY	mmu03010	21	122	7.36035676852e-11	6.67339013679e-9
Citrate cycle (TCA cycle)	KEGG PATHWAY	mmu00020	11	31	7.0935990415e-10	4.82364734822e-8
Ribosome biogenesis in eukaryotes	KEGG PATHWAY	mmu03008	17	91	1.28347044104e-9	6.98207919925e-8
Valine, leucine and isoleucine degradati...	KEGG PATHWAY	mmu00280	11	51	2.32735897776e-7	0.0000105506940325
RNA transport	KEGG PATHWAY	mmu03013	19	172	9.12417783994e-7	0.000035453948178
Fatty acid metabolism	KEGG PATHWAY	mmu00071	10	50	0.00000173333706721	0.0000589334602852
Gene Expression	Reactome	None	50	529	0.00000266347180933	0.0000804960369041
Protein processing in endoplasmic retic...	KEGG PATHWAY	mmu04141	17	172	0.0000151522846237	0.000412142141764
Butanoate metabolism	KEGG PATHWAY	mmu00650	7	31	0.000027701285477	0.000684977240886
fatty acid beta-oxidation I	BioCyc	FAO-PWY	8	32	0.0000598549786922	0.00135671285036
Parkinson's disease	KEGG PATHWAY	mmu05012	14	151	0.000167263648776	0.00349967018978

Displaying 1 - 20 of 272 Page 1 of 14

◆ 通过“Identify”获得处于不同通路中的Targets clusters

➤ 高通量组学数据库分析——DAVID

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home **Start Analysis** Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

Shortcut to DAVID Tools

- Functional Annotation**
Gene-annotation enrichment analysis, functional annotation clustering, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and [more](#)
- Gene Functional Classification**
Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)
- Gene ID Conversion**
Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)
- Gene Name Batch Viewer**
Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched

Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.7

2003 - 2012

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an [update to the sixth version](#) of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- ☑ Identify enriched biological themes, particularly GO terms
- ☑ Discover enriched functional-related gene groups
- ☑ Cluster redundant annotation terms
- ☑ Visualize genes on BioCarta & KEGG pathway maps
- ☑ Display related many-genes-to-many-terms on 2-D view.
- ☑ Search for other functionally related genes not in the list
- ☑ List interacting proteins

What's Important in DAVID?

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

Statistics of DAVID

DAVID Bioinformatic Resources Citations

➤ 高通量组学数据库分析——DAVID

Upload List Background

Upload Gene List

[Demolist 1](#) [Demolist 2](#)

[Upload Help](#)

Step 1: Enter Gene List

A: Paste a list

Q99Q2D9
O35218
P28658
O80W00-1

Clear

B: Choose From a File

Multi-List File ?

Step 2: Select Identifier

UNIPROT_ACCESSION

Step 3: List Type

Gene List

Background

Step 4: Submit List

Analysis Wizard

[Tell us how you like the tool](#)

[Contact us for questions](#)

← Step 1. Submit your gene list through left panel.

An example:

Copy/paste IDs to "box A" -> Select Identifier as "Affy_ID" -> List Type as "Gene List" -> Click "Submit" button

1007_s_at
1053_at
117_at
121_at
1255_g_at
1294_at
1316_at
1320_at
1405_i_at
1431_at
1438_at
1487_at
1494_f_at
1598_g_at

➤ 高通量组学数据库分析——DAVID

Upload List Background

Upload Gene List

Demolist 1 Demolist 2

Upload Help

Step 1: Enter Gene List

A: Paste a list

Or

B: Choose From a File

Multi-List File ?

Step 2: Select Identifier

AFFYMETRIX_3PRIME_IVT_ID ▼

Gene ID Conversion Tool

[Help and Tool Manual](#)

You are either not sure which identifier type your list contains, or less than 80% of your list has mapped to your chosen identifier type. Please use the Gene Conversion Tool to determine the identifier type.

