

# 筛选差异蛋白和对应配体蛋白

2024-03-08

LiChuang Huang



@ 立效研究院

# Contents

<b>1 摘要</b>	<b>1</b>
1.1 需求	1
1.2 结果	1
1.2.1 寻找差异致癌膜蛋白及对应配体蛋白	1
1.2.2 耐药差异膜蛋白及对应抗体蛋白或互作抑制其表达蛋白	1
<b>2 前言</b>	<b>2</b>
<b>3 材料和方法</b>	<b>2</b>
3.1 材料	2
3.2 方法	2
<b>4 分析结果</b>	<b>2</b>
<b>5 结论</b>	<b>2</b>
<b>6 附：分析流程——寻找差异致癌膜蛋白及对应配体蛋白</b>	<b>2</b>
6.1 结直肠癌差异蛋白	2
6.1.1 数据来源	2
6.1.2 差异蛋白	3
6.2 膜蛋白筛选	5
6.2.1 Unitmp	5
6.2.2 与高表达差异蛋白 (DPs-Up) 交集	6
6.3 以蛋白互作筛选配体蛋白	7
6.3.1 PPI	7
6.3.2 富集分析	7
6.3.3 根据富集结果缩减 PPI	10
<b>7 附：分析流程——耐药差异膜蛋白及对应抗体蛋白或互作抑制其表达蛋白</b>	<b>11</b>
7.1 结肠癌差异蛋白	11
7.1.1 数据来源	11
7.2 预测耐药性	13
7.2.1 使用 RNA 数据集预测	13
7.3 差异蛋白筛选	16
7.3.1 元数据	16
7.3.2 差异蛋白	17
7.4 膜蛋白筛选	18
7.4.1 Unitmp	18
7.4.2 与高表达差异蛋白 (TCGA-dps-Up) 交集	18
7.5 以蛋白互作筛选配体蛋白	19
7.5.1 TCGA-COAD 的 RNA-seq 差异表达	19

7.5.2	PPI . . . . .	22
7.5.3	富集分析 . . . . .	24
7.5.4	通过关联分析筛选负相关性互作蛋白 . . . . .	26
<b>Reference</b>		<b>28</b>

## List of Figures

1	PPS Cancer vs Control DEGs . . . . .	4
2	PPS Intersection of DPs Up with TransMemPs . . . . .	6
3	PPS DPS filtered and formated PPI network . . . . .	7
4	PPS PPI KEGG enrichment . . . . .	8
5	PPS PPI GO enrichment . . . . .	8
6	PPS PPI hsa04217 visualization . . . . .	9
7	PPS PPI hsa04612 visualization . . . . .	9
8	PPS DPS filtered by KEGG and formated PPI network . . . . .	10
9	PPS DPS filtered by KEGG Top MCC score . . . . .	11
10	QQ plot for distribution of the transformed IC50 data . . . . .	14
11	Estimate prediction accuracy . . . . .	15
12	TCGA Resistance vs Non resistance DEPs . . . . .	17
13	TCGA Intersection of DPs Up with TransMemPs . . . . .	19
14	TCGA RNA Resistance vs Non resistance DEGs . . . . .	20
15	TCGA RNA DEGs type . . . . .	21
16	TCGA raw PPI network . . . . .	22
17	TCGA DPS filtered and formated PPI network logFC . . . . .	23
18	TCGA DPS filtered and formated PPI network MCC . . . . .	24
19	TCGA KEGG enrichment . . . . .	25
20	TCGA GO enrichment . . . . .	25
21	TCGA TSC1 in hsa04151 visualization . . . . .	26
22	TCGA RNA correlation heatmap . . . . .	27

## List of Tables

1	PUBLISHED ProteomicsProfShao2022 metadata used sample . . . . .	3
2	PPS data Cancer vs Control DPs . . . . .	4
3	UniTmp data of htp all . . . . .	5
4	TCGA COAD clinical metadata . . . . .	12
5	TCGA COAD protein metadata . . . . .	12
6	TCGA COAD RNA metadata . . . . .	13
7	Predicted drug sensitivity . . . . .	15
8	TCGA COAD proteome metadata . . . . .	16
9	TCGA data Resistance vs Non resistance DEPs . . . . .	18

10	TCGA RNA data Resistance vs Non resistance DEGs . . . . .	20
11	TCGA RNA TSC1 negtive correlated . . . . .	27

# 1 摘要

- 研究对象：乳腺癌或结直肠癌
- 耐药：5-氟尿嘧啶或顺铂

## 1.1 需求

1、寻找差异致瘤膜蛋白及对应配体蛋白；2、耐药差异膜蛋白及对应抗体蛋白或互作抑制其表达蛋白。

现需要利用数据库分析正常组与疾病组间的差异表达膜蛋白 AA（在癌中高表达的）和对应靶向癌细胞特异性高表达的膜蛋白 AA'；以及非耐药组与耐药组间的差异表达膜蛋白 XX（在耐药组中高表达的）和对应 XX 的抗体蛋白或相互作用能抑制其表达的蛋白 XX'。

## 1.2 结果

### 1.2.1 寻找差异致瘤膜蛋白及对应配体蛋白

这一部分思路较简单，找到可用的蛋白质组数据<sup>1</sup>，筛选差异蛋白，再以 UniTmp 数据库筛选跨膜蛋白（膜蛋白受体主要为跨膜蛋白），再借助 STRINGdb 构建 PPI 网络，寻找互作蛋白，再结合富集分析 (6.3.3)，进一步缩小范围。

注意，以上 PPI 构建的来源是所有筛选到的差异蛋白 (拟从差异蛋白中找到候选的配体蛋白)。

差异瘤膜蛋白为：AIFM1, TFRC, ITGAM, PECAM1 最终筛选的蛋白和对应候选配体关系见：Fig. 8

### 1.2.2 耐药差异膜蛋白及对应抗体蛋白或互作抑制其表达蛋白

这部分思路稍复杂。由于无法直接获得包含耐药性分组的蛋白表达数据，因此需要另寻思路，即，获取 TCGA-COAD 的基因表达数据和蛋白定量数据，以 pRRophetic 根据基因表达数据分析耐药性 (顺铂 Cisplatin)，再对样本分组，随后分析蛋白定量数据。

后续和上一部分近似：筛选差异蛋白，再以 UniTmp 数据库筛选跨膜蛋白，再借助 STRINGdb 构建 PPI 网络，寻找互作蛋白，再结合富集分析。到这里，筛选蛋白 (TSC1) 和互作蛋白关系见 Fig. 17。后续的富集分析结果可能有一定参考价值，富集到 TSC1 (hsa04151, Fig. 21)

注意，以上 PPI 构建的来源是筛选的膜蛋白 (TSC1)，和 RNA-seq 的 DEGs (拟从差异蛋白中找到候选的配体蛋白)。

但这部分还需要指定互作抑制的蛋白，因此又结合了关联分析，挖掘 RNA 表达中呈负相关性的蛋白 (Fig. 22)。

最终可参考的表格：Tab. 11

## 2 前言

## 3 材料和方法

### 3.1 材料

Other data obtained from published article (e.g., supplementary tables):

- Supplementary file from article refer to<sup>1</sup>.

### 3.2 方法

Mainly used method:

- R package `ClusterProfiler` used for gene enrichment analysis<sup>2</sup>.
- R package `Limma` and `edgeR` used for differential expression analysis<sup>3,4</sup>.
- R Package `pRRophetic` was used for Prediction of Clinical Chemotherapeutic Response<sup>5</sup>.
- R package `STEINGdb` used for PPI network construction<sup>6,7</sup>.
- R package `TCGAbiolinks` used for abtain TCGA dataset<sup>8</sup>.
- The UNIfied database of TransMembrane Proteins (UniTmp) was used for transmembrane protein information retrieving<sup>9,10</sup>.
- The MCC score was calculated referring to algorithm of `CytoHubba`<sup>7</sup>.
- R version 4.3.2 (2023-10-31); Other R packages (eg., `dplyr` and `ggplot2`) used for statistic analysis or data visualization.

## 4 分析结果

## 5 结论

## 6 附：分析流程——寻找差异数致癌膜蛋白及对应配体蛋白

### 6.1 结直肠癌差异蛋白

#### 6.1.1 数据来源

Proteomics profiling of colorectal cancer progression identifies PLOD2 as a potential therapeutic target<sup>1</sup>

Table 1 (下方表格) 为表格 PUBLISHED\_ProteomicsProfShao2022 metadata used sample 概览。

(对应文件为 `Figure+Table/PUBLISHED-ProteomicsProfShao2022-metadata-used-sample.csv`)

注：表格共有 31 行 5 列，以下预览的表格可能省略部分数据；表格含有 2 个唯一‘group’。

1. sample: 样品名称
2. group: 分组名称

Table 1: PUBLISHED ProteomicsProfShao2022 metadata used sample

sample	Gender	Age	Pathology.Type	group
P1	Female	74	Normal Colon	Control
P2	Female	49	Normal Colon	Control
P5	Male	51	Normal Colon	Control
P6	Female	56	Normal Colon	Control
P7	Male	53	Normal Colon	Control
P8	Male	70	Normal Colon	Control
P9	Male	62	Normal Colon	Control
P11	Male	48	Normal Colon	Control
P13	Female	43	Normal Colon	Control
P14	Female	61	Normal Colon	Control
P15	Female	81	Normal Colon	Control
P16	Male	67	Normal Colon	Control
P17	Male	60	Normal Colon	Control
P18	Female	59	Normal Colon	Control
P19	Male	64	Normal Colon	Control
...	...	...	...	...

