

华大基因
蛋白质组与转录组关联分析
帮助文档

华大基因质谱研究部

版本号: BGI_V_1.2

2015 年 8 月 31 日

目录

1 文件夹及文件说明.....	2
2 文件格式说明.....	3
2.1 General 文件夹.....	3
2.2 SubCorrelation 文件夹	3
2.3 SubCorrelation2Function 文件夹	4
2.3.1 COG 文件夹	4
2.3.2 GO 文件夹.....	5
2.3.3 Pathway 文件夹	5
2.4 Venn 文件夹	7
2.5 GOCorrelation2Diff 文件夹	7
2.6 PathwayCorrelation2Diff 文件夹	9
2.7 IntegrationPathwayMap 文件夹	11
2.8 cluster 文件夹	12
3 联系我们	13
参考文献	14

1 文件夹及文件说明

Correlation 文件夹下的文件夹及文件内容见表 1.1。

表 1.1 关联分析结果文件夹及文件说明

序号	文件夹或文件名称	内容说明
1	General	鉴定、定量、差异 3 层次关联结果
2	SubCorrelation	关联结果细分成 5 类分析结果
3	SubCorrelation2Function	关联结果及细分 5 类结果的蛋白功能注释结果
4	Venn	鉴定、定量及差异关联数量 venn 图
5	GOCorrelation2Diff	GO 富集关联分析结果
6	PathwayCorrelation2Diff	Pathway 富集关联分析结果
7	IntegrationPathwayMap	Pathway 总通路图和差异通路图整合分析结果
8	Cluster	蛋白组与转录组聚类分析结果
9	Pipeline	结题报告素材
△10	cor_num.xls	关联结果统计表
△11	Parameters2Correlation.xls	关联分析参数
△12	SpearmanCoefficient_AllGroup*	相关性系数统计图
△13	Correlation_analysis_help_ch.pdf	关联分析帮助文档

注：1.其中 Pipeline 文件夹下的内容不需要查看，该文件夹下的内容仅为结题报告提供素材。

2.△标记的条目为文件，未标记的为文件夹。

2 文件格式说明

2.1 General 文件夹

该文件夹下的***_Cor_All.xls ,***_Cor_DEPs.xls ,***_Cor_Quant.xls 文件内容截图如图 2.1.1，表格含义见表 2.1.1。

图 2.1.1 general 文件夹下 Excel 文件内容截图

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	NO.	Protein ID	Protein Gr	Protein De	Unique	Pep	Protein Seq	Protein Re	Protein Sig	DEPs	Correlatio	Gene ID	Gene Descr	Gene Name	Gene Ratio	Gene Sig	DEGs
2	1	Bra000657	NA	[mRNA]	1	MACDKPRGI	----	NA	NA	NA	Bra000657	Bra000657	NA	NA	1.4770877	0.8545189	+
3	3	Bra000424	NA	[mRNA]	1	MACLRGTSRF	0.9166667	NA	NA	NA	Bra000424	Bra000424	NA	NA	-0.485776	0.3390305	NA
4	5	Bra032649	Bra033236	[mRNA]	1	MANASGMHV	0.9983333	NA	NA	NA	Bra032649	Bra032649	NA	NA	0.1371733	0.1801612	NA
5	8	Bra034388	NA	[mRNA]	2	MSLLTLFFS	1.2363333	NA	NA	NA	Bra034388	Bra034388	NA	NA	-0.771452	0.6512782	NA
6	7	Bra020541	NA	[mRNA]	3	MDSQQRSLCF	0.7446667	NA	NA	NA	Bra020541	Bra020541	NA	NA	-0.371065	0.5097842	NA

表格 title 含义见下：

表 2.1.1 general 文件夹下 Excel 文件 title 含义

列名	含义
NO.	编号
Protein ID	蛋白 ID
Protein Group ID	蛋白 Group ID
Protein Description	蛋白质描述信息
Unique Pep Number	蛋白 unique 肽段数量
Protein Seq	蛋白质氨基酸序列
Protein Ratio(A/B)	蛋白 ratio 值
Protein Sig	蛋白是否显著
DEPs	蛋白是否是差异蛋白
Correlation ID	关联用 ID
Gene ID	基因 ID
Gene Description	基因描述信息
Gene Name	基因名
Gene Ratio(A/B)	基因 ratio 值
Gene Sig	基因表达是否显著
DEGs	基因是否是差异表达基因

