

Report of Analysis

Huang LiChuang of Wie-Biotech

Contents

1 第一部分	1
1.1 ETCM 中药丹参的化合物以及靶点基因	1
1.1.1 获取 ETCM 网站数据	1
1.1.2 获取 HERB 网站数据	2
1.1.3 获取 PubChem 数据库关于化合物的别名	2
1.1.4 以 HERB 的数据补充 ETCM 的数据	2
1.1.5 总结	3
2 第二部分	3
2.1 在 genecards 网站上检索胃癌相关的基因	3
3 第三部分	5
3.1 使用 CellMiner 数据库的 NCI-60 数据集	5
3.2 药物敏感性分析	6
4 第四部分	7
4.1 使用 BiomaRt 注释靶点基因	7
4.2 蛋白互作和 Hubgenes 筛选	13
4.2.1 蛋白互作	13
4.2.2 Hubgenes 筛选	14
4.3 使用 clusterProfiler 富集分析	15
Reference	24

1 第一部分

1.1 ETCM 中药丹参的化合物以及靶点基因

1.1.1 获取 ETCM 网站数据

通过编写 R 函数以快速获取 ETCM 网站的中药和对应靶点数据。

丹参的 96 种化合物和相关靶基因概览。其中，靶点基因（非重复）共 216 个。(对应文件为 ./96_components.txt, ./components_and_target_genes.csv)

```
## # A tibble: 532 x 3
##   components          genes links
##   <chr>              <chr> <chr>
## 1 Sitosterol,Î' -Sitosterol AKR1C1 /ETCM/index.php/Home/Index/jyjb_details.html?gene=AKR1C1
## 2 Sitosterol,Î' -Sitosterol AKR1C2 /ETCM/index.php/Home/Index/jyjb_details.html?gene=AKR1C2
## 3 Sitosterol,Î' -Sitosterol AR      /ETCM/index.php/Home/Index/jyjb_details.html?gene=AR
## 4 Sitosterol,Î' -Sitosterol CLEC4E /ETCM/index.php/Home/Index/jyjb_details.html?gene=CLEC4E
## 5 Sitosterol,Î' -Sitosterol ESR1    /ETCM/index.php/Home/Index/jyjb_details.html?gene=ESR1
## 6 Sitosterol,Î' -Sitosterol ESR2    /ETCM/index.php/Home/Index/jyjb_details.html?gene=ESR2
## 7 Sitosterol,Î' -Sitosterol GABRA1 /ETCM/index.php/Home/Index/jyjb_details.html?gene=GABRA1
## 8 Sitosterol,Î' -Sitosterol GABRA2 /ETCM/index.php/Home/Index/jyjb_details.html?gene=GABRA2
## 9 Sitosterol,Î' -Sitosterol GABRA3 /ETCM/index.php/Home/Index/jyjb_details.html?gene=GABRA3
## 10 Sitosterol,Î' -Sitosterol GABRA4 /ETCM/index.php/Home/Index/jyjb_details.html?gene=GABRA4
## # i 522 more rows
```

1.1.2 获取 HERB 网站数据

由于 ETCM 缺少相当一部分化合物的靶点数据，因而使用 HERB 数据库补充。（对于丹参，HERB 包含更多的化合物）（对应文件为 `./HERB_compounds_target.tsv`）

1.1.3 获取 PubChem 数据库关于化合物的别名

ETCM 数据库是个封闭的网站，不包含和任何其它数据库相同的 ID 信息。为了以 HERB 数据库的靶点数据补充 ETCM 的化合物靶点数据，根据 PubChem CID（HERB 数据库提供）搜索 PubChem 获得化合物的别名。（对应文件为 `./synos.tsv`）

1.1.4 以 HERB 的数据补充 ETCM 的数据

以下是可以在 HERB 找到靶点基因的化合物，但是在 ETCM 找不到靶点基因的化合物：

```
##      cid          syno
## 1: 11425923 Dihydratanthinone I
## 2: 11600642      Danshensu
## 3:   68081      Isoimperatorin
## 4: 44425165 Neocryptotanshinone
## 5:   160254      Cryptotanshinone
## 6: 11629084      Danshensuan B
## 7: 3082765      Dehydromiltirone
## 8:   626608 Isocryptotanshinone
## 9:   622085      Przewaquinone B
## 10:  126072      Tanshindiol C
## 11:   389885      Salvilenone
## 12:  5321622      Tanshinol A
## 13:   114917      Tanshinone I
```

尽管如此，由于 ETCM 缺乏其他数据库的索引，还是有一部分的化合物不知道来源，所以无法从其他数据库得到靶点基因的补充：

## [1] "Methylene Tanshinquinone"	"Dihydroisotanshinone I"	"Tanshinlactone"
## [4] "Salvianolic Acid B"	"Danshexinkum D"	"Danshexinkum A"
## [7] "Methyl Tanshinonate"	"Ethyl Lithospermate"	"Salvinone"
## [10] "Isotanshinoneii B"	"Prgewaquinone A"	"Danshenol A"
## [13] "Danshenspiroketalactone"	"Danshenxinkun A"	"Danshenxinkun B"
## [16] "Danshenxinkun C"	"Danshenxinkun D"	"Î" 1-Dehydrotanshinone"
## [19] "Dihydroisotanshinone I"	"1,2-Dihydrotanshinquinone"	"Epidanshenspiroketal"
## [22] "3-Î' -Hydroxymethylenetanshinquinone"	"3Î' -Hydroxytanshinone Iia"	"3Î' -Hydroxytanshinone"
## [25] "Isotanshinone I"	"Isotanshinone Iia"	"Lithospermate B"
## [28] "Magnesium Lithospermate B"	"Methyl Tanshinonate"	"Monomethyl Lithospermate"
## [31] "Nortanshinone"	"Salvianolic Acid B"	"Salvianolic Acid C"
## [34] "Salvianolic Acid G"	"Salvinone"	"Tanshindiol A"
## [37] "Tanshindiol B"	"Tanshinlactone"	"Tanshinol B"
## [40] "Tanshinone Iia"	"Tanshinone Iib"	"Tanshinone Vi"

1.1.5 总结

HERB 数据库包含更多的化合物和靶点信息，所以以下分析以 HERB 数据库为主。HERB 记录的丹参的化合物有 330 个，能找到靶点基因信息的化合物有 187 个。(对应文件为 ./HERB_compounds_of_danshen.xlsx, ./HERB_targets_of_compounds.xlsx)

2 第二部分

2.1 在 genecards 网站上检索胃癌相关的基因

在网站 Genecards 检索胃癌，获取相关数据后，根据 Relevance score 进行筛选 (> 5)。

所有胃癌相关基因概览，基因数量为 4475 (对应文件为 ./all_gastric_Cancer_related_genes.csv)

```
## # A tibble: 4,475 x 8
```