### 6.1.2 差异蛋白

Figure 1 (下方图) 为图 PPS Cancer vs Control DEGs 概览。

(对应文件为 [Figure+Table/PPS-Cancer-vs-Control-DEGs.pdf](#))

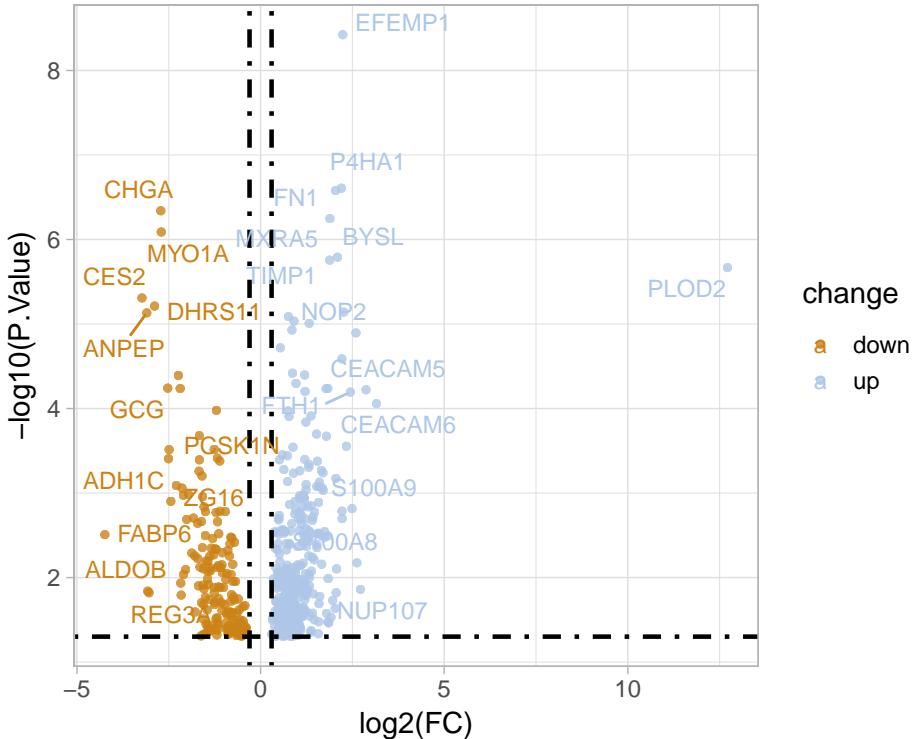


Figure 1: PPS Cancer vs Control DEGs

Table 2 (下方表格) 为表格 PPS data Cancer vs Control DPs 概览。

(对应文件为 [Figure+Table/PPS-data-Cancer-vs-Control-DPs.csv](#))

注：表格共有 509 行 8 列，以下预览的表格可能省略部分数据；表格含有 509 个唯一 ‘Gene\_name’。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
2. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
3. t: moderated t-statistic (omitted for ‘topTableF’ )
4. P.Value: raw p-value
5. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’ )

Table 2: PPS data Cancer vs Control DPs

Gene_name	logFC	rownames	AveExpr	t	P.Value	adj.P.Val	B
EFEMP1	2.2381...	Q12805	5.4470...	8.2464...	3.7745...	1.8231...	10.666...
P4HA1	2.2015...	P13674	4.8852...	6.6539...	2.4728...	0.0004...	6.8779...
FN1	2.0416...	P02751	10.551...	6.6301...	2.6380...	0.0004...	6.8187...
CHGA	-2.716...	P10645	6.4443...	-6.428...	4.5698...	0.0005...	6.3149...

Gene_name	logFC	rownames	AveExpr	t	P.Value	adj.P.Val	B
MXRA5	1.8876...	Q9NR99	5.3248...	6.3504...	5.6622...	0.0005...	6.1181...
MYO1A	-2.702...	Q9UBC5	6.3290...	-6.217...	8.1508...	0.0006...	5.7832...
BYSL	2.0924...	Q13895	3.1862...	5.9689...	1.6215...	0.0010...	5.1497...
TIMP1	1.8810...	P01033	5.1452...	5.9401...	1.7564...	0.0010...	5.0760...
PLOD2	12.715...	O00469	-5.707...	5.8680...	2.1455...	0.0011...	4.8915...
CES2	-3.227...	O00748	6.7662...	-5.570...	4.9174...	0.0023...	4.1256...
DHRS11	-2.886...	Q6UWP2	5.2052...	-5.491...	6.1300...	0.0026...	3.9219...
AEBP1	2.2654...	Q8IUX7	5.3752...	5.4328...	7.2234...	0.0027...	3.7701...
ANPEP	-3.097...	P15144	5.2534...	-5.425...	7.3809...	0.0027...	3.7501...
TKT	0.7599...	P29401	8.6760...	5.3887...	8.1722...	0.0028...	3.6559...
PTMS	0.9114...	P20962	8.2150...	5.3443...	9.2531...	0.0029...	3.5410...
...	...	...	...	...	...	...	...

## 6.2 膜蛋白筛选

受体蛋白主要分为：

- 离子通道受体 (Ligand-gated ion channel, LICs, LGIC)
- 催化受体 (酶受体) (catalytic receptor)
  - 鸟苷酸酰化酶受体
  - 酪氨酸激酶受体
- G 蛋白偶联受体 (G protein-coupled receptors) (GPCRs) (<https://gpcrdb.org/>)

以上都是跨膜蛋白类型。因此以下筛选将从跨膜蛋白出发。

### 6.2.1 UniTmp

UniTmp: unified resources for transmembrane proteins<sup>9</sup>

Table 3 (下方表格) 为表格 UniTmp data of htp all 概览。

(对应文件为 [Figure+Table/UniTmp-data-of-htp-all.xlsx](#))

注：表格共有 5499 行 5 列，以下预览的表格可能省略部分数据；表格含有 5499 个唯一‘id’。

1. evidence: 证据，相关文献中的描述。

Table 3: UniTmp data of htp all

id	transmembrane	evidence	Protein_name	Gene_Name
ACD10_HUMAN	yes	Exists	Acyl-CoA dehydrogenase	ACAD10
ASPH_HUMAN	yes	3D	Aspartyl/asparaginyl	ASPH

id	transmembrane	evidence	Protein_name	Gene_Name
ATP8_HUMAN	yes	3D	ATP synthase prot...	MT-ATP8
BFAR_HUMAN	yes	Exists	Bifunctional apop...	BFAR
BAMBI_HUMAN	yes	Exists	BMP and activin m...	BAMBI
ATRAP_HUMAN	yes	Exists	Type-1 angiotensi...	AGTRAP
AOFA_HUMAN	yes	3D	Amine oxidase [fl...	MAOA
BAP29_HUMAN	yes	3D	B-cell receptor-a...	BCAP29
BAP31_HUMAN	yes	3D	B-cell receptor-a...	BCAP31
C144C_HUMAN	yes	Prediction	Putative coiled-c...	CCDC144CP
CJ105_HUMAN	yes	Prediction	Uncharacterized p...	C10orf105
CJ111_HUMAN	yes	Prediction	Putative uncharac...	RPP38-DT
CLM2_HUMAN	yes	Exists	CMRF35-like molec...	CD300E
CK024_HUMAN	yes	Experiment	Uncharacterized p...	C11orf24
CK087_HUMAN	yes	Exists	Uncharacterized p...	C11orf87
...	...	...	...	...

### 6.2.2 与高表达差异蛋白 (DPs-Up) 交集

Figure 2 (下方图) 为图 PPS Intersection of DPs Up with TransMemPs 概览。

(对应文件为 [Figure+Table/PPS-Intersection-of-DPs-Up-with-TransMemPs.pdf](#))

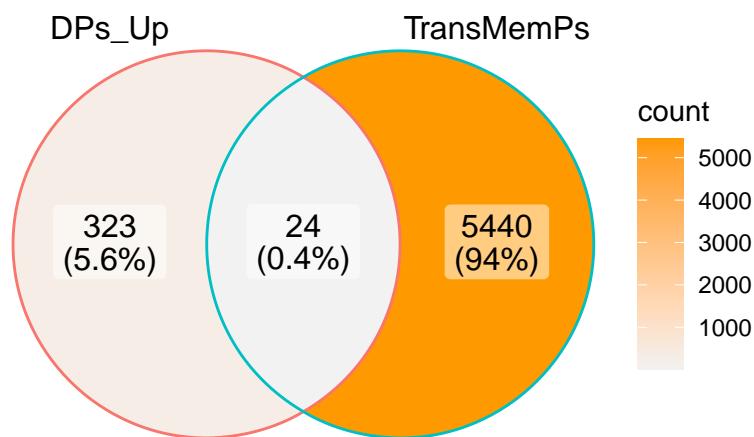


Figure 2: PPS Intersection of DPs Up with TransMemPs

#### Intersection :

CEACAM5, APP, LAMP1, MUC1, SFXN3, THY1, SLC25A6, TACSTD2, MTX1, AIFM1, NSDHL, PECAM1, HSD17B12, MRC2, ASPH, LEMD2, SSR3, LMO7, ITGAM, TFRC, SPINT2, SORT1, ACSL3, SFXN1

(上述信息框内容已保存至 Figure+Table/PPS-Intersection-of-DPs-Up-with-TransMemPs-content)

## 6.3 以蛋白互作筛选配体蛋白

### 6.3.1 PPI

Figure 3 (下方图) 为图 PPS DPS filtered and formated PPI network 概览。

(对应文件为 Figure+Table/PPS-DPS-filtered-and-formated-PPI-network.pdf)