Option 1 (Recommended):

Option 2: Convert the gene list to DAVID (Default) ▼

Option 3:

➤ 高通量组学数据库分析——DAVID

Home Start Analysis Shortcut to DAVID Tools Technical Center Downloads & APIs Term of Service Why DAVID? About Us

Upload List Background

Gene List Manager

Select to limit annotations by one or more species
[Help](#)

- Use All Species -
Mus musculus (369)
Unknown (121)

Select Species

List Manager [Help](#)

List_1

Select List to:

Use Rename
Remove Combine
Show Gene List

[View Unmapped Ids](#)

Analysis Wizard

[Tell us how you like the tool](#)
[Contact us for questions](#)

Step 1. Successfully submitted gene list
Current Gene List: List_1
Current Background: Mus musculus

Step 2. Analyze above gene list with one of DAVID tools

[Which DAVID tools to use?](#)

- [Functional Annotation Tool](#)
 - [Functional Annotation Clustering](#)
 - [Functional Annotation Chart](#)
 - [Functional Annotation Table](#)
- [Gene Functional Classification Tool](#)
- [Gene ID Conversion Tool](#)
- [Gene Name Batch Viewer](#)

Annotation & Reduce redundancy

Major gene functional groups

➤ 高通量组学数据库分析——DAVID

Annotation Summary Results

[Help and Tool Manual](#)

Current Gene List: List_1

369 DAVID IDs

Current Background: Mus musculus

Check Defaults

Clear All

- Functional_Categories (0 selected)
- Gene_Ontology (0 selected)
- General_Annotations (0 selected)
- Literature (0 selected)
- Main_Accessions (0 selected)
- Pathways (1 selected)

<input type="checkbox"/> BIOCARTA	9.5%	35	Chart	
<input type="checkbox"/> EC_NUMBER	30.4%	112	Chart	
<input checked="" type="checkbox"/> KEGG_PATHWAY	47.4%	175	Chart	
<input type="checkbox"/> PANTHER_PATHWAY	16.8%	62	Chart	
<input type="checkbox"/> REACTOME_PATHWAY	0.3%	1	Chart	

- Protein_Domains (0 selected)
- Protein_Interactions (0 selected)
- Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering

Functional Annotation Chart

Functional Annotation Table

Reduce redundancy

More details

➤ 高通量组学数据库分析——DAVID

Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List_1

Current Background: Mus musculus

369 DAVID IDs

Options

Classification Stringency High

Rerun using options

Create Sublist

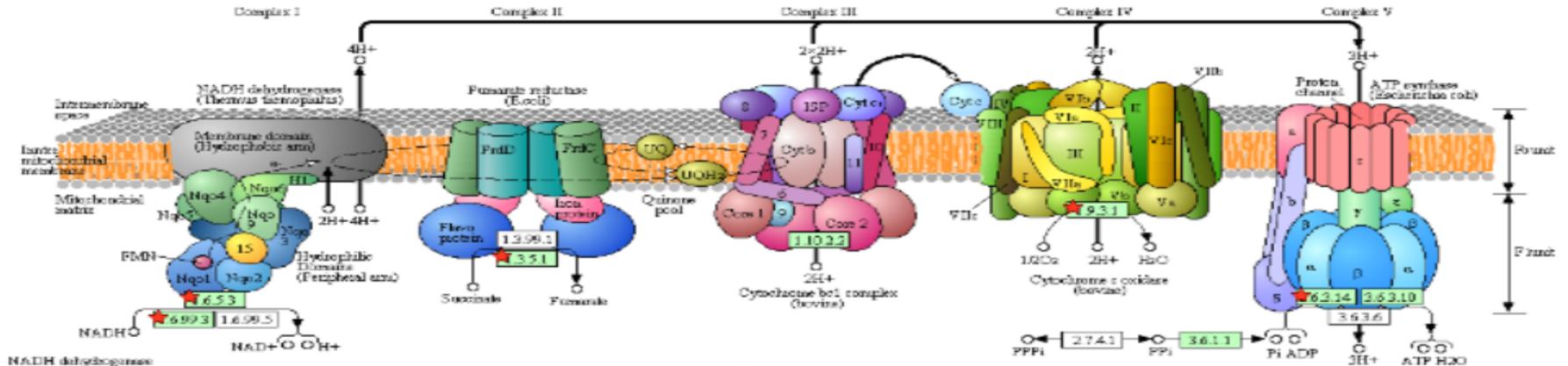
3 Cluster(s)

[Download File](#)