备注：1 “NA” 表示无相关信息或表达无显著性 or 无差异表达

2 “Protein Sig” 一列 “*” 表示统计检验具有显著性，

3 “DEPs 及 DEGs” 中 “+” 表示上调差异表达蛋白或基因，“-” 表示下调差异表达蛋白或基因。

2.2 SubCorrelation 文件夹

该文件夹下的内容为将关联结果细分成 5 种类型（见表 2.2.1）结果。每种分类结果中的 Excel 文件 title 含义同表 2.1.1 说明。

表 2.2.1 关联分析结果细分 5 类情况

文件夹名称	内容
Cor_DEPs_DEGs_Opposite	蛋白表达上调、基因表达下调 蛋白表达下调、基因表达上调
Cor_DEPs_DEGs_SameTrend	蛋白表达上调、基因表达上调或蛋白表达下调、基因表达下调
Cor_DEPs_NDEGs	蛋白表达有差异、基因表达无差异
Cor_NDEPs_DEGs	蛋白表达无差异、基因表达有差异
Cor_NDEPs_NDEGs	蛋白表达无差异、基因表达无差异

2.3 SubCorrelation2Function 文件夹

对鉴定、定量和差异层面的关联结果及细分成 5 种类型后的关联结果得到的蛋白进行 GO、COG 和 Pathway 分析，文件夹的名称后缀即可区分各种类型。

2.3.1 COG 文件夹

以 “***_Cor_All_COG” 文件夹为例进行说明：

- (1) ***_Cor_All.cog2protein.xls:按 COG 功能分类对每一个 COG 分类条目下的蛋白进行统计并列出对应的蛋白 ID；此文件 title 简单易懂不作说明；
- (2) ***_Cor_All.fa.blast.cog.xls: 该文件为分析中间文件，可以不看。Title 含义见表 2.3.1

表 2.3.1***_Cor_All.fa.blast.cog.xls 文件 Title 含义

序号	内容说明
1	蛋白 ID
2	COG 数据库中蛋白或结构域 ID
3	Blast 比对一致性
4	比对长度
5	错配数
6	空位数
7	蛋白比对起始氨基酸位置
8	蛋白比对氨基酸终止位置
9	COG 蛋白比对起始氨基酸位置
10	COG 蛋白比对氨基酸终止位置
11	期望值
12	比对得分

备注：1.序号为对文件从左往右依次的编号

2.主要看期望值和比对得分，比对一致性越高说明两条序列越相似。

- (3) ***_Cor_All.fa.cog.pdf 和***_Cor_All.fa.cog.png 为 COG 注释情况统计图
- (4) ***_Cor_All.protein2cog.xls: 对每个蛋白在 COG 数据库中的注释信息进行统计并列出相关信息；Title 含义见表 2.3.2:

表 2.3.2 ***_Cor_All.protein2cog.xls 文件 title 含义

列名	含义
----	----

Protein	蛋白 ID
Protein-or-Domain	COG 数据库中蛋白 ID
Score	Blast 比对得分
E-Value	Blast 期望值
COG-ID	COG ID
Function-Description	功能描述
Code	COG 数据库功能分类条码代号
Functional-Categories	功能分类信息

2.3.2 GO 文件夹

以 “***_Cor_All_GO” 文件夹为例进行说明：

(1) biological_process.png、cellular_component.png、molecular_function.png：本次关联结果蛋白在 GO 数据库中三个本体（Ontology）的注释情况统计图；这三个图的合并图为 “***_Cor_All.fa.GO.svg”，因此也只需要看***_Cor_All.fa.GO.svg；

(2) ***_Cor_All.fa.GO2protein.xls：以 GO 数据库中的 3 个本体为分类主体，分别列出对应 3 个本体下的 GO 条目及该条目下注释到的蛋白数量和蛋白 ID。该文件 Title 简单易懂不作说明。

(3) ***_Cor_All.fa.protein2GO.xls：该文件以蛋白为主体，列出该蛋白的描述信息及注释到的 GO 条目（以 GO 号的形式给出）；Title 从左往右依次为：蛋白 ID->蛋白描述信息->蛋白注释到的 GO 条目