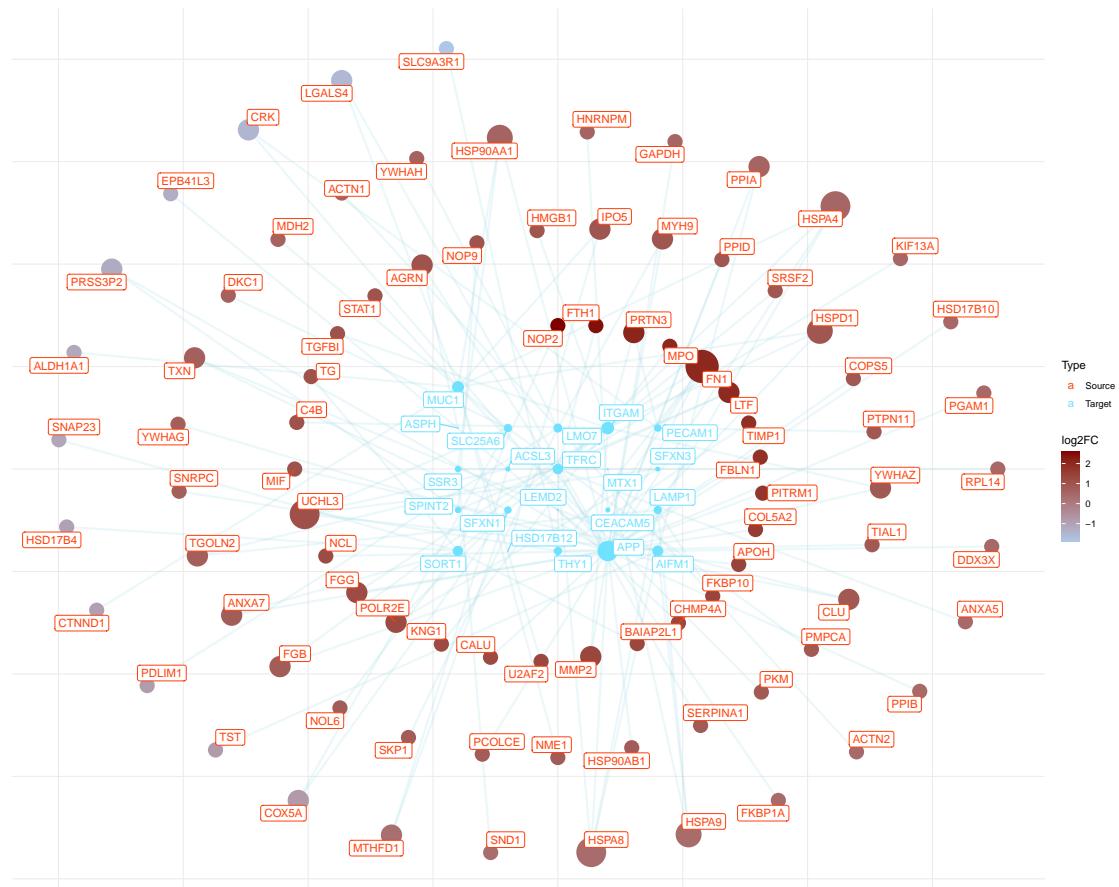


Figure 3: PPS DPS filtered and formated PPI network

### 6.3.2 富集分析

Figure 4 (下方图) 为图 PPS PPI KEGG enrichment 概览。

(对应文件为 Figure+Table/PPS-PPI-KEGG-enrichment.pdf)

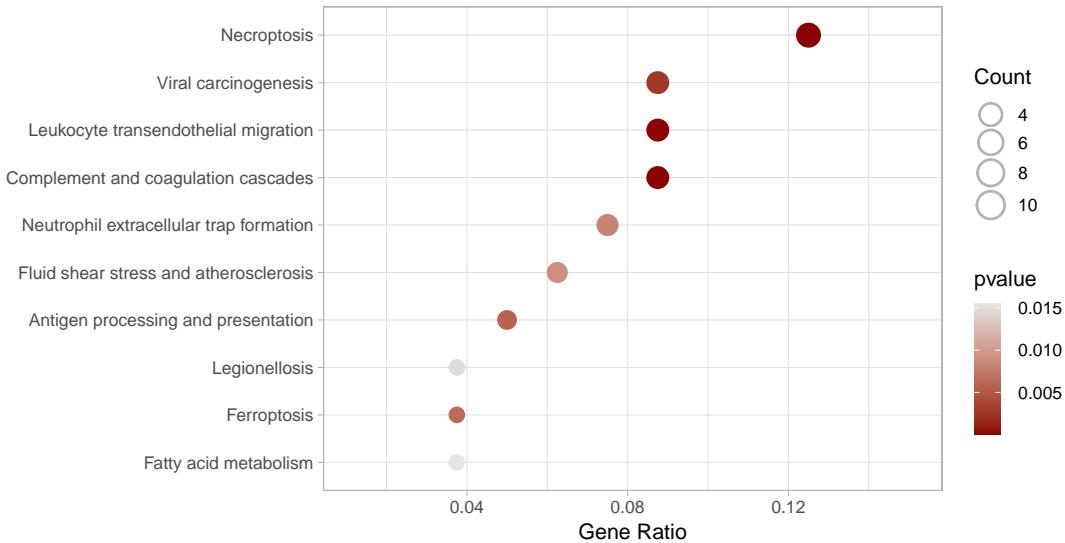
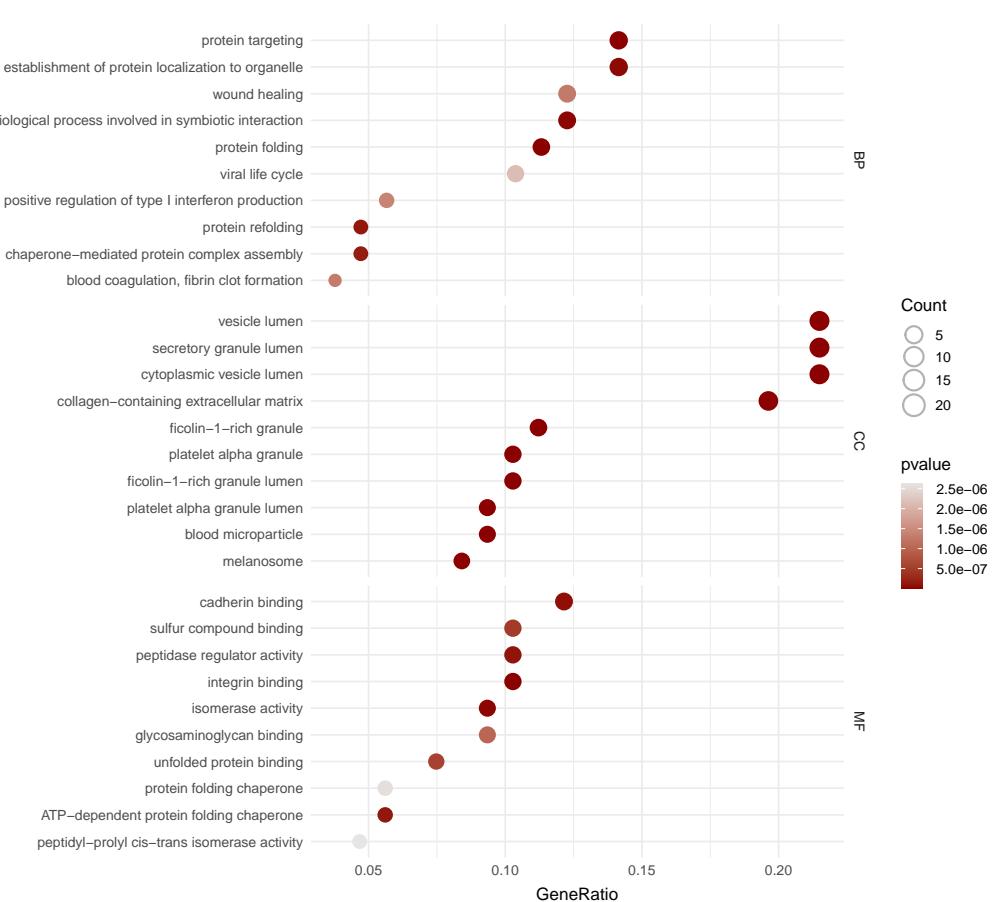


Figure 5 (下方图) 为图 PPS PPI GO enrichment 概览。

(对应文件为 Figure+Table/PPS-PPI-GO-enrichment.pdf)



坏死性凋亡信号通路在肿瘤发生发展、肿瘤坏死、肿瘤转移和肿瘤免疫反应中发挥作用；坏死性凋亡可能促进或抗肿瘤发生，具体取决于肿瘤的类型<sup>11</sup>。

Figure 6 (下方图) 为图 PPS PPI hsa04217 visualization 概览。

(对应文件为 Figure+Table/hsa04217.pathview.png)

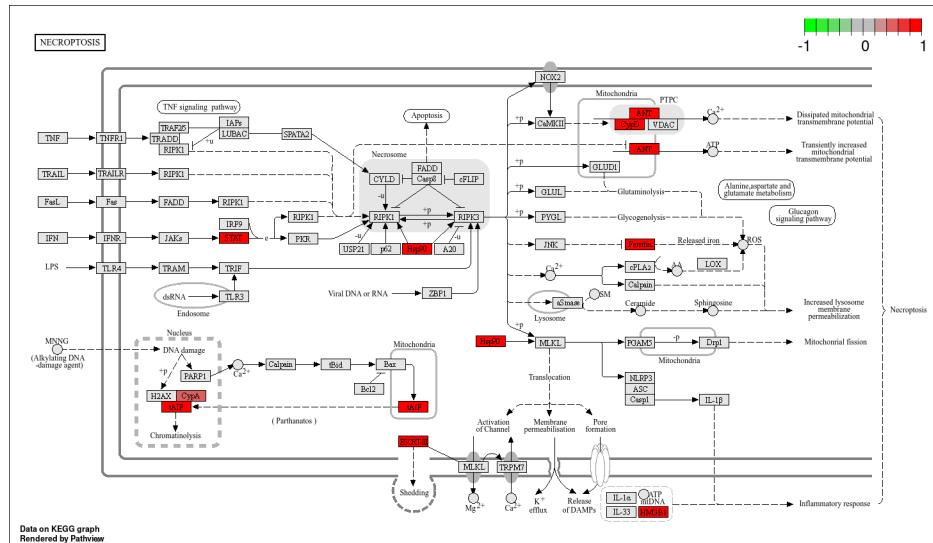


Figure 6: PPS PPI hsa04217 visualization

70-kDa 热休克蛋白 (HSP70) 在癌症中大量存在，通过抑制多种凋亡途径、调节坏死、绕过细胞衰老程序、干扰肿瘤免疫、促进血管生成和支持转移，为恶性细胞提供选择优势<sup>12</sup>