Annotation Cluster 1		Enrichment Score: 3.93			Count	P_Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	valine, leucine and isoleucine degradation	RT		11	8.0E-7	3.1E-5
<input type="checkbox"/>	KEGG_PATHWAY	Fatty acid metabolism	RT		10	6.1E-6	1.7E-4
<input type="checkbox"/>	KEGG_PATHWAY	Fatty acid elongation in mitochondria	RT		4	1.4E-3	1.3E-2
<input type="checkbox"/>	KEGG_PATHWAY	beta-Alanine metabolism	RT		4	2.8E-2	1.6E-1
Annotation Cluster 2		Enrichment Score: 3.33			Count	P_Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	Parkinson's disease	RT		16	9.9E-6	2.3E-4
<input type="checkbox"/>	KEGG_PATHWAY	Huntington's disease	RT		17	1.2E-4	1.9E-3
<input type="checkbox"/>	KEGG_PATHWAY	Oxidative phosphorylation	RT		12	1.8E-3	1.6E-2
<input type="checkbox"/>	KEGG_PATHWAY	Alzheimer's disease	RT		12	2.2E-2	1.4E-1
Annotation Cluster 3		Enrichment Score: 1.8			Count	P_Value	Benjamini
<input type="checkbox"/>	KEGG_PATHWAY	Butanoate metabolism	RT		8	1.0E-4	1.9E-3
<input type="checkbox"/>	KEGG_PATHWAY	Propanoate metabolism	RT		6	1.9E-3	1.5E-2
<input type="checkbox"/>	KEGG_PATHWAY	beta-Alanine metabolism	RT		4	2.8E-2	1.6E-1

高通量组学数据库分析——DAVID

OXIDATIVE PHOSPHORYLATION



Targets involved

NADH dehydrogenase

E	ND1	ND2	ND3	ND4	ND4L	ND5	ND6							
E	Ndufa1	Ndufa2	Ndufa3	Ndufa4	Ndufa5	Ndufa6	Ndufa7	Ndufa8	Ndufa9	Ndufa10	Ndufa11	Ndufa12	Ndufa13	
B/A	NuoA	NuoB	NuoC	NuoD	NuoE	NuoF	NuoG	NuoH	NuoI	NuoJ	NuoK	NuoL	NuoM	NuoN
B/A	NdhA	NdhB	NdhC	NdhD	NdhE	NdhF	NdhG	NdhH	NdhI	NdhJ	NdhK	NdhL	NdhM	NdhN
E	Ndufa1	Ndufa2	Ndufa3	Ndufa4	Ndufa5	Ndufa6	Ndufa7	Ndufa8	Ndufa9	Ndufa10	Ndufa11	Ndufa12	Ndufa13	
E	Ndufb1	Ndufb2	Ndufb3	Ndufb4	Ndufb5	Ndufb6	Ndufb7	Ndufb8	Ndufb9	Ndufb10	Ndufb11	Ndufb12	Ndufb13	

Succinate dehydrogenase / Fumarate reductase

E	SDHD	SDHE	SDHA	SDHB		
B/A	SdhC	SdhD	SdhA	SdhB		
			FrdA	FrdB	FrdC	FrdD

Cytochrome c oxidase

E	COX10	COX3	COX1	COX2	COX4	COX5A	COX5B	COX6A	COX6B	COX6D	COX7A	COX7B	COX7C	COX8	COX11	COX15	COX17
B/A	CycE	CycD	CycC	CycB	CycA	CoxD	CoxC	CoxA	CoxB	CoxD	CoxC	CoxB	CoxA				

Cytochrome c oxidase, cbb3-type

D	I	II	IV	III
---	---	----	----	-----

Cytochrome bc1 complex

B/A	CybA	CybB
-----	------	------

F-type ATPase (Bacteria)

bets	alpha	gamma	delta	epsilon	c	a	b
------	-------	-------	-------	---------	---	---	---

F-type ATPase (Eukaryotes)

beta	alpha	gamma	CSDF	delta	epsilon	c	a
b	e	SS	f	B			
d	r	s	i	k	g		

V-type ATPase (Fungi)

A	B	C	D	E	F	I	K
---	---	---	---	---	---	---	---

V-type ATPase (Eukaryotes)

A	B	C	D	E	F	G	H
I	AC3P	S4D	S1	3pd			

2019/04/22
© Komeilis Laboratories

genes analyzed in ad

VID Gene Name

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_beta-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_alpha-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_gamma-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_delta-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_epsilon-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_c-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_a-subunit](#)

[Fxn088g_Hs-transcriptome_mitochondrial_P1-sample_b-subunit](#)

➤ 高通量组学数据库分析——DAVID

Step 2. Analyze above gene list with one of DAVID tools

Which DAVID tools to use?