(4) ***_Cor_All_GO_detail.xls 该文件以蛋白为主体，按照 GO 数据库中的 3 个本体注释进行排列。如果看了（2）和（3）文件，该文件可以不看。

2.3.3 Pathway 文件夹

以 “***_Cor_All_Pathway” 文件夹为例进行说明：

(1) ***_Cor_All_map 文件夹：蛋白注释到的通路图，每个通路图都以 png 和 html 两种形式给出。红框标注的为本次关联结果分类后的蛋白。

(2) ***_Cor_All.htm：蛋白 pathway 注释信息，对于 Pathway 文件夹下的内容看此网页即可。该文件夹下的其他内容都会被此网页展现出来。网页示例如图 2.3.1，格式说明如表 2.3.3。

图 2.3.1 ***_Cor_All.htm 内容截图

#	Pathway	Diff Proteins with pathway annotation (4187)	Pathway ID
1	Metabolic pathways	1364 (32.58%)	ko01100
2	Biosynthesis of secondary metabolites	793 (18.94%)	ko01110
3	Ribosome	208 (4.97%)	ko03010
4	Protein processing in endoplasmic reticulum	155 (3.7%)	ko04141
5	Spliceosome	149 (3.56%)	ko03040
6	Plant-pathogen interaction	144 (3.44%)	ko04626
7	RNA transport	141 (3.37%)	ko03013
8	Starch and sucrose metabolism	127 (3.03%)	ko00500

***_Cor_All.htm 格式说明:

表 2.3.3 ***_Cor_All.htm 格式说明

列名	含义
Pathway	通路名, 点击可看到注释到该通路的各蛋白质情况及 Pathway 图
Diff Proteins with pathway annotation	鉴定的蛋白质注释到该通路的数目及比例
Pathway ID	KEGG 数据库中的 Pathway ID

(3) ***_Cor_All.ko: pathway 分析中间文件, 可以不看;

*.ko 为蛋白 ID 及对应 KO 号列表文件, 示例如图 2.3.2(该文件可以选择 Excel 打开, 也可以在终端直接打开)

图 2.3.2 ***_Cor_All.ko 文件内容截图

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	# Method:Condition: expect <= 1e-05; rank <= 5												
2	# Summary4187 succeed, 1535 fail												
3													
4	# query ko_id:rank:evaluate:score:identity:ko_definition												
5	Bra000000K12492 1 3e-172 603 aly:ARALYDRAFT_482750 ADP-ribosylation factor GTPase-activating protein 1												
6	Bra000001K00011 1 1e-145 513 ath:AT2G37790 aldehyde reductase [EC:1.1.1.21]!K00002 2 4e-129 459 osa:4338987 alcohol d												

title 含义见表 2.3.4:

表 2.3.4 ***_Cor_All.ko 文件 title 含义

列名	含义
query	蛋白 ID
ko_id	pathway ID
rank	blast 比对得分排名
evaluate	blast 期望值
score	blast 比对得分
identity	物种及基因 ID
ko_definition	pathway 描述信息

(4) ***_Cor_All.path: 分析中间文件, 可以不看; 此文件 title 含义见表 2.3.5;

表 2.3.5 ***_Cor_All.path 文件 title 含义

列名	含义
Pathway	通路名
Count	注释到该通路的蛋白质的数目
Pathway ID	KEGG 数据库中的 Pathway ID
Proteins	注释到该 Pathway 的蛋白
KOs	属于该 Pathway 的 KEGG Orthology

2.4 Venn 文件夹

该文件夹下的内容为鉴定、定量和差异层面关联分析结果 Venn 图；

2.5 GOCorrelation2Diff 文件夹

Gene Ontology (简称 GO) 是一个国际化的基因功能分类体系，提供了一套动态更新的标准词汇表 (Controlled Vocabulary) 来全面描述生物体中基因和基因产物的属性。GO 总共有三个本体 (Ontology)，分别描述基因的分子功能 (Molecular Function)、所处的细胞位置 (Cellular Component)、参与的生物过程 (Biological Process)。因此对蛋白质组与转录组分别在 GO 数据库中的 3 个本体条目中进行富集分析，再将两组学的富集结果按功能条目进行联合分析并统计相关数据。