Figure 7 (下方图) 为图 PPS PPI hsa04612 visualization 概览。

(对应文件为 Figure+Table/hsa04612.pathview.png)

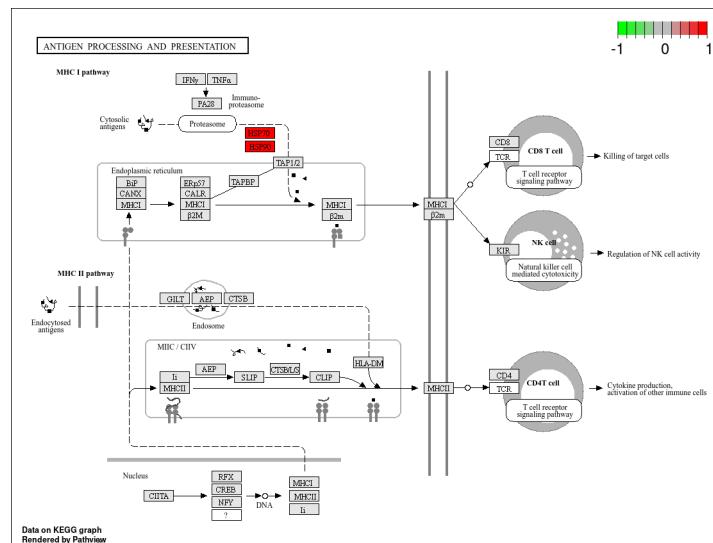


Figure 7: PPS PPI hsa04612 visualization

### 6.3.3 根据富集结果缩减 PPI

由于富集结果可以凸显肿瘤的性质，这里尝试根据 KEGG top 10 通路的富集基因缩减 PPI

Figure 8 (下方图) 为图 PPS DPS filtered by KEGG and formated PPI network 概览。

(对应文件为 Figure+Table/PPS-DPS-filtered-by-KEGG-and-formated-PPI-network.pdf)

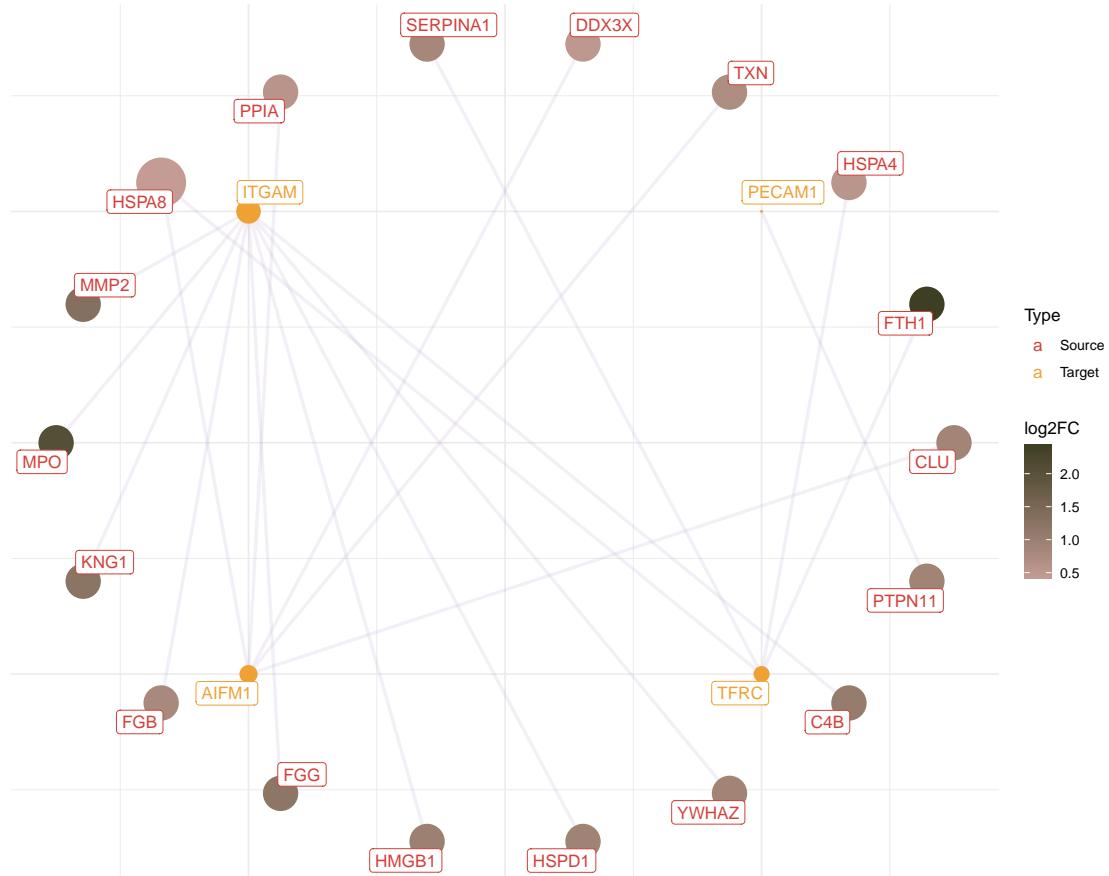


Figure 8: PPS DPS filtered by KEGG and formated PPI network

Figure 9 (下方图) 为图 PPS DPS filtered by KEGG Top MCC score 概览。

(对应文件为 Figure+Table/PPS-DPS-filtered-by-KEGG-Top-MCC-score.pdf)

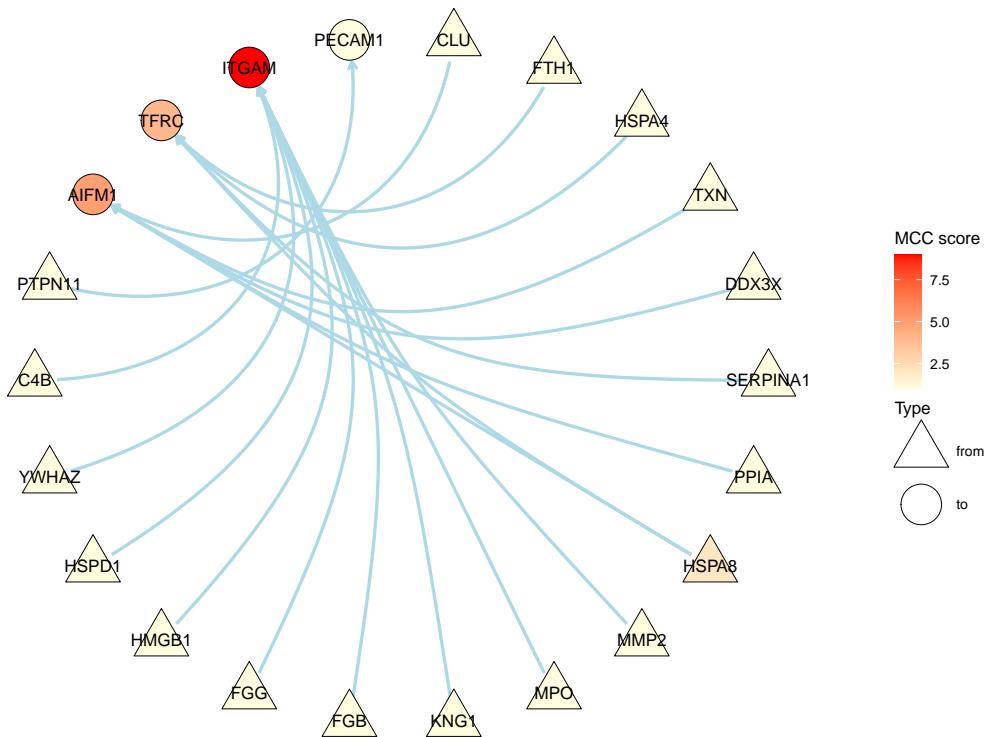


Figure 9: PPS DPS filtered by KEGG Top MCC score

## 7 附：分析流程——耐药差异膜蛋白及对应抗体蛋白或互作抑制其表达蛋白

### 7.1 结肠癌差异蛋白

注：由于无法直接获得包含耐药性分组的蛋白表达数据，因此这部分的内容另寻思路，即，获取 TCGA-COAD 的基因表达数据和蛋白定量数据，以 pRRophetic 根据基因表达数据分析耐药性，再对样本分组，随后分析蛋白定量数据。

#### 7.1.1 数据来源

共使用了 TCGA-COAD 的 RNA, protein, clinical 数据 (使用了三者都包含的病人的样本数据)。

Table 4 (下方表格) 为表格 TCGA COAD clinical metadata 概览。

(对应文件为 Figure+Table/TCGA-COAD-clinical-metadata.csv)

注：表格共有 459 行 19 列，以下预览的表格可能省略部分数据；表格含有 459 个唯一‘rownames’。

Table 4: TCGA COAD clinical metadata

rownames	id	data_f...	cases	access	file_name	submit...	data_c...	type	file_size
1	080f87...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	29118
2	77638e...	BCR XML	TCGA-C...	open	nation...	nation...	Clinical	clinic...	54636
3	c58a3b...	BCR XML	TCGA-D...	open	nation...	nation...	Clinical	clinic...	29261
4	e67964...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	28673
5	568223...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	31577
6	d77499...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	31559
7	0efd65...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	29185
8	7aaa95...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	52563
9	75bb5e...	BCR XML	TCGA-G...	open	nation...	nation...	Clinical	clinic...	54330
10	5d3d27...	BCR XML	TCGA-D...	open	nation...	nation...	Clinical	clinic...	29453
11	0f0694...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	24277
12	d05b7b...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	34479
13	ef4172...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	24264
14	f64700...	BCR XML	TCGA-G...	open	nation...	nation...	Clinical	clinic...	29321
15	849b28...	BCR XML	TCGA-A...	open	nation...	nation...	Clinical	clinic...	24215
...	...	...	...	...	...	...	...	...	...