- Functional Annotation Tool
 - Functional Annotation Clustering
 - Functional Annotation Chart
 - Functional Annotation Table
- Gene Functional Classification Tool** Major gene functional groups

Gene Functional Classification Result

Current Gene List: List_1
Current Background: Mus musculus
369 DAVID IDs

Options: Classification Stringency Medium
Rerun using options Create Sublist

Groups	Score	Related Genes
Gene Group 1	Enrichment Score: 25.8	RG
1		mitochondrial ribosomal protein S28
2		predicted gene 8680 ; predicted gene 9663 ; predicted gene 7434 ; predicted gene 6251 ; predicted gene 6520 ; ribosomal protein L32
3		mitochondrial ribosomal protein S31
4		predicted gene 4833 ; predicted gene 9616 ; predicted gene 13680 ; predicted gene 4859 ; predicted gene 8729 ; predicted gene 8231 ; ribosomal protein S4, X-linked
5		predicted gene 4754 ; ribosomal protein L7 ; predicted gene 12174 ; LOC100049084 ; predicted gene 6543 ; predicted gene 7059 ; predicted gene 4734
6		ribosomal protein L26 ; predicted gene 15772 ; predicted gene 9197 ; predicted gene 13268 ; predicted gene 11826 ; predicted gene 8253 ; predicted gene 7606
7		mitochondrial ribosomal protein S26
8		predicted gene 14217 ; predicted gene 7424 ; predicted gene 8934 ; ribosomal protein L5 ; similar to ribosomal protein L5 ; similar to 60S ribosomal protein L5 ; predicted gene 15616 ; predicted gene 15500 ; predicted gene 7625
9		predicted gene 6599 ; predicted gene 6199 ; predicted gene 6341 ; predicted gene 6301 ; predicted gene 11518 ; similar to ribosomal protein L27 ; ribosomal protein L27 ; predicted gene 7053 ; predicted gene 11552 ; predicted gene 15730
10		similar to 60S ribosomal protein L13a (Transplantation antigen P198) (Tum-P198 antigen) ; ribosomal protein L13A ; similar to 60S ribosomal protein L13a ; predicted gene 6533
11		mitochondrial ribosomal protein S7

◆ DAVID 综合多种数据库满足我们对结果的分类和注释功能的需求

➤ 高通量组学数据库分析——GENECODIS

Home | Analysis | Web Services | Help | Release info | Other GeneCodis sites



gene annotations
co-occurrence discovery

1. Select the organism [?]

Mus musculus

Species

2. Select the annotations [?] [Last update on May, 2011]

Annotations

- GO Biological Process
- GO Molecular Function
- GO Cellular Component
- GOSlim Process
- GOSlim Function
- GOSlim Component
- KEGG Pathways
- InterPro Motifs
- MicroRNA
- Transcription Factors

Databases

“GOSlim”

GO Levels

- Lowest Level
- Level 3
- Level 4
- Level 5
- Level 6
- Level 7

More specific

3. Paste your lists of genes [see allowed IDs]

Paste list of genes [?]

P81255
Q8PAC3
Q8BG81
P83158
P81358
Q211119

Or upload a file with the list of genes [?]

Browse...

