

# GEO Explorer

- A shiny app for GEO database exploring
- 数据下载
- 如何设置差异分析的分组
- 平台注释的问题
- **ID** 转换、功能富集

# GEO Explorer

gselD

platform

Choose Phenotype File

Browse...

No file selected

Analysis

- ☐ output DEG?
- ☐ output GO?
- ☐ output KEGG?
- ☐ output GO.pdf?
- ☐ output KEGG.pdf?

GO Enrichment

KEGG Pathway

DEG

KEGG

GO

Please contact: xuhoushi1996@163.com

# GEO Explorer

gseID

GSE41804

platform

GPL570

Choose Phenotype File

Browse...

phenoType.csv

Upload complete

Analysis

- ☐ output DEG?
- ☐ output GO?
- ☐ output KEGG?
- ☐ output GO.pdf?
- ☐ output KEGG.pdf?

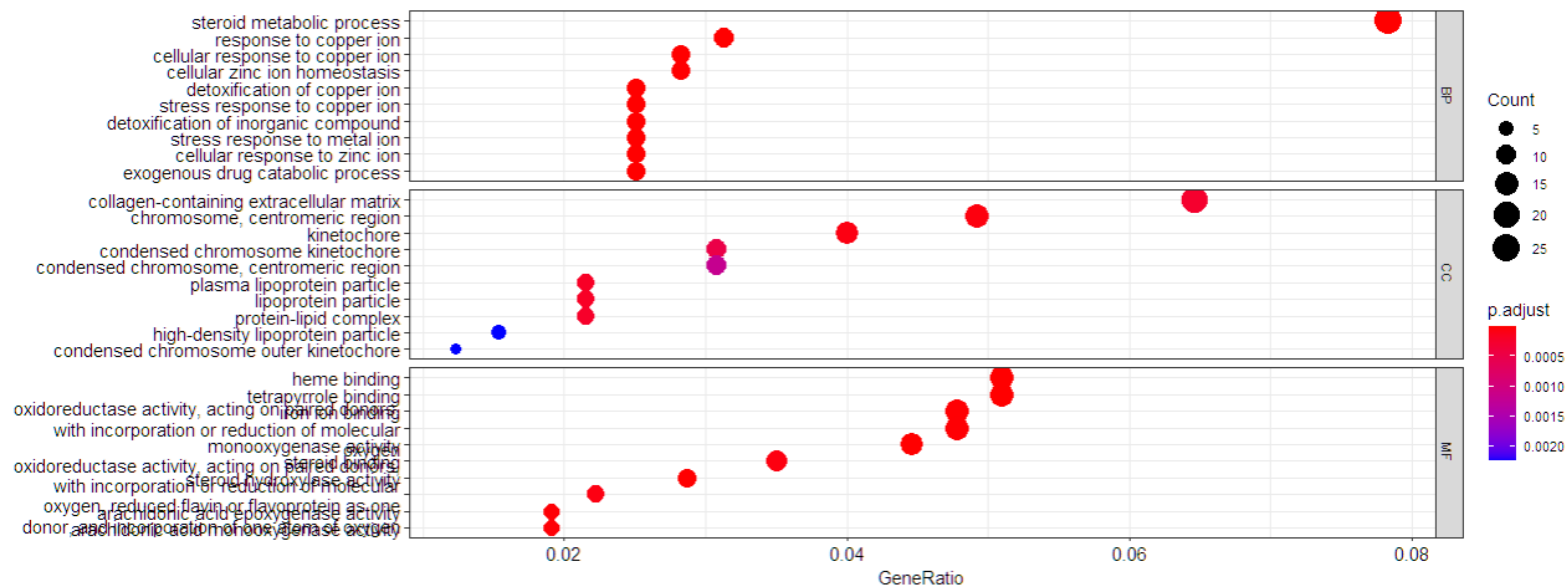
GO Enrichment

KEGG Pathway

DEG

KEGG

GO



Please contact: xuhoushi1996@163.com

# GEO Explorer

gseID

GSE41804

platform

GPL570

Choose Phenotype File

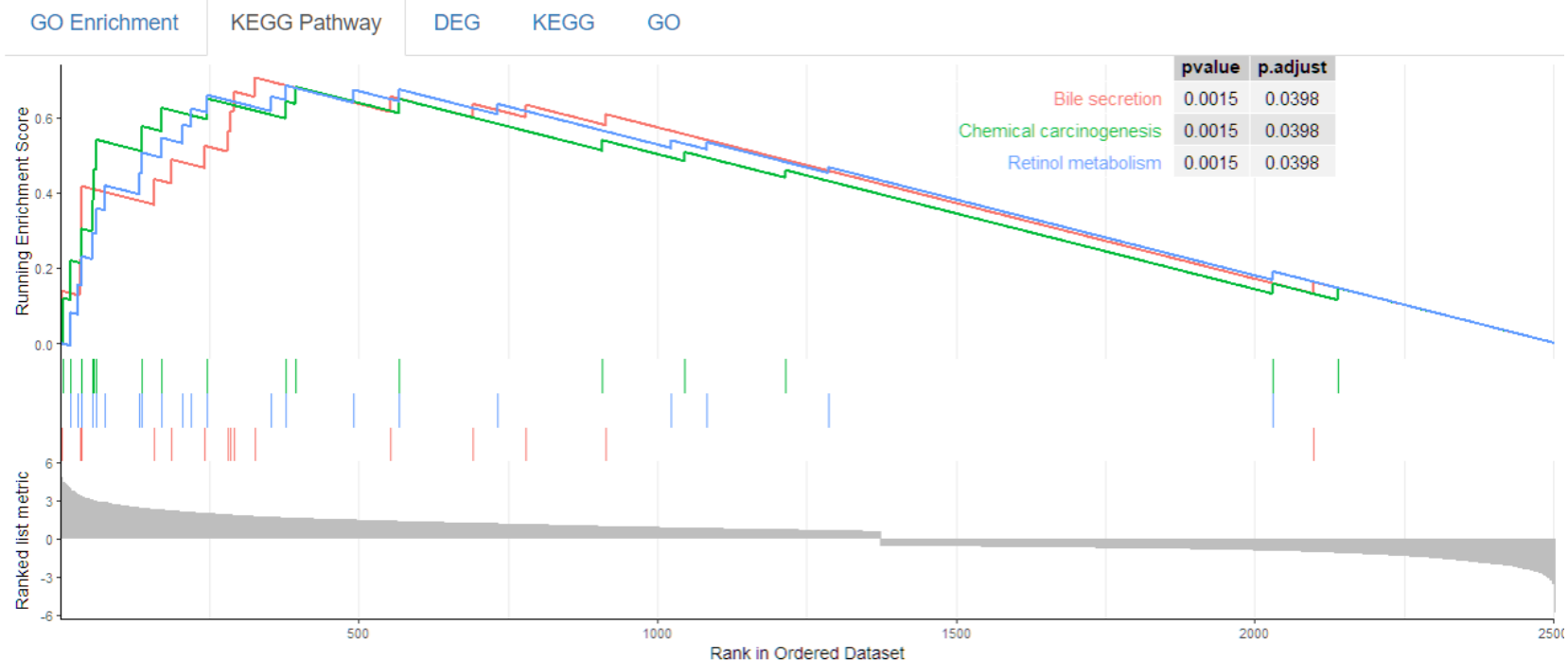
Browse...

phenoType.csv

Upload complete

Analysis

- ☐ output DEG?
- ☐ output GO?
- ☐ output KEGG?
- ☐ output GO.pdf?
- ☐ output KEGG.pdf?



Please contact: xuhoushi1996@163.com

# GEO Explorer

gseID

GSE41804

platform

GPL570

Choose Phenotype File

Browse...

phenoType.csv

Upload complete

Analysis

☐ output DEG?

☐ output GO?