Table 5 (下方表格) 为表格 TCGA COAD protein metadata 概览。

(对应文件为 Figure+Table/TCGA-COAD-protein-metadata.csv)

注：表格共有 363 行 24 列，以下预览的表格可能省略部分数据；表格含有 363 个唯一 ‘id’。

Table 5: TCGA COAD protein metadata

id	data_f...	cases	access	file_name	submit...	data_c...	type	platform	file_size
7b5dc5...	TSV	TCGA-C...	open	TCGA-C...	TCGA-C...	Proteo...	protei...	RPPA	22135
7d66e7...	TSV	TCGA-S...	open	TCGA-S...	TCGA-S...	Proteo...	protei...	RPPA	22022
8ecf75...	TSV	TCGA-A...	open	TCGA-A...	TCGA-A...	Proteo...	protei...	RPPA	24055
5e4ec1...	TSV	TCGA-A...	open	TCGA-A...	TCGA-A...	Proteo...	protei...	RPPA	23999
e45a96...	TSV	TCGA-A...	open	TCGA-A...	TCGA-A...	Proteo...	protei...	RPPA	24027
47c932...	TSV	TCGA-C...	open	TCGA-C...	TCGA-C...	Proteo...	protei...	RPPA	22100
44f0f7...	TSV	TCGA-A...	open	TCGA-A...	TCGA-A...	Proteo...	protei...	RPPA	24049
9fe863...	TSV	TCGA-C...	open	TCGA-C...	TCGA-C...	Proteo...	protei...	RPPA	22071
e06fe7...	TSV	TCGA-A...	open	TCGA-A...	TCGA-A...	Proteo...	protei...	RPPA	24065
cf9c71...	TSV	TCGA-Q...	open	TCGA-Q...	TCGA-Q...	Proteo...	protei...	RPPA	22099
335dae...	TSV	TCGA-G...	open	TCGA-G...	TCGA-G...	Proteo...	protei...	RPPA	22154
4f0d60...	TSV	TCGA-D...	open	TCGA-D...	TCGA-D...	Proteo...	protei...	RPPA	22155
f36bba...	TSV	TCGA-G...	open	TCGA-G...	TCGA-G...	Proteo...	protei...	RPPA	22109

id	data_f...	cases	access	file_name	submit...	data_c...	type	platform	file_size
f185df...	TSV	TCGA-A...	open	TCGA-A...	TCGA-A...	Proteo...	protei...	RPPA	24011
be6d80...	TSV	TCGA-F...	open	TCGA-F...	TCGA-F...	Proteo...	protei...	RPPA	22109
...	...	...	...	...	...	...	...	...	...

Table 6 (下方表格) 为表格 TCGA COAD RNA metadata 概览。

(对应文件为 `Figure+Table/TCGA-COAD-RNA-metadata.csv`)

注: 表格共有 524 行 29 列, 以下预览的表格可能省略部分数据; 表格含有 524 个唯一 ‘id’。

Table 6: TCGA COAD RNA metadata

id	data_f...	cases	access	file_name	submit...	data_c...	type	file_size	create...
efdcda...	TSV	TCGA-C...	open	a45a75...	9a7d8c...	Transc...	gene_e...	4240354	2022-0...
66905c...	TSV	TCGA-A...	open	40b98f...	7070f7...	Transc...	gene_e...	4233852	2022-0...
8820d9...	TSV	TCGA-A...	open	58369b...	a6da34...	Transc...	gene_e...	4210387	2022-0...
5302ca...	TSV	TCGA-C...	open	45a267...	c5db3d...	Transc...	gene_e...	4209007	2022-0...
a7affc...	TSV	TCGA-A...	open	3ef878...	97c2f0...	Transc...	gene_e...	4208716	2022-0...
3456d4...	TSV	TCGA-A...	open	524723...	4d8cb5...	Transc...	gene_e...	4216110	2022-0...
ab6d59...	TSV	TCGA-G...	open	085e42...	228266...	Transc...	gene_e...	4231843	2022-0...
3db2cc...	TSV	TCGA-A...	open	e1d0ac...	2b637e...	Transc...	gene_e...	4210406	2022-0...
a761bd...	TSV	TCGA-A...	open	d837c3...	738864...	Transc...	gene_e...	4201059	2022-0...
04981c...	TSV	TCGA-A...	open	4bafe8...	53e5cd...	Transc...	gene_e...	4214788	2022-0...
8ac00c...	TSV	TCGA-A...	open	0ae6d3...	0e343a...	Transc...	gene_e...	4219902	2022-0...
e929b2...	TSV	TCGA-C...	open	a82c20...	99e86d...	Transc...	gene_e...	4223537	2022-0...
c403bb...	TSV	TCGA-A...	open	e5ae4a...	8e4b41...	Transc...	gene_e...	4211960	2022-0...
756b10...	TSV	TCGA-A...	open	b3e428...	fe34e7...	Transc...	gene_e...	4229447	2022-0...
700b39...	TSV	TCGA-A...	open	fd618c...	afa933...	Transc...	gene_e...	4229363	2022-0...
...	...	...	...	...	...	...	...	...	...

## 7.2 预测耐药性

### 7.2.1 使用 RNA 数据集预测

以 pRRophetic 预测药物敏感性 (Cisplatin)。

Figure 10 (下方图) 为图 QQ plot for distribution of the transformed IC50 data 概览。

(对应文件为 `Figure+Table/QQ-plot-for-distribution-of-the-transformed-IC50-data.pdf`)

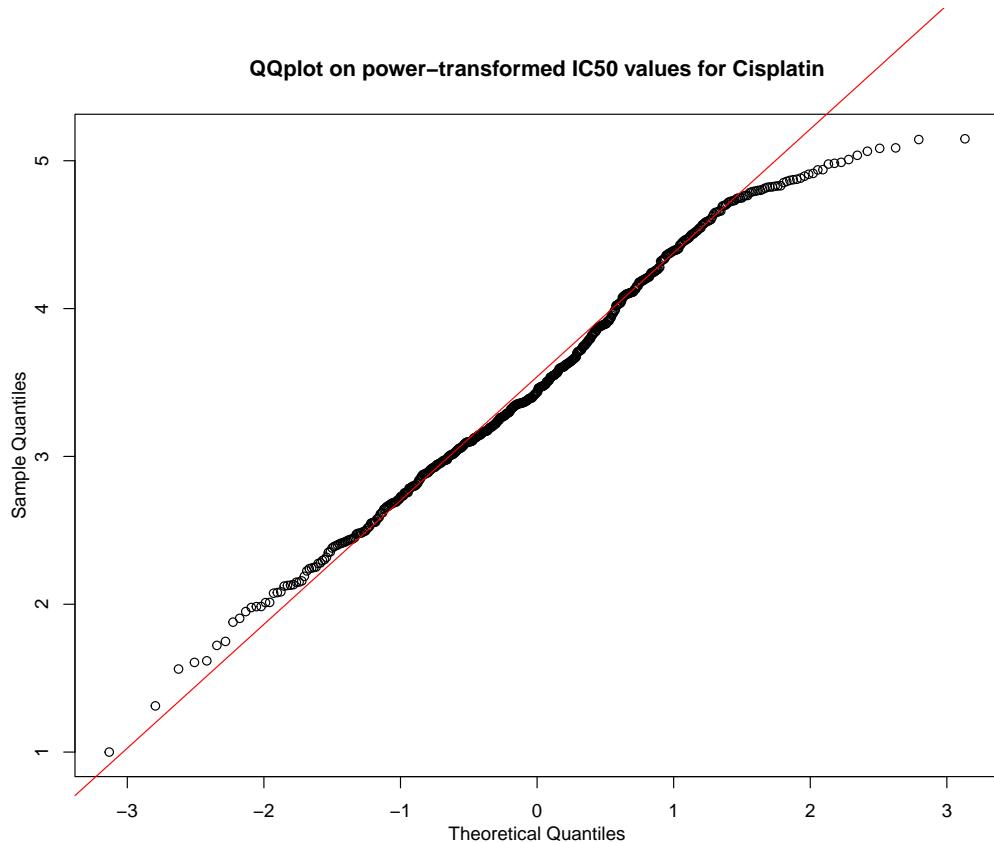


Figure 10: QQ plot for distribution of the transformed IC50 data

Fig. 10 表明, Cisplatin 的  $IC_{50}$  数据特征基本符合正太分布, 可以用于线形预测 Cisplatin 敏感性。

Figure 11 (下方图) 为图 Estimate prediction accuracy 概览。

(对应文件为 [Figure+Table/Estimate-prediction-accuracy.pdf](#))

**Measured phenotype Vs. C.V. predicted phenotype**

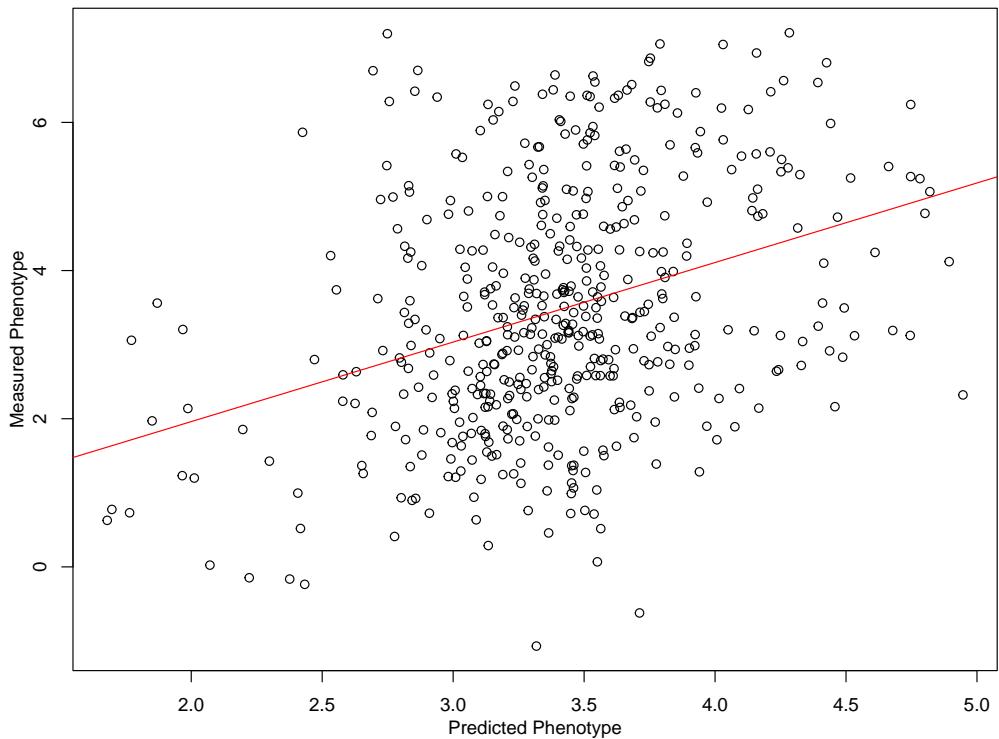


Figure 11: Estimate prediction accuracy

**k-means clustering :**

Centers = 3

Table 7 (下方表格) 为表格 Predicted drug sensitivity 概览。

(对应文件为 Figure+Table/Predicted-drug-sensitivity.csv)