按照差异蛋白和差异基因在各 GO 条目中富集显著性情况进行分类：

- 1) Both_of_Significance: 蛋白质组与转录组两个组学都显著富集
- 2) Either_of_Significance: 蛋白质组与转录组任意一个组学数据显著富集
- 3) None_of_Significance: 蛋白质组与转录组两个组学均无显著富集

依据上述分类情况，按照“1)->2)->3)”的优先级别进行排序（对于同一分类的条目，按照蛋白质组富集 P-Value 值由小到大进行排序），输出富集到的全部 GO 和一定数量的 GO 条目数量图。按照上述分类情况以 P-Value 值的大小为坐标输出所有的条目散点图。

1.DOC 文件夹

该文件夹为 GO 富集关联结果的图、表信息。

该文件夹下文件的后缀为 ***_C_*** 的文件即为 Cellular Component 的分析结果，***_P_*** 的文件为 Biological Process 的分析结果，***_F_*** 的文件为 Molecular Function 的分析结果。

以 ***_C_*** 的文件进行数据说明：

(1) ***_C_GO_detail.xls

该文件为蛋白质组与转录组在 GO 本体 Molecular Function 条目中富集关联的详细结果。该文件内容部分截图见图 2.5.1

图 2.5.1 ***_C_GO_detail.xls 内容截图

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	No.	GO_ID	GO_Term	Number of	Number of	Number of	Proteomics	Transcriptom	Protein	II	Gene ID	Correlation	IDs		
2	1	GO:001602C	membrane	170	746	33	6.72E-12	0.0002	Bra013741	Bra034471	Bra029025	(Protein:1.041 + @@Gene:1.847 +); Bra0165			
3	2	GO:004434	chloroplas	92	110	7	4.52E-07	9.24E-09	Bra007912	Bra025682	Bra013741	(Protein:0.965 + @@Gene:-1.232 -); Bra033			
4	3	GO:001989	extrinsic	9	3	0	0.0005411		1	Bra015896	Bra009751	(-1.858 -); Bra021603	(1.320 +); Bra028520	(2.224 +)	
5	4	GO:003031	external e	36	232	14	0.038387	1.18E-15	Bra006103	Bra007071	Bra028409	(Protein:-0.902 - @@Gene:-1.531 -); Bra01			
6	5	GO:003109	C organelle	32	272	11	0.7544626	5.53E-15	Bra004806	Bra012716	Bra004806	(Protein:-0.883 - @@Gene:-1.076 -); Bra01			
7	6	GO:003198	C membrane-b	2	5	0	0.9825403		1	Bra027209	Bra015543	(1.388 +); Bra036311	(1.488 +); Bra006372	(1.570 +);	
8	7	GO:000953	chloroplas	52	95	7	1.17E-11	1.99E-07	Bra013741	Bra025682	Bra013741	(Protein:0.965 + @@Gene:-1.232 -); Bra015			
9	8	GO:000573	cytoplasm	240	867	48	5.65E-05	4.29E-26	Bra013741	Bra032945	Bra013741	(Protein:0.965 + @@Gene:-1.232 -); Bra021			

该文件的 title 含义见表 2.5.1

表 2.5.1 ***_C_GO_detail.xls 文件 title 含义

列名	含义
No.	编号
GO_ID	GO ID
GO_Term	GO Term
Number of Protein	该 GO 条目下蛋白数量
Number of Transcript	该 GO 条目下基因数量

Number of Correlation	该 GO 条目下蛋白和基因关联上的数量
Proteomics(P-Value)	蛋白质组中经超几何检验该 GO 条目的 P 值
Transcriptome(P-Value)	转录组中经超几何检验该 GO 条目的 P 值
Protein ID	该 GO 条目下蛋白 ID
Gene ID	该 GO 条目下基因 ID
Correlation IDs	该 GO 条目下蛋白和基因关联上的 ID

注：1. Protein ID 或 Gene ID 列中蛋白或基因 ID 后括号内内容为：LOG2 (Ratio) 值及上调 (+标识) 或下调 (-标识) 信息

2. Correlation IDs 一列中, 如果蛋白和基因 ID 相同, 则 ID 号后的括号内内容为: 蛋白 LOG2(ratio) 值, 蛋白上下调信息, 基因 LOG2 (ratio) 值, 基因上下调信息。如果蛋白和基因 ID 号不同, 则会将关联上的蛋白 ID 前采用 “P—>” 进行标识, 基因 ID 前会采用 “G—>” 标识, 且都会附上表达量信息和上下调信息。不同的关联组之间会用 “;” 隔开。