[Yeast example] [Human example]

➤ 高通量组学数据库分析——GENECODIS

— Singular annotations results: GOSlim Process

Download results in TAB delimited format

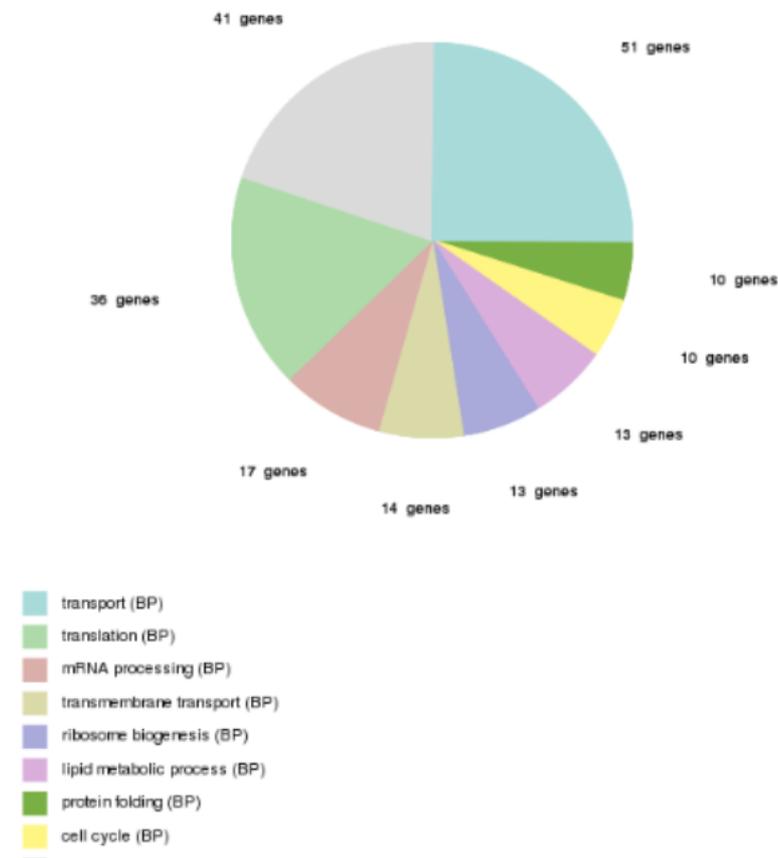
NGR = Number of annotated genes in the reference list (Total number of genes in the reference list);

NG = Number of annotated genes in the Input list (Total number of genes in the Input list);

Hyp = Hypergeometric pValue; Hyp* = Corrected hypergeometric pValue

Genes	NGR	NG	Hyp	Hyp*	Annotations
36 genes	271(36814)	36(327)	2.84972e-31	8.54916e-30	GO:0006412 :translation (BP)
13 genes	56(36814)	13(327)	2.26401e-15	3.39601e-14	GO:0042254 :ribosome biogenesis (BP)
51 genes	1637(36814)	51(327)	6.29504e-15	6.29504e-14	GO:0006810 :transport (BP)
17 genes	257(36814)	17(327)	1.96471e-10	1.47353e-09	GO:0006397 :mRNA processing (BP)
13 genes	229(36814)	13(327)	1.64115e-07	9.84691e-07	GO:0006629 :lipid metabolic process (BP)
10 genes	130(36814)	10(327)	2.7699e-07	1.38495e-06	GO:0006457 :protein folding (BP)
9 genes	110(36814)	9(327)	6.52241e-07	2.79532e-06	GO:0006950 :response to stress (BP)
14 genes	432(36814)	14(327)	3.8007e-05	0.000142526	GO:0055085 :transmembrane transport (BP)
8 genes	174(36814)	8(327)	0.00017589	0.000586301	GO:0016192 :vesicle-mediated transport (BP)
8 genes	279(36814)	8(327)	0.0036881	0.0110643	GO:0051301 :cell division (BP)
6 genes	203(36814)	6(327)	0.00986557	0.0269061	GO:0007067 :mitosis (BP)
5 genes	150(36814)	5(327)	0.0111557	0.0278892	GO:0008283 :cell proliferation (BP)
10 genes	493(36814)	10(327)	0.0134208	0.030971	GO:0007049 :cell cycle (BP)
3 genes	60(36814)	3(327)	0.016367	0.0350722	GO:0007059 :chromosome segregation (BP)
2 genes	26(36814)	2(327)	0.0222137	0.0444274	GO:0006913 :nudeocytoplasmic transport (BP)