##	Gene.Symbol	Description	Category	Uniprot.ID	Gifts	GC.Id	Relevance
##	<chr>	<chr>	<chr>	<chr>	<int>	<chr>	
##	1	CDH1	Cadherin 1	Protein Coding	P12830	56	GC16P068737
##	2	BRCA2	BRCA2 DNA Repair Associated	Protein Coding	P51587	54	GC13P032315
##	3	BRCA1	BRCA1 DNA Repair Associated	Protein Coding	P38398	57	GC17M043044
##	4	TP53	Tumor Protein P53	Protein Coding	P04637	60	GC17M007661
##	5	APC	APC Regulator Of WNT Signaling Pathway	Protein Coding	P25054	56	GC05P112707
##	6	CHEK2	Checkpoint Kinase 2	Protein Coding	O96017	61	GC22M028687
##	7	PALB2	Partner And Localizer Of BRCA2	Protein Coding	Q86YC2	51	GC16M023603
##	8	ATM	ATM Serine/Threonine Kinase	Protein Coding	Q13315	60	GC11P108222
##	9	MLH1	MutL Homolog 1	Protein Coding	P40692	56	GC03P036993

```
## 10 BRIP1          BRCA1 Interacting Helicase 1          Protein Coding Q9BX63          56 GC17M061679
## # i 4,465 more rows
```

根据丹参靶点基因过滤数据集，即，将筛选过胃癌基因数据和丹参数据根据基因合并。

韦恩图见 Figure 1，说明：韦恩图分三个区域，左侧加上中间区域对应 187 个化合物所有靶点基因；中间区域加上右侧区域对应所有胃癌相关基因。(对应文件为 `./figs/venn_plot.pdf`)

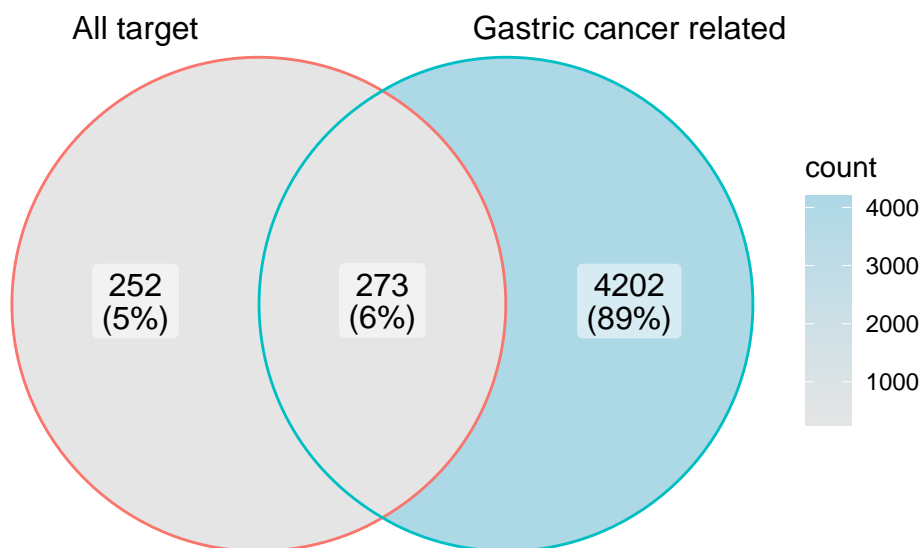


Figure 1: 靶点基因和胃癌相关基因交集韦恩图

187 个化合物和胃癌相关基因的交集数据概览 (包含交集基因以及对应化合物)，化合物和靶点基因一一对应，化合物共 187 个，非重复基因 273 个，与上述韦恩图一致。(对应文件为 `./gastric_Cancer_related_genes_Intersect_with_targetGenes_components.csv`)

```
## # A tibble: 2,083 x 9
##   components          Gene.Symbol Description Category Uniprot.ID Gifts GC.Id
##   <chr>              <chr>      <chr>      <chr>      <chr>      <int> <chr>
## 1 (-)-Epicedrol      ACHE        Acetylchol~ Protein~  P22303      53 GC07~
## 2 (-)-Epicedrol      DPP4        Dipeptidyl~ Protein~  P27487      56 GC02~
## 3 (-)-beta-Phellandrene ACHE        Acetylchol~ Protein~  P22303      53 GC07~
## 4 (-)-beta-Phellandrene DPP4        Dipeptidyl~ Protein~  P27487      56 GC02~
## 5 (-)-beta-Phellandrene NR3C1       Nuclear Re~ Protein~  P04150      56 GC05~
## 6 (-)-beta-Phellandrene PRSS1       Serine Pro~ Protein~  P07477      52 GC07~
## 7 (1R,4R,5S)-1-isopropyl-4-methyl-4-bicyc~ ACHE        Acetylchol~ Protein~  P22303      53 GC07~
## 8 (1R,4R,5S)-1-isopropyl-4-methyl-4-bicyc~ DPP4        Dipeptidyl~ Protein~  P27487      56 GC02~
## 9 (1R,4S,4aR,8aR)-4-isopropyl-1,6-dimethy~ ACHE        Acetylchol~ Protein~  P22303      53 GC07~
## 10 (1R,4S,4aR,8aR)-4-isopropyl-1,6-dimethy~ AR          Androgen R~ Protein~  P10275      58 GC0X~
## # i 2,073 more rows
```

3 第三部分

3.1 使用 CellMiner 数据库的 NCI-60 数据集

下载并预处理 NCI-60 的数据以备药物敏感性分析。

Cisplatin 活性数据, 包含 60 个癌细胞的活性 IC50 Z-score :

```
## # A tibble: 1 x 61
##   `Drug name` `BR:MCF7` `BR:MDA-MB-231` `BR:HS 578T` `BR:BT-549` `BR:T-47D` `CNS:SF-268` `CNS:SF-295`
##   <chr>      <chr>      <chr>      <chr>      <chr>      <chr>      <chr>      <chr>
## 1 Cisplatin  0.26      -1.8      -0.56      -0.29      -1.32      1.42      1.15
## # i 51 more variables: `CNS:SNB-75` <chr>, `CNS:U251` <chr>, `CO:COLO 205` <chr>, `CO:HCC-2998` <chr>,
## #   `CO:HCT-15` <chr>, `CO:HT29` <chr>, `CO:KM12` <chr>, `CO:SW-620` <chr>, `LE:CCRF-CEM` <chr>, `LE:
## #   `LE:K-562` <chr>, `LE:MOLT-4` <chr>, `LE:RPMI-8226` <chr>, `LE:SR` <chr>, `ME:LOX IMVI` <chr>, `ME:
## #   `ME:M14` <chr>, `ME:SK-MEL-2` <chr>, `ME:SK-MEL-28` <chr>, `ME:SK-MEL-5` <chr>, `ME:UACC-257` <chr>,
## #   `ME:MDA-MB-435` <chr>, `ME:MDA-N` <chr>, `LC:A549/ATCC` <chr>, `LC:EKVX` <chr>, `LC:HOP-62` <chr>,
## #   `LC:NCI-H226` <chr>, `LC:NCI-H23` <chr>, `LC:NCI-H322M` <chr>, `LC:NCI-H460` <chr>, `LC:NCI-H522` <chr>,
## #   `OV:OVCAR-3` <chr>, `OV:OVCAR-4` <chr>, `OV:OVCAR-5` <chr>, `OV:OVCAR-8` <chr>, `OV:SK-OV-3` <chr>
```