(2) ***_C_GO_pro2trans.xls 及***_C_Statistical.xls

这两个文件为阅读的方便按照不同的方式对 GO 富集关联的结果进行的展示, 这两个文件的最终整合结果即为 (1) 中的***_C_GO_detail.xls 文件内容, 且这两个文件中的 title 含义同表 2.5.1。因此这两个文件可不查看。

(3) FunGO2CorStatistical.xls

该文件为本次所有比对方案进行以 GO 条目关联后的统计结果。文件格式说明见表 2.5.2

表 2.5.2 FunGO2CorStatistical.xls 格式说明

列名	含义
Group Name	比较组名称
Functional Classification	Gene Ontology
Type	类型, 以是否显著来区分
Number of GO Term[Proteomics]	蛋白质组富集 GO 条目数
Number of GO Term[Transcriptome]	转录组富集 GO 条目数
Number of Correlation	蛋白质组与转录组关联上的 GO 条目数

注：1.Type 类型中 ALL 表示所有的 GO 条目情况；Significant 表示是否具有显著性的 GO 条目情况。

2.HTML 文件夹

该文件夹下的内容为 GO 富集关联结果*.html 文件, html 文件打开后, 将鼠标放置在数据框中的文字处, 即可显示相关详细信息。点击 GO 条目名称, 可以链接至 GO 数据库网站, 查询该 GO 条目的详细信息。

将每个比较组按 GO 三个本体进行富集分析, 并进行关联, 将关联结果按照 GO 条目的富集 P-Value 值的情况, 根据设定的显著性阈值, 将结果进行分类, 分类信息见本部分章节开始处。并将每种结果都用 HTML 形式展现, 便于查阅。使用示意图见图 2.5.2 及图 2.5.3

图 2.5.2 GO 本体 Cellular Component 富集关联结果部分截图 1

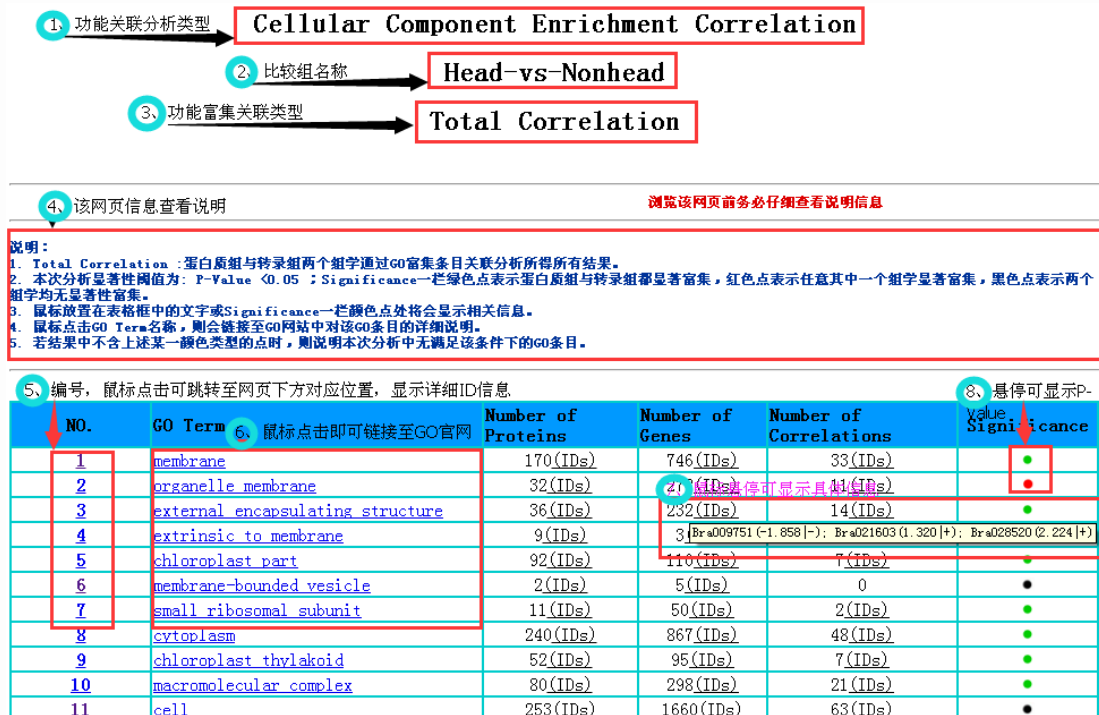
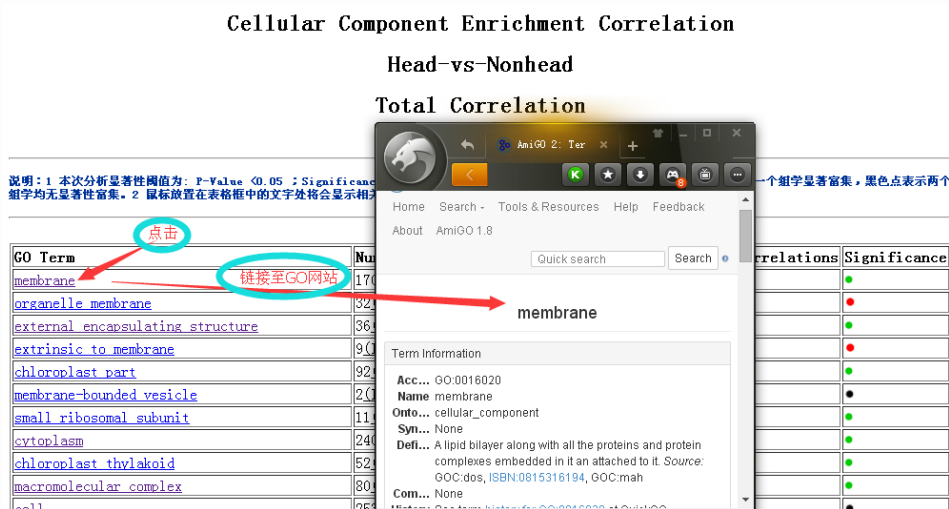