Number of genes per concurrent annotations

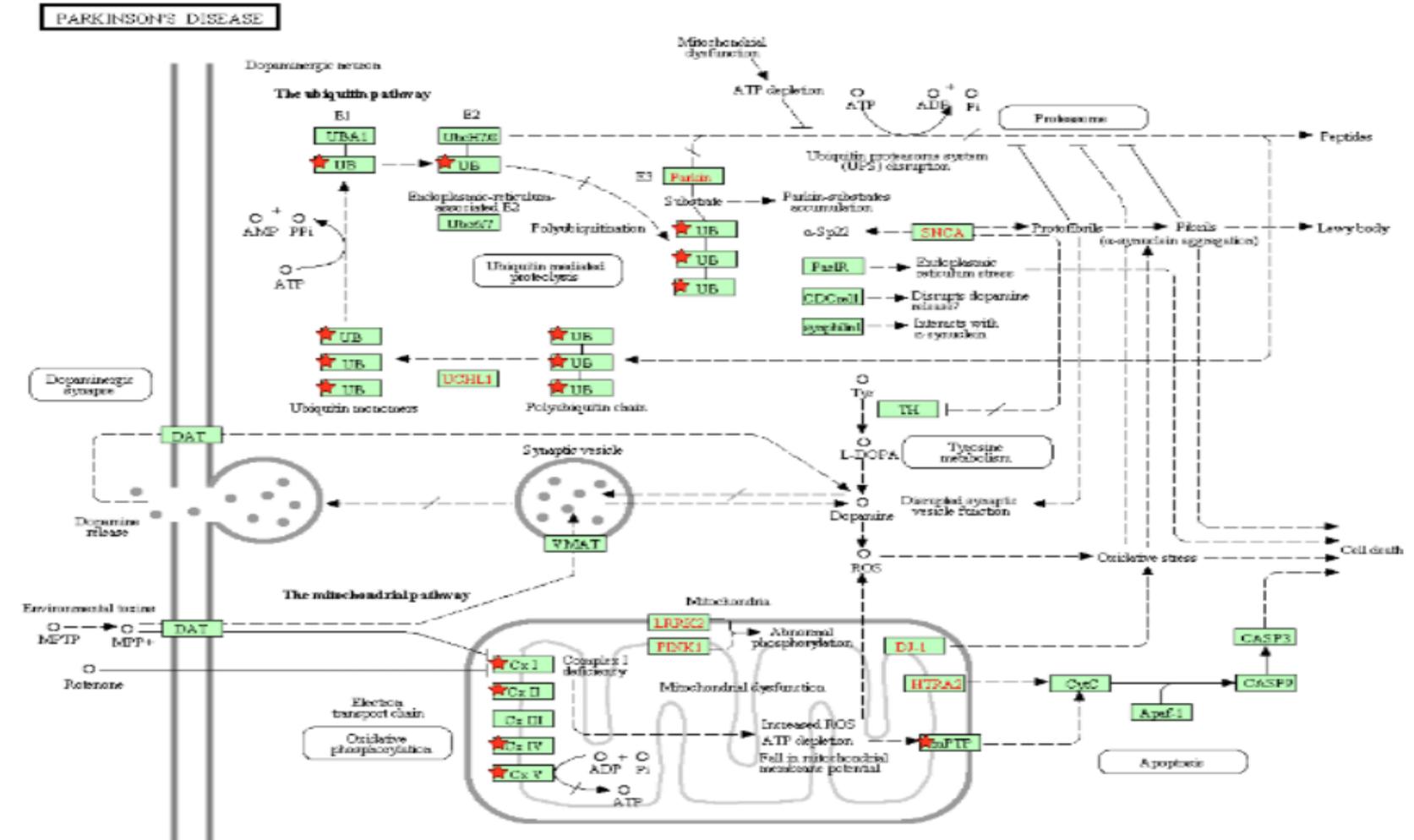


◆ GENECODIS每一类注释都会配有Detailed Table和Pie Chart

数据分析

➤ 蛋白相互作用分析

Parkinson Disease



数据分析

➤ 蛋白相互作用分析

The screenshot shows the STRING 9.0 web interface. At the top, there are navigation links: Home · Download · Help/Info. The main title is "STRING - Known and Predicted Protein-Protein Interactions".