NCI-60 表达数据, 包含 60 个癌细胞的基因表达数据 (FPKM):

```
## # A tibble: 270 x 67
##   `Gene name d` `Entrez gene id e` `Chromosome f` `Start f` `End f` `Cytoband f` `BR:MCF7` `BR:MDA-MB-231`
##   <chr>      <dbl> <chr>      <dbl>      <dbl> <chr>      <dbl>
## 1 PARK7      11315 1      8021713  8045342 1p36.23  5.42
## 2 PIK3CD      5293 1      9711789  9789172 1p36.2   0.132
## 3 MTOR        2475 1      11166587 11322608 1p36.2   2.54
## 4 CTRC        11330 1      15764937 15773153 1p36.21  0
## 5 CASP9        842 1      15817895 15851285 1p36.21  0.669
## 6 LCK          3932 1      32716839 32751768 1p34.3   0
## 7 AKR1A1      10327 1      46016454 46035723 1p33-p32 3.38
## 8 JUN          3725 1      59246462 59249785 1p32-p31 0.76
## 9 GCLM         2730 1      94350755 94375154 1p22.1   1.02
## 10 VCAM1       7412 1      101185195 101204601 1p32-p31 0
## # i 260 more rows
## # i 58 more variables: `BR:BT-549` <dbl>, `BR:T-47D` <dbl>, `CNS:SF-268` <dbl>, `CNS:SF-295` <dbl>,
## #   `CNS:SNB-19` <dbl>, `CNS:SNB-75` <dbl>, `CNS:U251` <dbl>, `CO:COLO 205` <dbl>, `CO:HCC-2998` <dbl>,
## #   `CO:HCT-15` <dbl>, `CO:HT29` <dbl>, `CO:KM12` <dbl>, `CO:SW-620` <dbl>, `LE:CCRF-CEM` <dbl>, `LE:
## #   `LE:K-562` <dbl>, `LE:MOLT-4` <dbl>, `LE:RPMI-8226` <dbl>, `LE:SR` <dbl>, `ME:LOX IMVI` <dbl>, `ME:
## #   `ME:M14` <dbl>, `ME:SK-MEL-2` <dbl>, `ME:SK-MEL-28` <dbl>, `ME:SK-MEL-5` <dbl>, `ME:UACC-257` <dbl>,
## #   `ME:MDA-MB-435` <dbl>, `ME:MDA-N` <dbl>, `LC:A549/ATCC` <dbl>, `LC:EKVX` <dbl>, `LC:HOP-62` <dbl>
```

3.2 药物敏感性分析

将药物活性数据和基因表达数据关联分析 (Pearson)。

其中有显著性意义的有 39 个基因 ($p < 0.05$)，可视化见 Figure 2 (对应文件为 `figs/pearsonTest.pdf`)。这意味着，与顺铂协作的靶点基因有 39 个。