注：表格共有 357 行 3 列，以下预览的表格可能省略部分数据；表格含有 357 个唯一‘sample’。

1. sample: 样品名称

Table 7: Predicted drug sensitivity

sample	sensitivity	kmeans_group
TCGA-3L-AA1B-01A	4.14816093273104	2
TCGA-4N-A93T-01A	4.40190686256072	2

sample	sensitivity	kmeans_group
TCGA-4T-AA8H-01A	4.15648471482818	2
TCGA-5M-AAT6-01A	2.62116649131769	3
TCGA-A6-2671-11A	4.03864919730421	2
TCGA-A6-2672-01A	2.76158089020952	3
TCGA-A6-2676-01A	2.10216420226229	3
TCGA-A6-2677-01A	3.02480062454453	1
TCGA-A6-2678-11A	3.85525962252971	2
TCGA-A6-2680-01A	4.07942930425798	2
TCGA-A6-2681-01A	3.95239422637806	2
TCGA-A6-2683-11A	3.64057626378351	1
TCGA-A6-2684-01C	2.93670056076102	3
TCGA-A6-2685-01A	4.32047341962156	2
TCGA-A6-2686-11A	3.87165483393886	2
...	...	...

## 7.3 差异蛋白筛选

### 7.3.1 元数据

根据 Tab. 7 k-means 聚类结果，将样品分为三组：耐药组、中等组、低耐药性组（非耐药组）。

**k-means clustering :**

Centers = 3

Table 8 (下方表格) 为表格 TCGA COAD proteome metadata 概览。

(对应文件为 **Figure+Table/TCGA-COAD-proteome-metadata.csv**)

注：表格共有 178 行 4 列，以下预览的表格可能省略部分数据；表格含有 2 个唯一 ‘group’。

1. sample: 样品名称
2. group: 分组名称

Table 8: TCGA COAD proteome metadata

sample	group	sensitivity	kmeans_group
TCGA-3L-AA1B	Resistance	4.14816093273104	2
TCGA-4N-A93T	Resistance	4.40190686256072	2
TCGA-4T-AA8H	Resistance	4.15648471482818	2
TCGA-5M-AAT6	Non_resistance	2.62116649131769	3

sample	group	sensitivity	kmeans_group
TCGA-A6-2671	Resistance	4.03864919730421	2
TCGA-A6-2672	Non_resistance	2.76158089020952	3
TCGA-A6-2676	Non_resistance	2.10216420226229	3
TCGA-A6-2678	Resistance	3.85525962252971	2
TCGA-A6-2680	Resistance	4.07942930425798	2
TCGA-A6-2681	Resistance	3.95239422637806	2
TCGA-A6-2684	Non_resistance	2.93670056076102	3
TCGA-A6-2685	Resistance	4.32047341962156	2
TCGA-A6-2686	Resistance	3.87165483393886	2
TCGA-A6-3808	Non_resistance	2.93134891159939	3
TCGA-A6-3809	Non_resistance	1.43580359549368	3
...	...	...	...

### 7.3.2 差异蛋白

Figure 12 (下方图) 为图 TCGA Resistance vs Non resistance DEPs 概览。

(对应文件为 Figure+Table/TCGA-Resistance-vs-Non-resistance-DEPs.pdf)

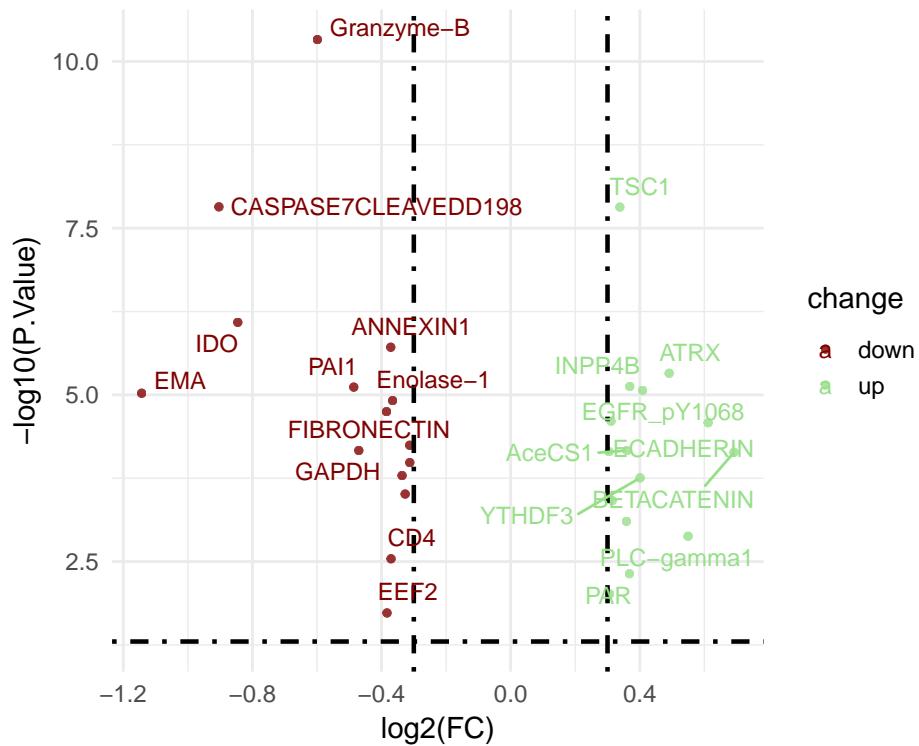


Figure 12: TCGA Resistance vs Non resistance DEPs

Table 9 (下方表格) 为表格 TCGA data Resistance vs Non resistance DEPs 概览。

(对应文件为 Figure+Table/TCGA-data-Resistance-vs-Non-resistance-DEPs.csv)

注：表格共有 30 行 9 列，以下预览的表格可能省略部分数据；表格含有 30 个唯一 ‘rownames’。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
2. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
3. t: moderated t-statistic (omitted for ‘topTableF’ )
4. P.Value: raw p-value
5. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’ )

Table 9: TCGA data Resistance vs Non resistance DEPs

rownames	logFC	AveExpr	t	P.Value	adj.P.Val	B	peptid...	Gene_name
AGID00319	-0.598...	0.0565...	-7.018...	4.6950...	2.1691...	14.789...	Granzy...	Granzy...
AGID00015	-0.903...	0.7403...	-5.932...	1.5134...	2.3504...	9.2486...	CASPAS...	CASPAS...
AGID00155	0.3382...	0.2485...	5.9309...	1.5262...	2.3504...	9.2406...	TSC1	TSC1
AGID00366	-0.845...	1.0622...	-5.114...	8.1696...	5.3919...	5.4545...	IDO	IDO
AGID00193	-0.371...	-0.473...	-4.922...	1.9302...	0.0001...	4.6336...	ANNEXIN1	ANNEXIN1
AGID00268	0.4910...	-0.646...	4.7228...	4.7425...	0.0002...	3.7904...	ATRX	ATRX
AGID02153	0.3689...	0.4742...	4.6151...	7.4803...	0.0003...	3.3539...	INPP4B	INPP4B
AGID00053	-0.486...	0.1941...	-4.609...	7.6492...	0.0003...	3.3328...	PAI1	PAI1
AGID00450	0.4087...	0.5620...	4.5837...	8.5582...	0.0003...	3.2270...	EGFR_p...	EGFR_p...
AGID00235	-1.143...	-0.300...	-4.562...	9.4475...	0.0003...	3.1408...	EMA	EMA
AGID00394	-0.365...	0.1766...	-4.503...	1.2172...	0.0003...	2.9023...	Enolase-1	Enolase-1
AGID00031	-0.384...	-0.126...	-4.410...	1.7819...	0.0004...	2.5371...	FIBRON...	FIBRON...
AGID00290	0.3109...	-0.549...	4.3319...	2.4809...	0.0006...	2.2336...	Lasu1	Lasu1
AGID00148	0.6118...	0.3484...	4.3171...	2.6150...	0.0006...	2.1771...	ECADHERIN	ECADHERIN
AGID00301	-0.312...	-0.132...	-4.125...	5.6927...	0.0011...	1.4564...	B7-H3	B7-H3
...	...	...	...	...	...	...	...	...