The search interface includes several tabs: "search by name", "search by protein sequence", "multiple names" (highlighted with a red box), and "multiple sequences". Below these is a text input field labeled "list of names:" with a note "(one per line; examples: #1 #2 #3)". The field contains the following protein IDs: O55042, Q9CQZ5, P56391, Q9D3D9, and Q8K2B3. There is a "Browse..." button next to the input field.

Below the input field is an "organism:" dropdown menu (highlighted with a red box) currently set to "Mus musculus". Underneath, there are radio buttons for "interactors wanted": "COGs" and "Proteins" (selected). There are also "Reset" and "GO!" buttons.

To the right of the search interface is a box titled "What it does ...". It explains that STRING is a database of known and predicted protein interactions, including direct (physical) and indirect (functional) associations. It lists four sources: Genomic Context, High-throughput Experiments, (Conserved) Coexpression, and Previous Knowledge. It also states that STRING quantitatively integrates interaction data from these sources for a large number of organisms, covering 5'214'234 proteins from 1133 organisms.

At the bottom, there are navigation tabs: "More Info", "Funding / Support", "Acknowledgements", and "Use Scenarios". The "More Info" tab is active, showing text about the development of STRING at CPR, EMBL, SIB, KU, TUD, and UZH, along with references and a "What's New?" section stating it is version 9.0 covering more than 1100 organisms.

◆ 手动添加一些与PD密切相关的蛋白Uniprot AC

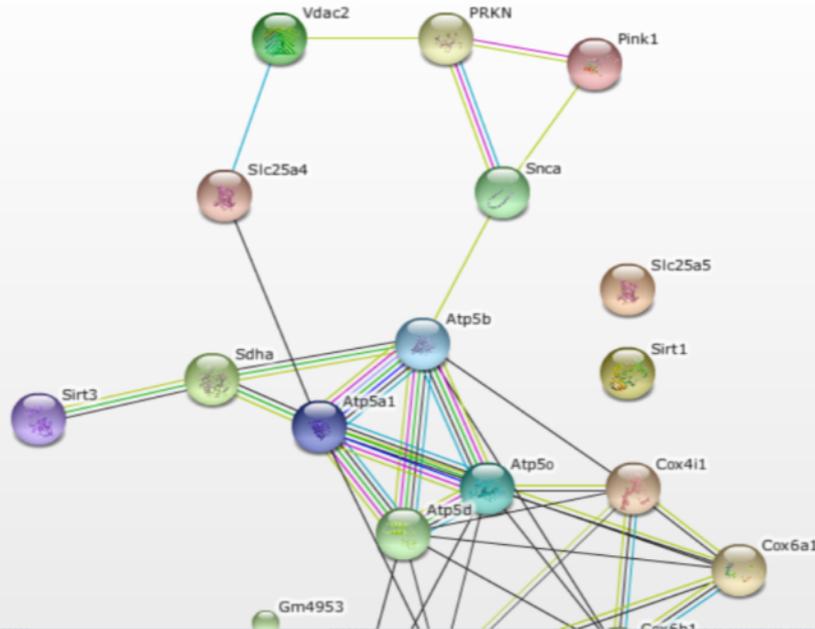
数据分析

➤ 蛋白相互作用分析

Home · Download · Help/Info

STRING  打开视频播放器 x

Save Layout Clustering GO-Enrichment Options



This is the **evidence view**. Different line colors represent the types of evidence for the association.

confidence evidence actions interactive advanced more less

(requires Flash player 10 or better)

Your Input:

-  Cox6c cytochrome c oxidase, subunit VIc Gene; This protein is one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the terminal oxidase in mitochondrial electron transport (By similarity) (76 aa)
-  Slc25a5 solute carrier family 25 (mitochondrial carrier, adenine nucleotide translocator), member 5 Gene; Catalyzes the exchange of ADP and ATP across the mitochondrial inner membrane (288 aa)

总结

➤ 数据分析策略与技巧

- ◆ 粗筛数据清除障碍
- ◆ 综合利用不同的高通量数据库所得的注释和分类信息进行整合
- ◆ 将上调和下调的蛋白分别进行分析
- ◆ 添加或删除一些无用条目
- ◆ 利用近缘物种的同源蛋白进行分析
- ◆ 以相关文献报道为参考判断数据分析的可信度
- ◆ 结合文献积累以及直觉判断对自己研究有用的信息

Thanks for your attention!