图 2.5.3 GO 本体 Cellular Component 富集关联结果部分截图 2



2.6 PathwayCorrelation2Diff 文件夹

该文件夹下分析内容为蛋白质组与转录组数据分别在 KEGG 数据库中进行富集分析, 将两组学富集分析数据按照通路进行关联分析, 在两个组学都富集成功的通路进行整合关联分析。按照差异蛋白和差异基因在各代谢通路中富集显著性情况进行分类:

- 1) Both_of_Significance: 蛋白质组与转录组两个组学都显著富集
- 2) Either_of_Significance: 蛋白质组与转录组任意一个组学数据显著富集
- 3) None_of_Significance: 蛋白质组与转录组两个组学均无显著富集

依据上述分类情况, 按照“1)->2)->3)”的优先级别进行排序(对于同一分类的代谢通

路，按照蛋白质组富集 P-Value 值由小到大进行排序），输出富集到的全部代谢通路图和一定数量的通路统计数量图。按照上述分类情况以 P-Value 值的大小为坐标输出所有的条目散点图。

1.DOC 文件夹

(1) ***_Pathway_detail.xls

该文件为蛋白质组与转录组按照富集通路进行关联分析的详细结果。该文件的内容截图见图 2.6.1

图 2.6.1 ***_Pathway_detail.xls 内容截图

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	No.	Pathway_ID	Pathway_Term	Number of Protein	Number of Transcript	Number of Correlation	Proteomics(P-Value)	Transcriptome(P-Value)	Protein ID	Gene ID	Correlation IDs					
2	1	ko00380	Tryptophan	7	35	6	0.0142488	2.84E-04	Bra034674	Bra028711	Bra034674	(Protein:1.823 + @@Gene:-2.485 -);	Bra035006	(Protein:0.869 + @@Gene:1.558 +)		
3	2	ko00040	Pentose ar	4	35	1	0.4025184	4.08E-01	Bra018090	Bra025376	Bra018090	(Protein:0.869 + @@Gene:1.558 +)				
4	3	ko00380	Phenylalanine	5	17	3	0.1483217	9.99E-01	Bra015404	Bra034523	Bra015404	(Protein:1.012 + @@Gene:5.462 +);	Bra038526	(Protein:0.738 + @@Gene:-1.342 -);	Bra008421	(Protein:0.738 + @@Gene:-1.342 -);
5	4	ko00710	Carbon fix	7	28	2	0.171326	8.97E-04	Bra003221	Bra040406	Bra035516	(Protein:0.738 + @@Gene:-1.342 -);	Bra008421	(Protein:0.738 + @@Gene:-1.342 -);		
6	5	ko00750	Vitamin B6	1	6	0	0.4889046	9.03E-02	Bra036724	Bra031486	(3.658 +);	Bra021157	(-2.207 -);	Bra002473	(-1.215 -);	Bra027140
7	6	ko00960	Tropone, f	3	6	1	0.2425555	9.99E-01	Bra036490	Bra034523	Bra036490	(Protein:-0.778 - @@Gene:1.659 +)				
8	7	ko00862	Fatty acid	1	4	0	0.4332286	1.00E+00	Bra033983	Bra035683	(3.963 +);	Bra020290	(-2.118 -);	Bra017491	(3.294 +);	Bra004251
9	8	ko00330	Arginine s	7	21	3	0.0600645	2.65E-01	Bra035529	Bra034648	Bra018090	(Protein:0.869 + @@Gene:1.558 +);	Bra035392	(Protein:0.869 + @@Gene:1.558 +);		

文件格式说明见表 2.6.1

表 2.6.1 ***_Pathway_detail.xls 格式说明

列名	含义
No.	编号
Pathway_ID	KEGG 数据库中的 Pathway ID
Pathway_Term	Pathway 名称
Number of Protein	该通路条目下蛋白数量
Number of Transcript	该通路条目下基因数量
Number of Correlation	该通路条目下蛋白和基因关联上的数量
Proteomics(P-Value)	蛋白质组中经超几何检验该通路条目的 P 值
Transcriptome(P-Value)	转录组中经超几何检验该通路条目的 P 值
Protein ID	该通路条目下蛋白 ID
Gene ID	该通路条目下基因 ID
Correlation IDs	该通路条目下蛋白和基因关联上的 ID

注：1. Protein ID 或 Gene ID 列中蛋白或基因 ID 后括号内内容为：LOG2 (Ratio) 值及上调 (+标识) 或下调 (-标识) 信息

2. Correlation IDs 一列中，如果蛋白和基因 ID 相同，则 ID 号后的括号内容为：蛋白 LOG2 (ratio) 值，蛋白上下调信息，基因 LOG2 (ratio) 值，基因上下调信息。如果蛋白和基因 ID 号不同，则会将关联上的蛋白 ID 前采用 “P—>” 进行标识，基因 ID 前会采用 “G—>” 标识，且都会附上表达量信息和上下调信息。不同的关联组之间会用 “;” 隔开。

(2) ***_pathway_pro2trans.xls 及 ***_Statistical.xls

这两个文件为阅读的方便按照不同的方式对 Pathway 富集关联的结果进行的展示，这两个文件的最终整合结果即为 (1) 中的 ***_Pathway_detail.xls 文件内容，且这两个文件中的 title 含义同表 2.6.1。因此这两个文件可不查看。

2.HTML 文件夹

该文件夹下的*.html 文件为：Pathway 富集关联结果按照本部分章节起始处的分类情况进行分析的结果。使用方法见 “2.5 GOCorrelation2Diff 文件夹” HTML 部分说明。

2.8 cluster 文件夹

对每一组对应的蛋白组和转录组的关联结果进行 heatmap 分析, 每一个比对方案分别做定量和差异两个层面的聚类。聚类采用 heatmap.2, 有关说明和参数如下:

A heat map is a false color image (basically `image(t(x))`) with a dendrogram added to the left side and/or to the top. Typically, reordering of the rows and columns according to some set of values (row or column means) within the restrictions imposed by the dendrogram is carried out.