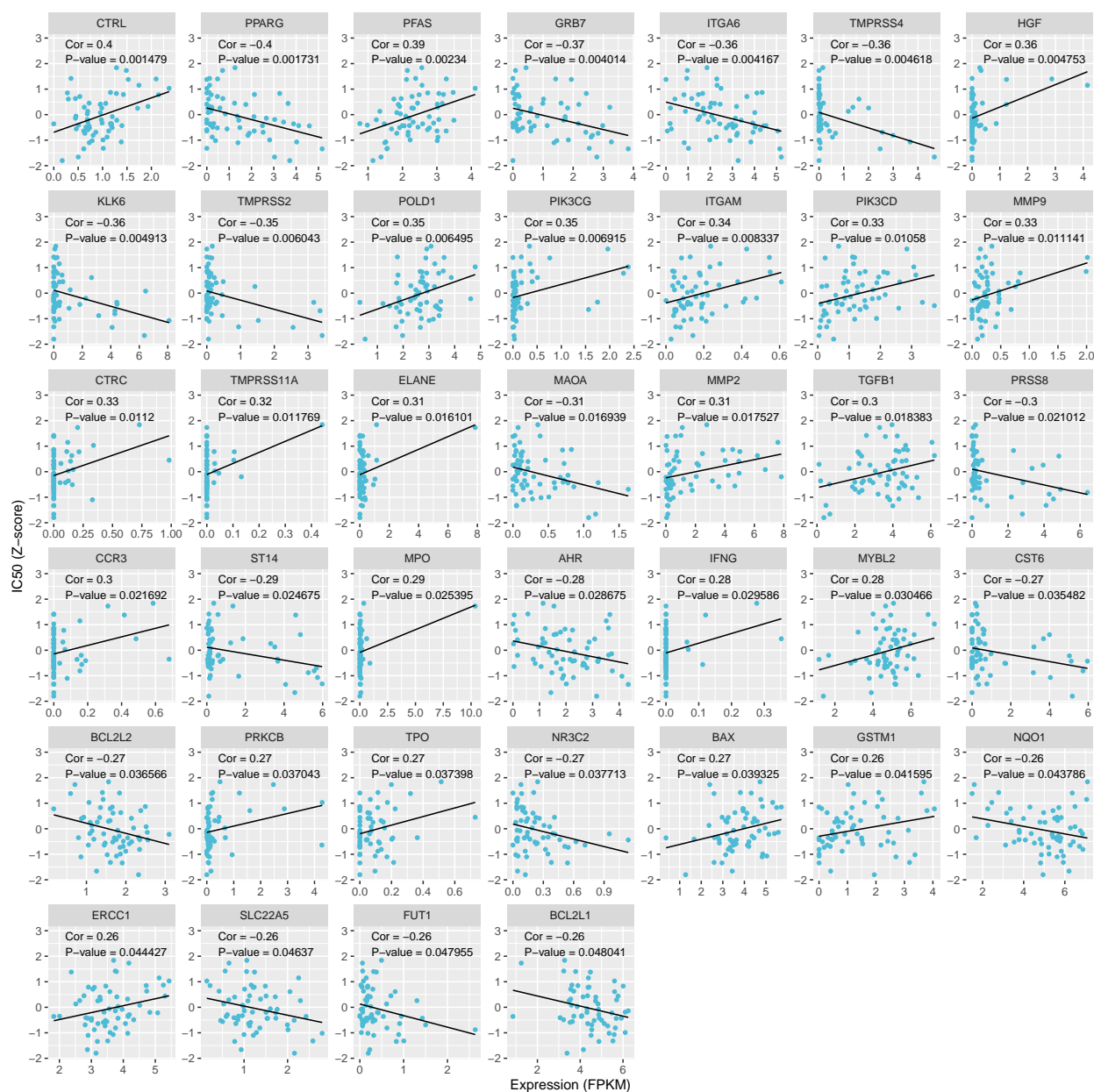


Figure 2: 关联性分析回归曲线图

将这 39 个显著基因的分析数据与 187 个化合物及其靶点基因数据合并，得到作用于显著基因的化合物数据。

关联性分析(Pearson)结果概览(已包含基因和对应化合物数据),(对应文件为 `./pearsonTest_allResults.csv`, `./pearsonTest_results_with_components.csv`) 其中，协同顺铂靶基因的化合物共有 73 个。

```
## # A tibble: 117 x 11
```

```
##      name      cor p.value components      Description      Category Uniprot.ID Gifts GC.Id
##      <chr>      <dbl>      <dbl> <chr>      <chr>      <chr>      <chr>      <int> <chr>
##  1 AHR      -0.283    0.0287 kaempferol      Aryl Hydrocarbo~ Protein~ P35869      53 GC07~
##  2 BAX      0.267    0.0393 aloemodin      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  3 BAX      0.267    0.0393 cryptotanshinone      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  4 BAX      0.267    0.0393 kaempferol      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  5 BAX      0.267    0.0393 rhein      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  6 BAX      0.267    0.0393 aucubin      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  7 BAX      0.267    0.0393 tanshinone i      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  8 BAX      0.267    0.0393 beta-sitosterol      BCL2 Associated~ Protein~ Q07812      57 GC19~
##  9 BAX      0.267    0.0393 protocatechuic acid      BCL2 Associated~ Protein~ Q07812      57 GC19~
## 10 BCL2L1 -0.256    0.0480 15,16-dihydrotanshinone i BCL2 Like 1      Protein~ Q07817      54 GC20~
## # i 107 more rows
```

4 第四部分

4.1 使用 BiomaRt 注释靶点基因

使用 R 包 biomaRt 获取靶点基因的 Entrezgene id 以便后续分析。

```
## # A tibble: 39 x 3
##      ensembl_gene_id entrezgene_id hgnc_symbol
##      <chr>      <int> <chr>
##  1 ENSG00000106546      196 AHR
##  2 ENSG00000087088      581 BAX
##  3 ENSG00000171552      598 BCL2L1
##  4 ENSG00000129473      599 BCL2L2
##  5 ENSG00000183625     1232 CCR3
##  6 ENSG00000175315     1474 CST6
##  7 ENSG00000162438    11330 CTRC
##  8 ENSG00000141086     1506 CTRL
##  9 ENSG00000277571     1991 ELANE
## 10 ENSG00000012061     2067 ERCC1
## # i 29 more rows
```

将注释数据与筛选的化合物的靶点基因数据合并，并按照化合物分组。

各个化合物包含的显著性靶点基因数量信息：

```
## $`^((2R)-3-(3,4-dihydroxyphenyl)-2-[(Z)-3-(3,4-dihydroxyphenyl)acryloyl]oxy-propionic acid`
## [1] 1
##
## $`^((2S,3S)-2-(3,4-dihydroxyphenyl)-7-hydroxy-4-[(E)-3-hydroxy-3-oxoprop-1-enyl]-2,3-dihydrobenzofuran
## [1] 1
```

```

##
## $(6S)-6-(hydroxymethyl)-1,6-dimethyl-8,9-dihydro-7H-naphtho[8,7-g]benzofuran-10,11-dione`
## [1] 1
##
## $(E)-3-(3-hydroxy-4,5-dimethoxy-phenyl)acrylic acid`
## [1] 1
##
## $(E)-3-[2-(3,4-dihydroxyphenyl)-7-hydroxy-benzofuran-4-yl]acrylic acid`
## [1] 1
##
## $(R)-p-Menth-1-en-4-ol`
## [1] 2
##
## $1-methyl-8,9-dihydro-7H-naphtho[5,6-g]benzofuran-6,10,11-trione`
## [1] 1
##
## $1,2-DT-Quinone`
## [1] 1
##
## $1,2,5,6-tetrahydrotanshinone`
## [1] 1
##
## $15,16-dihydrotanshinone i`
## [1] 1
##
## $2-isopropyl-8-methylphenanthrene-3,4-dione`
## [1] 1
##
## $3-beta-Hydroxymethyllenetanshiquinone`
## [1] 1
##
## $3beta-Hydroxytanshinone IIA`
## [1] 1
##
## $3-hydroxytanshinone a`
## [1] 1
##
## $4-methylenemiltirone`
## [1] 1
##
## $7-oxoroleanone2`
## [1] 1

```



```

##
## $aloeemodin
## [1] 3
##
## $\alpha$-amyrin`
## [1] 2
##
## $aucubin
## [1] 1
##
## $\beta$-sitosterol`
## [1] 4
##
## $\text{caffeic acid}$`
## [1] 2
##
## $carnosol
## [1] 2
##
## $\text{chlorogenic acid}$`
## [1] 1
##
## $cryptotanshinone
## [1] 3
##
## $cyanidol
## [1] 1
##
## $\text{dan-shexinkum b}$`
## [1] 1
##
## $\text{dan-shexinkum d}$`
## [1] 1
##
## $\text{Danshenol A}$`
## [1] 1
##
## $danshensu
## [1] 1
##
## $\text{Dehydrotanshinone II A}$`
## [1] 1

```

```

##
## $dihydroisotanshinone
## [1] 1
##
## $dihydrotanshinlactone
## [1] 1
##
## $`dihydrotanshinone i`
## [1] 2
##
## $dihydrotanshinone
## [1] 1
##
## $dimethylolithospermate
## [1] 1
##
## $DTY
## [1] 1
##
## $EIC
## [1] 1
##
## $`ferulic acid`
## [1] 1
##
## $formyltanshinone
## [1] 1
##
## $GLY
## [1] 2
##
## $`isoferulic acid`
## [1] 2
##
## $`isotanshinone i`
## [1] 1
##
## $kaempferol
## [1] 5
##
## $labiatenicacid
## [1] 1