☐ output KEGG?

☐ output GO.pdf?

☐ output KEGG.pdf?

GO Enrichment

KEGG Pathway

DEG

KEGG

GO

Show 10 entries

Search:

	logFC	adj.P.Val	P.Value	t	B
SPINK1	-5.7923792548	0.0000651124473316196	3.21542949785776e-8	-6.80591665613207	8.75514881835453
CXCL14	5.4889730216	0.000217458438990145	4.18996181055938e-7	6.01808280538391	6.35857615002355
FREM2	4.80326086745	0.0000106894455719776	5.86526506006999e-10	8.0560863135682	12.4691466842158
CFTR	4.7970624541	0.00104983160426024	0.00000775732909229335	5.11983324022693	3.62962117045163
CRHBP	4.45484068	0.000228699422727889	4.81032164859758e-7	5.97578189483354	6.2294994403017
NAT2	4.43530403575	0.000866959106008147	0.00000556265453470451	5.22294480617059	3.94044290084134
CLEC1B	4.42453189585	0.000171620499378264	2.5111367078667e-7	6.1749156348297	6.83705842667887
MARCO	4.39339422065	0.000132504063507939	1.67220491669827e-7	6.29950172466634	7.21690373199107
FCN2	4.29069491165	0.0000455595245821334	1.16659047855486e-8	7.11908540806285	9.69891170491655
KCNJ16	4.26533293885	0.000675085133941802	0.00000336888545101435	5.37792069609354	4.40930885764482

Showing 1 to 10 of 2,668 entries

Previous

1

2

3

4

5

...