## 7.4 膜蛋白筛选

### 7.4.1 Unitmp

UniTmp: unified resources for transmembrane proteins<sup>9</sup>

见 Tab. 3

### 7.4.2 与高表达差异蛋白 (TCGA-dps-Up) 交集

Figure 13 (下方图) 为图 TCGA Intersection of DPs Up with TransMemPs 概览。

(对应文件为 Figure+Table/TCGA-Intersection-of-DPs-Up-with-TransMemPs.pdf)

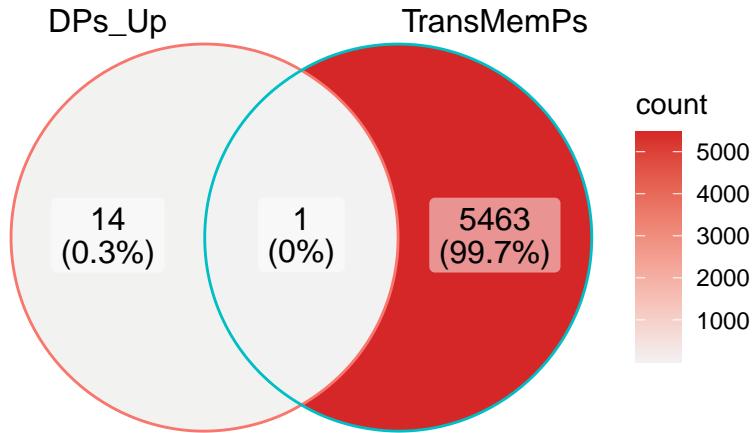


Figure 13: TCGA Intersection of DPs Up with TransMemPs

**Intersection :**

TSC1

(上述信息框内容已保存至 [Figure+Table/TCGA-Intersection-of-DPs-Up-with-TransMemPs-content](#))

## 7.5 以蛋白互作筛选配体蛋白

### 7.5.1 TCGA-COAD 的 RNA-seq 差异表达

为了筛选 Fig. 13 的配体，以 TCGA-COAD 的差异表达基因作为候选。

Figure 14 (下方图) 为图 TCGA RNA Resistance vs Non resistance DEGs 概览。

(对应文件为 [Figure+Table/TCGA-RNA-Resistance-vs-Non-resistance-DEGs.pdf](#))

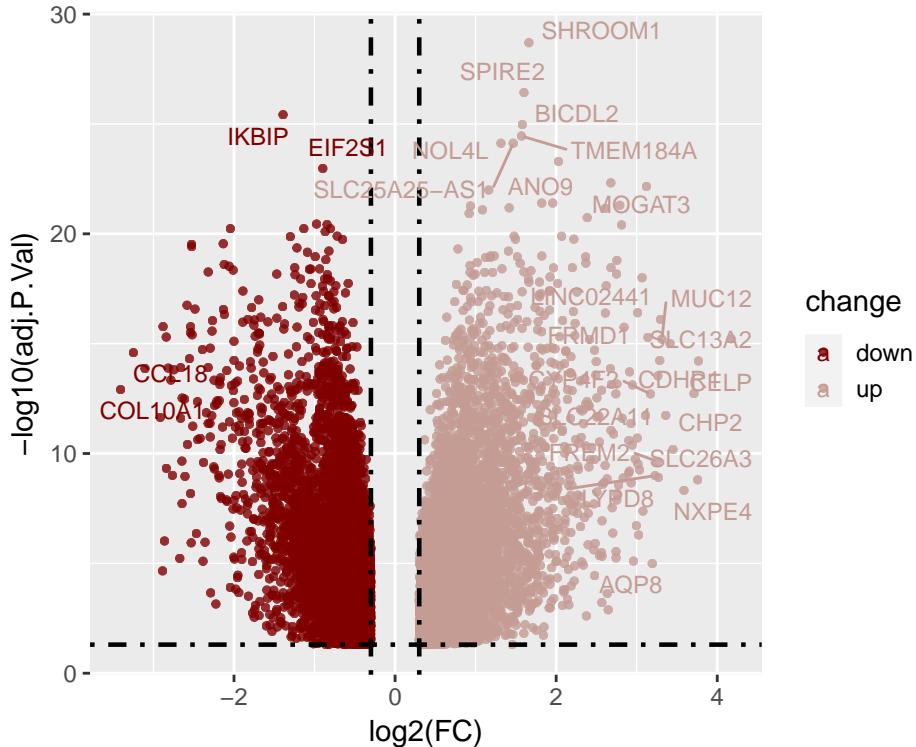


Figure 14: TCGA RNA Resistance vs Non resistance DEGs

Table 10 (下方表格) 为表格 TCGA RNA data Resistance vs Non resistance DEGs 概览。

(对应文件为 [Figure+Table/TCGA-RNA-data-Resistance-vs-Non-resistance-DEGs.csv](#))

注：表格共有 10108 行 22 列，以下预览的表格可能省略部分数据；表格含有 10108 个唯一 ‘rownames’。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
2. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
3. t: moderated t-statistic (omitted for ‘topTableF’ )
4. P.Value: raw p-value
5. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’ )
6. gene\_id: GENCODE/Ensembl gene ID
7. gene\_name: GENCODE gene name
8. strand: genomic strand

Table 10: TCGA RNA data Resistance vs Non resistance DEGs

rownames	gene_id	seqnames	start	end	width	strand	source	type	score
ENSG00...	ENSG00...	chr5	132822141	132830659	8519	-	HAVANA	gene	NA

rownames	gene_id	seqnames	start	end	width	strand	source	type	score
ENSG00...	ENSG00...	chr16	89818179	89871319	53141	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr12	98613405	98645113	31709	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr16	3027682	3036944	9263	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr7	1542235	1560821	18587	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr9	128108581	128118693	10113	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr20	32443059	32585074	142016	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr11	417933	442011	24079	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr14	67360328	67386516	26189	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr7	101195007	101201038	6032	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr12	56230049	56237846	7798	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr3	48403854	48430086	26233	-	HAVANA	gene	NA
ENSG00...	ENSG00...	chr3	42489299	42537573	48275	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr17	79074824	79088599	13776	+	HAVANA	gene	NA
ENSG00...	ENSG00...	chr2	219054424	219060921	6498	-	HAVANA	gene	NA
...	...	...	...	...	...	...	...	...	...

Figure 15 (下方图) 为图 TCGA RNA DEGs type 概览。

(对应文件为 Figure+Table/TCGA-RNA-DEGs-type.pdf)

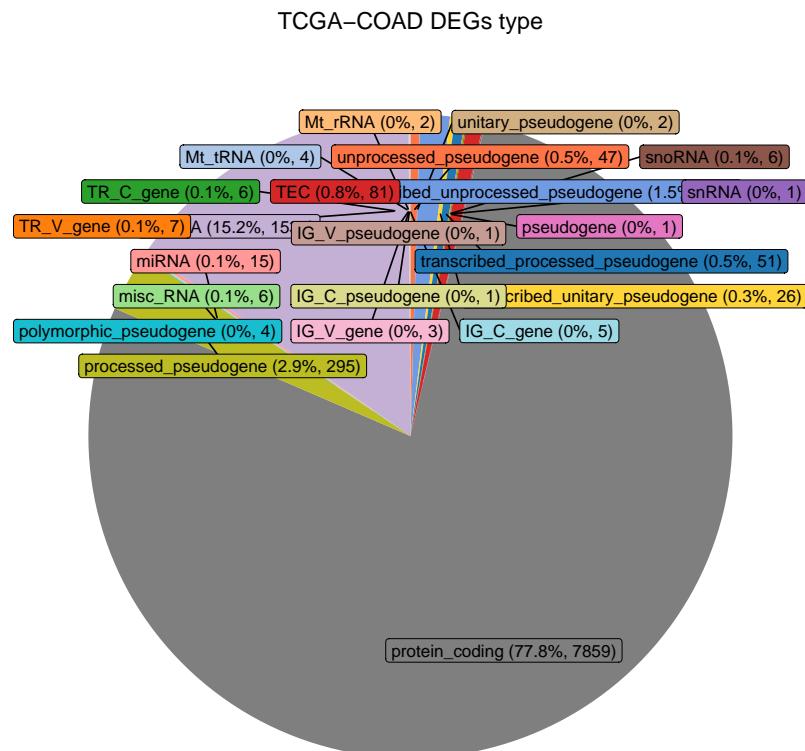


Figure 15: TCGA RNA DEGs type

### 7.5.2 PPI

以 Fig. 13 和 Tab. 10 top 2000 构建 PPI 网络。

Figure 16 (下方图) 为图 TCGA raw PPI network 概览。

(对应文件为 Figure+Table/TCGA-raw-PPI-network.pdf)

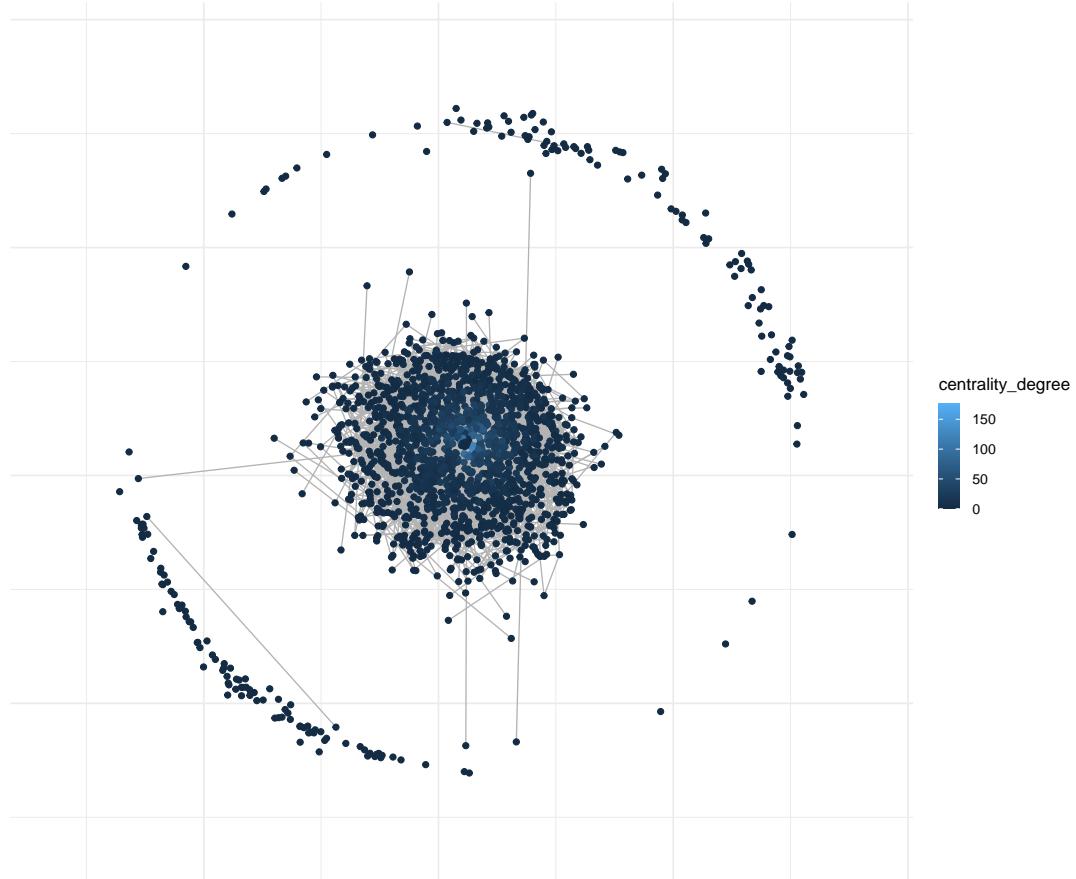


Figure 16: TCGA raw PPI network

将 PPI 网络过滤，凸显 Fig. 13 交集蛋白 (MCC 筛选高分相连的其他蛋白)