```

```

##
## $Methylenetanshinquinone
## [1] 1
##
## $methylosmarinate
## [1] 1
##
## $methyltanshinonate
## [1] 1
##
## $`Mono-O-methylwightin`
## [1] 1
##
## $Nortrachelogenin
## [1] 1
##
## $`oleanolic acid`
## [1] 2
##
## $`palmitic acid`
## [1] 3
##
## $PHA
## [1] 1
##
## $Poriferasterol
## [1] 1
##
## $`prolithospermic acid`
## [1] 1
##
## $`protocatechuic acid`
## [1] 2
##
## $`przewalskin a`
## [1] 1
##
## $`przewalskin b`
## [1] 1
##
## $`Przewaquinone B`
## [1] 1

```

```

##
## $`przewaquinone f`
## [1] 1
##
## $rhein
## [1] 1
##
## $`Rosemary acid`
## [1] 1
##
## $rutin
## [1] 2
##
## $`Sal A`
## [1] 2
##
## $`salvianic acid c`
## [1] 1
##
## $`salvianolic acid a`
## [1] 4
##
## $Salvigenin
## [1] 1
##
## $`salvilenone `
## [1] 1
##
## $Tanshilactone
## [1] 1
##
## $tanshinaldehyde
## [1] 1
##
## $`Tanshinol A`
## [1] 1
##
## $`tanshinone i`
## [1] 18
##
## $`tanshinone `
## [1] 1

```

```
##
## $`Z-8-Hexadecen-1-ol acetate`
## [1] 1
```

除了 tanshinone i, 其他化合物都不超过 5 个靶点基因。

4.2 蛋白互作和 Hubgenes 筛选

4.2.1 蛋白互作

R 包 STRINGdb 提供网站<https://www.string-db.org/>的 API, 用以绘制蛋白质互作网络。

可视化的蛋白质互作网络图为见图3 (对应文件为 **figs/protein_interaction.pdf**)