This heatmap provides a number of extensions to the standard R heatmap function. `distfun` function used to compute the distance (dissimilarity) between both rows and columns. Defaults to `dist`. `hclustfun` . function used to compute the hierarchical clustering when `Rowv` or `Colv` are not dendrograms. Defaults to `hclust`

See More ,Please pay a visit "<http://127.0.0.1:29667/library/gplots/html/heatmap.2.html>"

3 联系我们

服务热线: 400-706-6615/ 0755-25653043

技术支持: BGI-MS@genomics.cn

参考文献

- [1] Ritchie MD , et al.Methods of integrating data to uncover genotype–phenotype interactions. Nat Rev Genet. 2015 Feb;16(2):85-97.
- [2] Schwanhäusser B , et al. Global quantification of mammalian gene expression control. Nature.2013 Mar 7;495(7439):126-7.
- [3] Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses . Nat Rev Genet. 2012 Mar 13;13(4):227-32 .
- [4] Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples FEBS Letters 583 (2009) 3966–3973.
- [5] Mary Muers.Transcriptome to proteome and back to genome. Nature Reviews Genetics ,2011.
- [6] Bouchal P., et al.Combined proteomics and transcriptomics identifies carboxypeptidase B1 and NF-κB associated proteins as putative biomarkers of metastasis in low grade breast cancer. Mol Cell Proteomics. 2015 Apr 22.
- [7] Petersen HO, Höger SK, Looso M, et al. A Comprehensive Transcriptomic and Proteomic Analysis of Hydra Head Regeneration. Mol Biol Evol. 2015 .
- [8] Wang J, Mei H, Zheng C, et al. The metabolic regulation of sporulation and parasporal crystal formation in *Bacillus thuringiensis* revealed by transcriptomics and proteomics. Mol Cell Proteomics. 2013 May;12(5):1363-76.
- [9] B. FUTCHER, et al.A Sampling of the Yeast Proteome MOLECULAR AND CELLULAR BIOLOGY,Nov. 1999, 7357–7368.
- [10] 吴松峰, 朱云平, 贺福初. 转录组与蛋白质组比较研究进展. 生物化学与生物物理进展.2005,3 (2) .
- [11] Lan P, Li W, Schmidt W. Complementary proteome and transcriptome profiling in phosphate-deficient *Arabidopsis* roots reveals multiple levels of gene regulation. Molecular & Cellular Proteomics. 2012 Nov;11(11):1156-66.
- [12] Dyhrman ST, Jenkins BD, Rynearson TA, Saito MA, Mercier ML, et al. The Transcriptome and Proteome of the Diatom *Thalassiosira pseudonana* Reveal a Diverse Phosphorus Stress Response. PLoS ONE 7(3),2012.
- [13] Lackner DH, Schmidt MW, Wu S, Wolf DA, Bähler J. Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. Genome Biology. 2012;13:R25.
- [14] Gonzalez, LMG,et al.Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy .Plant Cell Environ. 2012 Apr;35(4):682-701.
- [15] Kuss C , Gan CS , Gunalan K , Bozdech Z , Sze SK , Preiser PR. Quantitative proteomics reveals new insights into erythrocyte invasion by *Plasmodium falciparum* .Mol Cell Proteomics. 2012 ,11(2).

- [16] Kocharunchitt C, King T, Gobius K, Bowman JP, Ross T. Integrated Transcriptomic and proteomic Analysis of the Physiological Response of *Escherichia coli* O157:H7 Sakai to steady-state conditions of cold and water activity stress. *Molecular & Cellular Proteomics*. 2012 .
- [17] 曹文君等.基因表达谱富集分析方法研究进展.生物技术通讯.2008.
- [18] Croken et al.Gene Set Enrichment Analysis (GSEA) of *Toxoplasma gondii* expression datasets links cell cycle progression and the bradyzoite developmental program. *BMC Genomics* .2014, 15:515.
- [19] Wang SH, You ZY, Ye LP, et al. Quantitative proteomic and transcriptomic analyses of molecular mechanisms associated with low silk production in silkworm *Bombyx mori*. *J Proteome Res*. 2014 Feb 7;13(2):735-51.
- [20] Chen Q, Guo W, Feng L, et al.Transcriptome and proteome analysis of *Eucalyptus* infected with *Calonectria pseudoreteaudii*. *J Proteomics*. 2015 Feb 6;115:117-31.
- [21] Trevisan S, Manoli A, Ravazzolo L, et al.Nitrate sensing by the maize root apex transition zone: a merged transcriptomic and proteomic survey. *J Exp Bot*. 2015 Apr 23. pii: erv165.
- [22] Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, et al. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLoS Genet* 7(6).2011.