267

Next

Please contact: xuhoushi1996@163.com

# GEO Explorer

gselD

GSE41804

platform

GPL570

Choose Phenotype File

Browse...phenoType.csv

Upload complete

Analysis

- ☐ output DEG?
- ☐ output GO?
- ☐ output KEGG?
- ☐ output GO.pdf?
- ☐ output KEGG.pdf?

GO EnrichmentKEGG PathwayDEGKEGGGO

Show10▼entries

Search:

ID		Description	setSize	enrichmentScore	NES	pvalue	p.adjust	
hsa04976	hsa04976	Bile secretion	16	0.705098038486201	2.40332872624676	0.00146198830409357	0.0339912280701754	0.021
hsa00830	hsa00830	Retinol metabolism	21	0.684207611638665	2.50989567639784	0.00143884892086331	0.0339912280701754	0.021
hsa05204	hsa05204	Chemical carcinogenesis	17	0.680488175307307	2.35361986220042	0.00145985401459854	0.0339912280701754	0.021
hsa00350	hsa00350	Tyrosine metabolism	11	0.655373214381801	1.98212650518469	0.00303490136570561	0.047040971168437	0.033
hsa00982	hsa00982	Drug metabolism - cytochrome P450	16	0.652805597534487	2.22508978833305	0.00146198830409357	0.0339912280701754	0.021
hsa04978	hsa04978	Mineral absorption	19	0.640211194415164	2.27339440388785	0.00145772594752187	0.0339912280701754	0.021
hsa00980	hsa00980	Metabolism of xenobiotics by cytochrome P450	16	0.610745246484795	2.08172695877424	0.00292397660818713	0.047040971168437	0.033

# GEO Explorer

gseID

GSE41804

platform

GPL570

Choose Phenotype File

Browse...

phenoType.csv

Upload complete

Analysis

☐ output DEG?

☐ output GO?

☐ output KEGG?

☐ output GO.pdf?

☐ output KEGG.pdf?

GO EnrichmentKEGG PathwayDEGKEGGGO

Show10▼ entriesSearch:

ONTOLOGY		ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	
GO:0010273	BP	GO:0010273	detoxification of copper ion	8/319	15/18670	3.86009500483114e-11	6.85745877608252e-8	5
GO:1990169	BP	GO:1990169	stress response to copper ion	8/319	15/18670	3.86009500483114e-11	6.85745877608252e-8	5
GO:0061687	BP	GO:0061687	detoxification of inorganic compound	8/319	17/18670	1.41542023268443e-10	1.25724702168195e-7	1
GO:0097501	BP	GO:0097501	stress response to metal ion	8/319	17/18670	1.41542023268443e-10	1.25724702168195e-7	1
GO:0008202	BP	GO:0008202	steroid metabolic process	25/319	331/18670	5.84508964788565e-10	4.15352070378754e-7	3
GO:0071280	BP	GO:0071280	cellular response to copper ion	9/319	29/18670	8.22516264436986e-10	4.87066714590769e-7	4
GO:0071294	BP	GO:0071294	cellular response to zinc ion	8/319	23/18670	2.61066033418589e-9	0.00000121375485854949	0.000
GO:0046688	BP	GO:0046688	response to copper ion	10/319	44/18670	2.73291271274864e-9	0.00000121375485854949	0.000

# 数据下载

- 目前工具是**GEOquery**
- 考虑到下载问题/时间，设计上传数据和平台文件接口



# 如何设置差异分析的分组

- **BART**的解决方案是读取**pData**，用户选择分组后再提交一次
- 目前我的解决方法是上传一个分组文件

# 平台注释的问题

- 先做差异分析再注释
- 平台文件繁多 **ImaGEO**的解决方案是预先将平台文件整理好 (**GPL570.txt**)
- 目前**GEO Explorer**的解决方案是先通过字符大致选择某几列, 后续会让用户自主勾选哪一列是**gene symbol** 然后再**submit**
- 通过正则表达式保留第一个基因名 探针多对一保留第一个
- 平台注释陈旧? 考虑后续做一个工具用**blast/seqman**比对

# ID 转换、功能富集

- *clusterProfiler* OR *biomart*
- 后续做一个子工具 允许用户根据富集结果自主绘图  
(**Hiplot**)

谢谢！