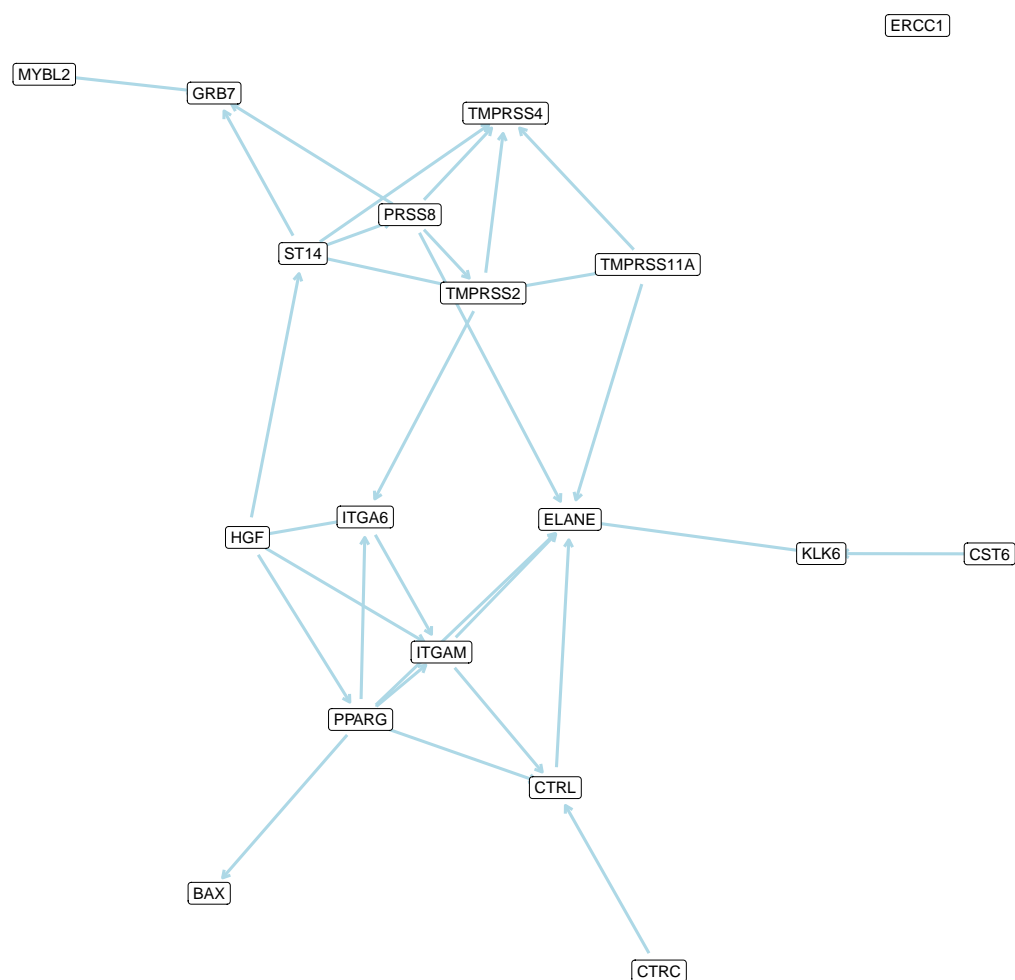


Figure 3: 药物敏感性基因的蛋白质互作图

4.2.2 Hubgenes 筛选

利用 Cytoscape 的插件 CytoHubba¹ 提供的 MCC 算法计算 Hub 基因得分（这里 MCC 算法被集成到 R 中，独立计算）。

以下为结果概览：（对应文件为 `./tanshinone.iMCC_score.xlsx`）

```
## # A tibble: 18 x 15
##   genes MCC_score STRING_id cor p.value components Description Category Uniprot.ID Gifts GC.Id
##   <chr>      <dbl> <chr>      <dbl>  <dbl> <chr>      <chr>      <chr>      <chr>      <int> <chr>
## 1 PPARG      13 9606.ENS~ -0.396 0.00173 tanshinon~ Peroxisome~ Protein~ P37231      60 GC03~
## 2 ITGAM      12 9606.ENS~  0.338 0.00834 tanshinon~ Integrin S~ Protein~ P11215      54 GC16~
## 3 ST14        9 9606.ENS~ -0.290 0.0247  tanshinon~ ST14 Trans~ Protein~ Q9Y5Y6      55 GC11~
## 4 PRSS8        9 9606.ENS~ -0.297 0.0210  tanshinon~ Serine Pro~ Protein~ Q16651      52 GC16~
## 5 TMPRS~      9 9606.ENS~ -0.350 0.00604 tanshinon~ Transmembr~ Protein~ O15393      54 GC21~
## 6 ELANE        9 9606.ENS~  0.310 0.0161  tanshinon~ Elastase, ~ Protein~ P08246      58 GC19~
## 7 TMPRS~      8 9606.ENS~ -0.361 0.00462 tanshinon~ Transmembr~ Protein~ Q9NRS4      46 GC11~
## 8 HGF          7 9606.ENS~  0.360 0.00475 tanshinon~ Hepatocyte~ Protein~ P14210      58 GC07~
## 9 ITGA6        7 9606.ENS~ -0.365 0.00417 tanshinon~ Integrin S~ Protein~ P23229      57 GC02~
## 10 CTRL        7 9606.ENS~  0.401 0.00148 tanshinon~ Chymotryps~ Protein~ P40313      47 GC16~
## 11 TMPRS~      3 9606.ENS~  0.323 0.0118  tanshinon~ Transmembr~ Protein~ Q6ZMR5      43 GC04~
## 12 GRB7        3 9606.ENS~ -0.366 0.00401 tanshinon~ Growth Fac~ Protein~ Q14451      50 GC17~
## 13 KLK6        2 9606.ENS~ -0.359 0.00491 tanshinon~ Kallikrein~ Protein~ Q92876      51 GC19~
## 14 MYBL2        1 9606.ENS~  0.280 0.0305  tanshinon~ MYB Proto~ Protein~ P10244      49 GC20~
## 15 BAX          1 9606.ENS~  0.267 0.0393  tanshinon~ BCL2 Assoc~ Protein~ Q07812      57 GC19~
## 16 CST6         1 9606.ENS~ -0.272 0.0355  tanshinon~ Cystatin E~ Protein~ Q15828      48 GC11~
## 17 CTRC         1 9606.ENS~  0.325 0.0112  tanshinon~ Chymotryps~ Protein~ Q99895      52 GC01~
## 18 ERCC1       NA 9606.ENS~  0.260 0.0444  tanshinon~ ERCC Excis~ Protein~ P07992      53 GC19~
## # i 2 more variables: ensembl_gene_id <chr>, entrezgene_id <int>
```

将结果可视化，见 Figure 4

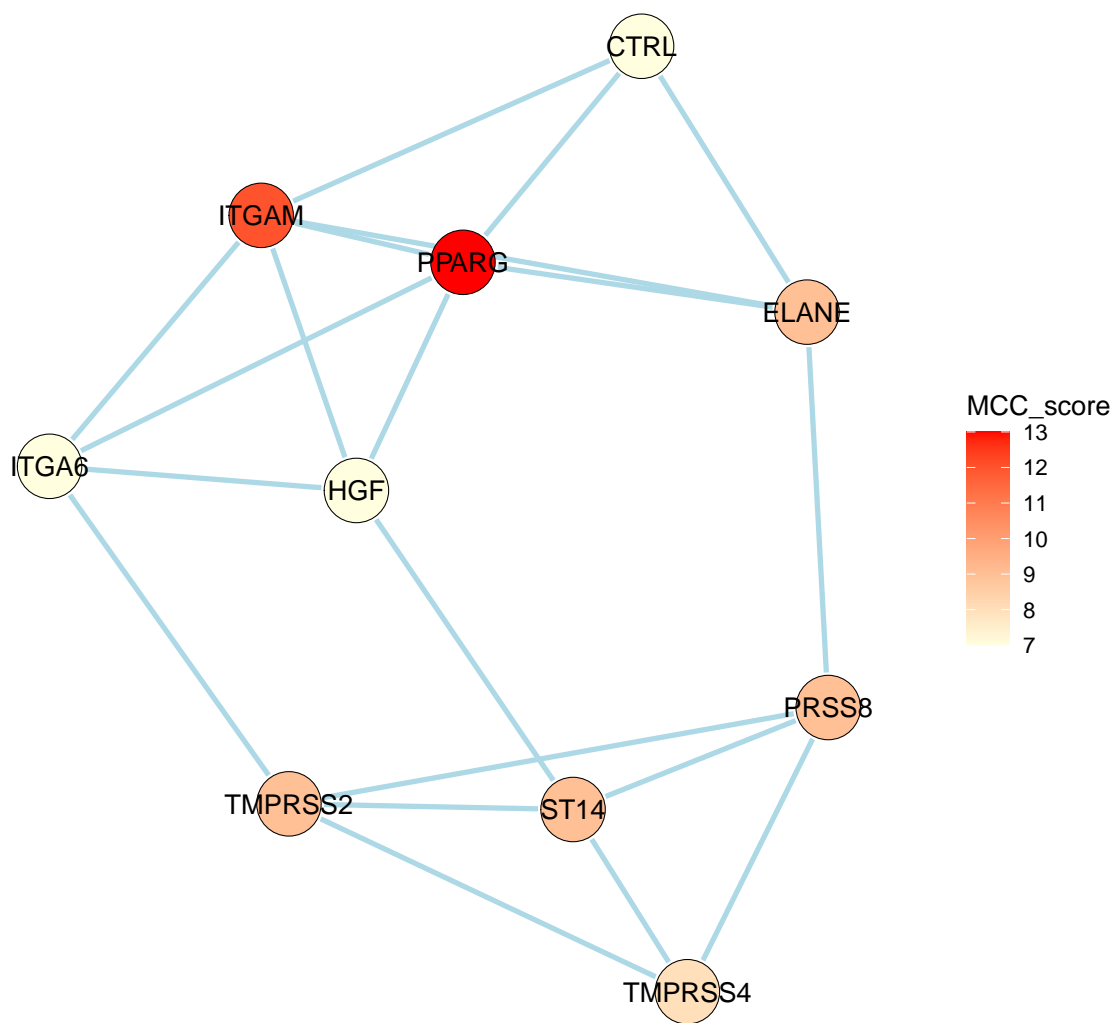


Figure 4: MCC score of gene targets of Tanshinone i

4.3 使用 clusterProfiler 富集分析

tanshinone i 的富集分析用 MCC top 10 的靶点基因进行，其他化合物的靶点基因直接以关联分析筛选过的基因富集。

以下为 tanshinone i 的 KEGG 和 GO 富集图：

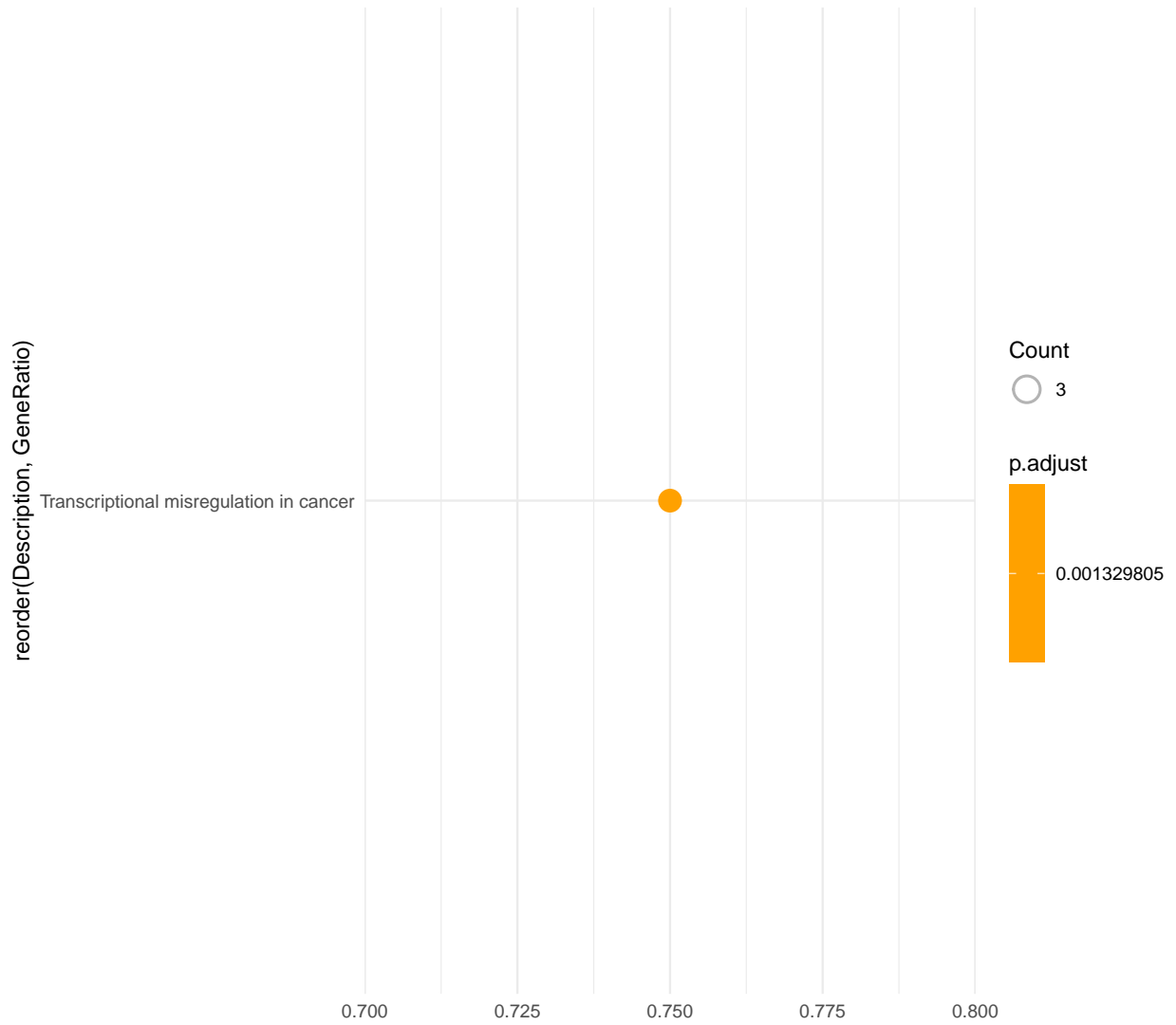


Figure 5: KEGG enrichment

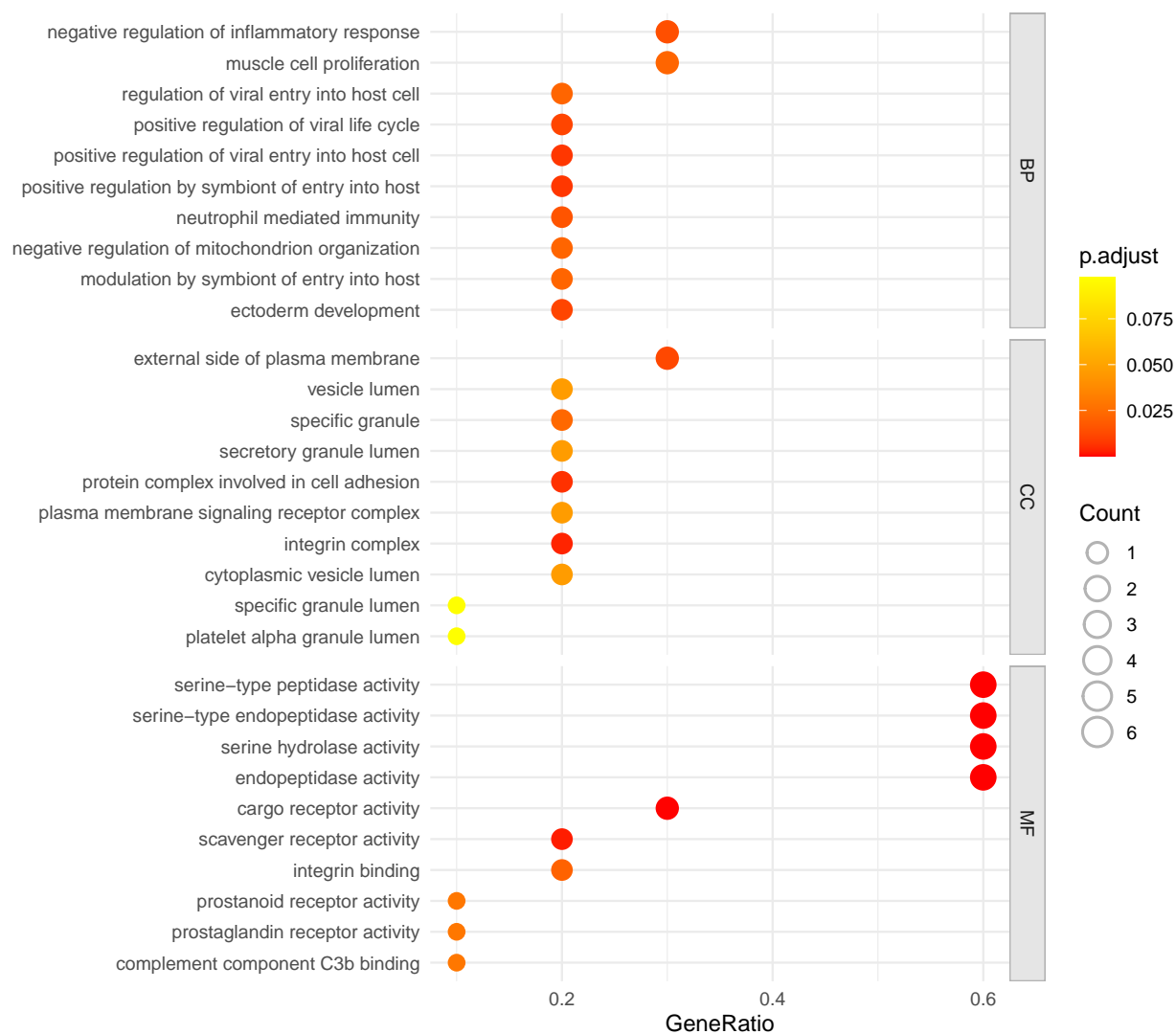


Figure 6: GO enrichment

图片数量较多, 不一一展示 (KEGG 富集共 73 个, GO 富集共 73)。(对应文件为 ./enrichGO, ./enrichKEGG)
对富集图的解释, 可以参考文献²。

说明: KEGG 富集分析都有结果; 但是对于 GO 富集分析 (BP, CC 或 MF) 中, 个别化合物有靶点基因, 但未映射到通路中的基因, 所以无结果, 这些是 (TRUE 表示有结果, 而 FALSE 表示无结果):

```
## $(2R)-3-(3,4-dihydroxyphenyl)-2-[(Z)-3-(3,4-dihydroxyphenyl)acryloyl]oxy-propionic acid`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $(2S,3S)-2-(3,4-dihydroxyphenyl)-7-hydroxy-4-[(E)-3-hydroxy-3-oxoprop-1-enyl]-2,3-dihydrobenzofuran
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $(6S)-6-(hydroxymethyl)-1,6-dimethyl-8,9-dihydro-7H-naphtho[8,7-g]benzofuran-10,11-dione`
```

```

## BP CC MF
## TRUE TRUE TRUE
##
## $(E)-3-(3-hydroxy-4,5-dimethoxy-phenyl)acrylic acid`
## BP CC MF
## TRUE TRUE TRUE
##
## $(E)-3-[2-(3,4-dihydroxyphenyl)-7-hydroxy-benzofuran-4-yl]acrylic acid`
## BP CC MF
## TRUE TRUE TRUE
##
## $(R)-p-Menth-1-en-4-ol`
## BP CC MF
## TRUE TRUE TRUE
##
## $`1-methyl-8,9-dihydro-7H-naphtho[5,6-g]benzofuran-6,10,11-trione`
## BP CC MF
## TRUE FALSE TRUE
##
## $`1,2-DT-Quinone`
## BP CC MF
## TRUE TRUE TRUE
##
## $`1,2,5,6-tetrahydrotanshinone`
## BP CC MF
## TRUE TRUE TRUE
##
## $`15,16-dihydrotanshinone i`
## BP CC MF
## TRUE TRUE TRUE
##
## $`2-isopropyl-8-methylphenanthrene-3,4-dione`
## BP CC MF
## TRUE TRUE TRUE
##
## $`3-beta-Hydroxymethyllenetanshiquinone`
## BP CC MF
## TRUE TRUE TRUE
##
## $`3beta-Hydroxytanshinone IIA`
## BP CC MF
## TRUE TRUE TRUE

```

```

##
## $`3-hydroxytanshinone a`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`4-methylenemiltirone`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`7-oxoroyleanone2`
##   BP   CC   MF
## TRUE FALSE TRUE
##
## $aloeemodin
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`alpha-amyrin`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $aucubin
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`beta-sitosterol`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`caffeic acid`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $carnosol
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`chlorogenic acid`
##   BP   CC   MF
## TRUE FALSE TRUE
##
## $cryptotanshinone

```

```

## BP CC MF
## TRUE TRUE TRUE
##
## $cyanidol
## BP CC MF
## TRUE TRUE TRUE
##
## $`dan-shexinkum b`
## BP CC MF
## TRUE TRUE TRUE
##
## $`dan-shexinkum d`
## BP CC MF
## TRUE TRUE TRUE
##
## $`Danshenol A`
## BP CC MF
## TRUE TRUE TRUE
##
## $danshensu
## BP CC MF
## TRUE TRUE TRUE
##
## $`Dehydrotanshinone II A`
## BP CC MF
## TRUE TRUE TRUE
##
## $dihydroisotanshinone
## BP CC MF
## TRUE TRUE TRUE
##
## $dihydrotanshinlactone
## BP CC MF
## TRUE TRUE FALSE
##
## $`dihydrotanshinone i`
## BP CC MF
## TRUE TRUE TRUE
##
## $dihydrotanshinone
## BP CC MF
## TRUE TRUE TRUE

```

```

##
## $dimethylolithospermate
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $DTY
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $EIC
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`ferulic acid`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $formyltanshinone
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $GLY
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`isoferulic acid`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`isotanshinone i`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $kaempferol
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $labiatenicacid
##   BP   CC   MF
## TRUE FALSE TRUE
##
## $Methylenetanshinquinone

```

```

## BP CC MF
## TRUE TRUE TRUE
##
## $methylosmarinate
## BP CC MF
## TRUE TRUE TRUE
##
## $methylosmarinate
## BP CC MF
## TRUE TRUE TRUE
##
## $`Mono-O-methylwightin`
## BP CC MF
## TRUE TRUE TRUE
##
## $Nortrachelogenin
## BP CC MF
## TRUE TRUE TRUE
##
## $`oleanolic acid`
## BP CC MF
## TRUE TRUE TRUE
##
## $`palmitic acid`
## BP CC MF
## TRUE TRUE TRUE
##
## $PHA
## BP CC MF
## TRUE TRUE TRUE
##
## $Poriferasterol
## BP CC MF
## TRUE FALSE TRUE
##
## $`prolithospermic acid`
## BP CC MF
## TRUE TRUE TRUE
##
## $`protocatechuic acid`
## BP CC MF
## TRUE TRUE TRUE

```

```

##
## $`przewalskin a`
##   BP   CC   MF
## TRUE FALSE TRUE
##
## $`przewalskin b`
##   BP   CC   MF
## TRUE FALSE TRUE
##
## $`Przewaquinone B`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`przewaquinone f`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $rhein
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`Rosemary acid`
##   BP   CC   MF
## TRUE FALSE TRUE
##
## $rutin
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`Sal A`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`salvianic acid c`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $`salvianolic acid a`
##   BP   CC   MF
## TRUE TRUE TRUE
##
## $Salvigenin

```

```

## BP CC MF
## TRUE TRUE TRUE
##
## $`salvilenone `
## BP CC MF
## TRUE TRUE TRUE
##
## $Tanshilactone
## BP CC MF
## TRUE TRUE TRUE
##
## $tanshinaldehyde
## BP CC MF
## TRUE TRUE TRUE
##
## $`Tanshinol A`
## BP CC MF
## TRUE TRUE TRUE
##
## $`tanshinone i`
## BP CC MF
## TRUE TRUE TRUE
##
## $`tanshinone `
## BP CC MF
## TRUE TRUE TRUE
##
## $`Z-8-Hexadecen-1-ol acetate`
## BP CC MF
## TRUE TRUE TRUE

```

Reference

1. Chin, C.-H. *et al.* CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**, S11 (2014).
2. Liu, Y. *et al.* Integrative analyses of biomarkers and pathways for adipose tissue after bariatric surgery. *Adipocyte* **9**, 384–400 (2020).