Figure 17 (下方图) 为图 TCGA DPS filtered and formated PPI network logFC 概览。

(对应文件为 Figure+Table/TCGA-DPS-filtered-and-formated-PPI-network-logFC.pdf)

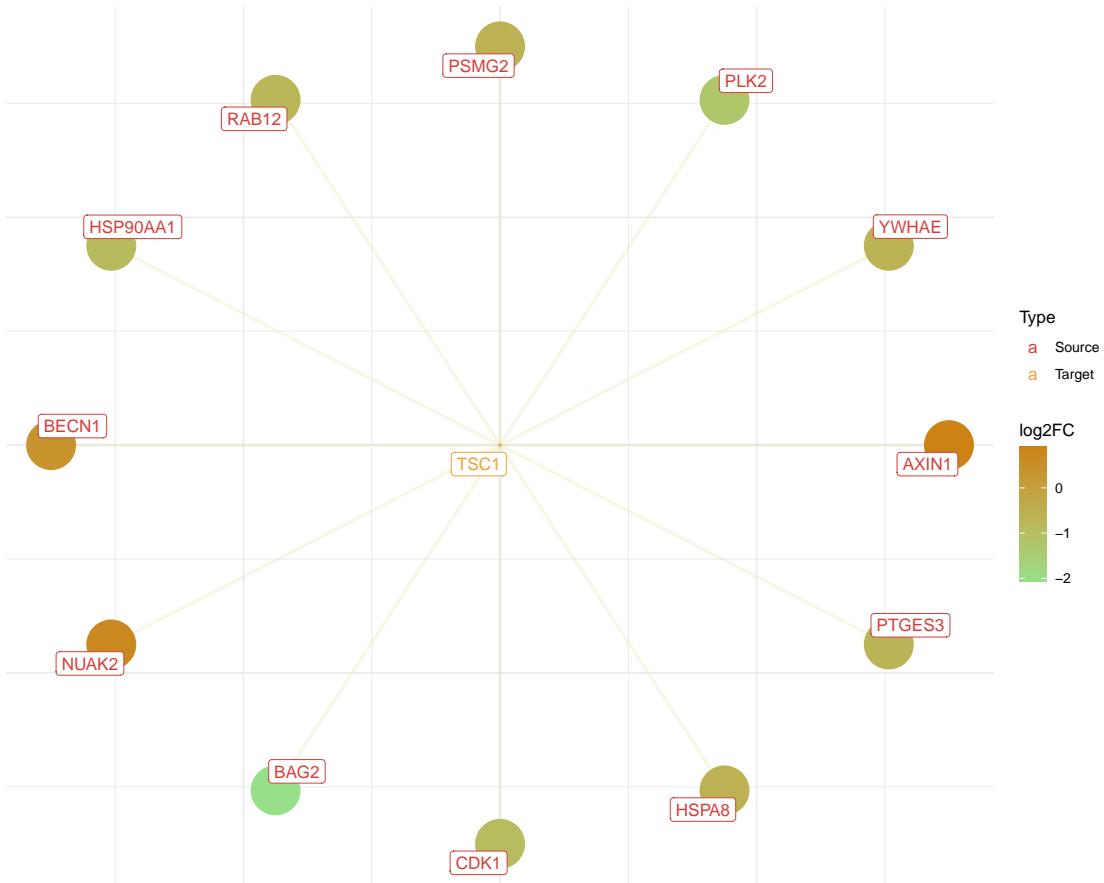


Figure 17: TCGA DPS filtered and formated PPI network logFC

Figure 18 (下方图) 为图 TCGA DPS filtered and formated PPI network MCC 概览。

(对应文件为 [Figure+Table/TCGA-DPS-filtered-and-formated-PPI-network-MCC.pdf](#))

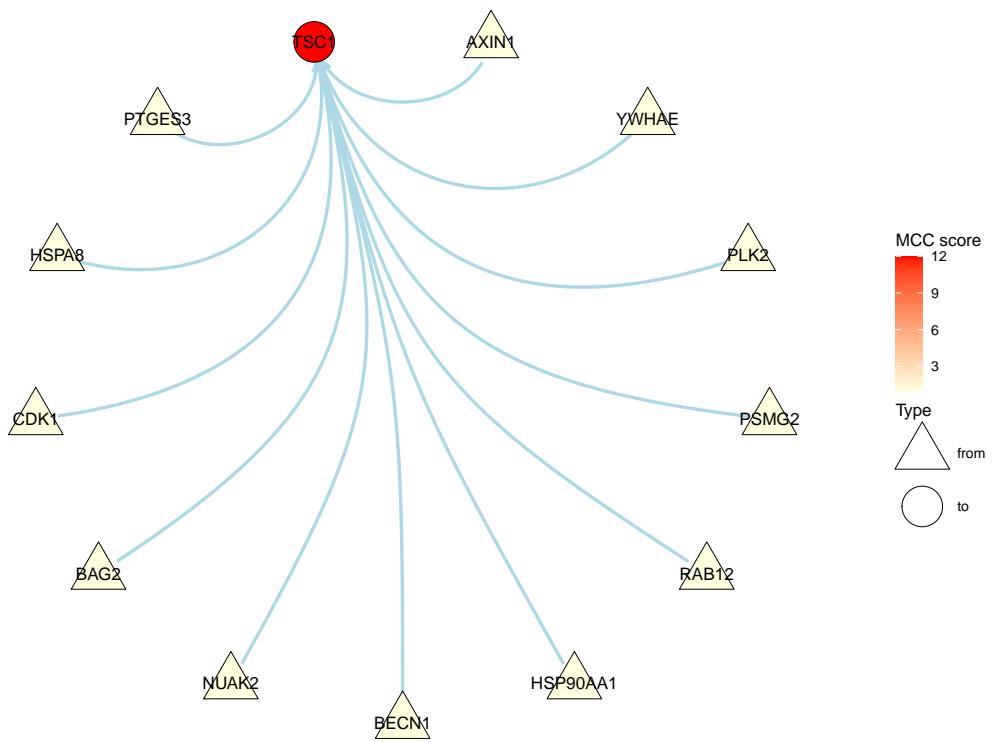


Figure 18: TCGA DPS filtered and formated PPI network MCC

### 7.5.3 富集分析

TSC1 在通路可见 Fig. 21

Figure 19 (下方图) 为图 TCGA KEGG enrichment 概览。

(对应文件为 Figure+Table/TCGA-KEGG-enrichment.pdf)

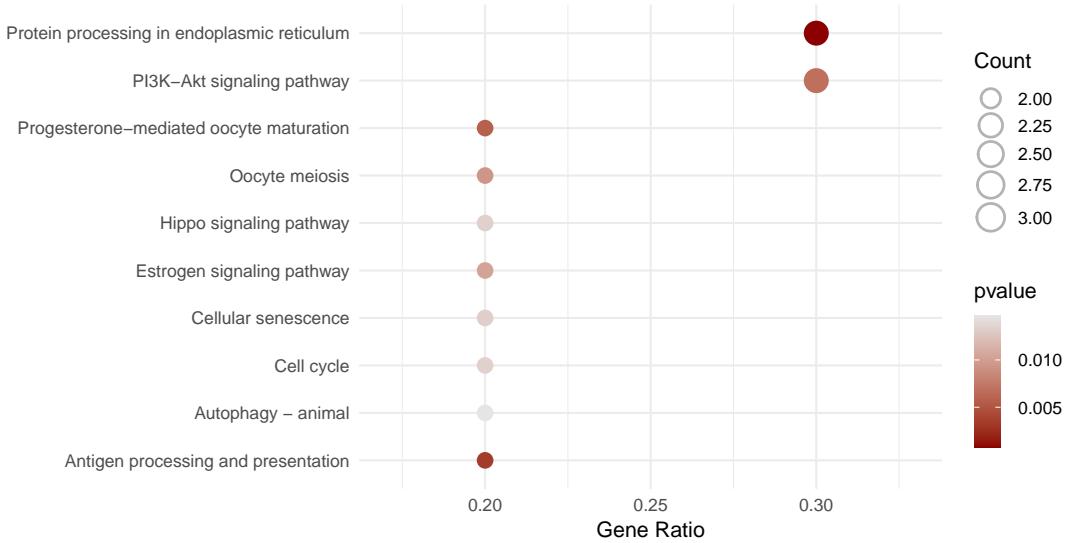


Figure 19: TCGA KEGG enrichment

Figure 20 (下方图) 为图 TCGA GO enrichment 概览。

(对应文件为 Figure+Table/TCGA-GO-enrichment.pdf)

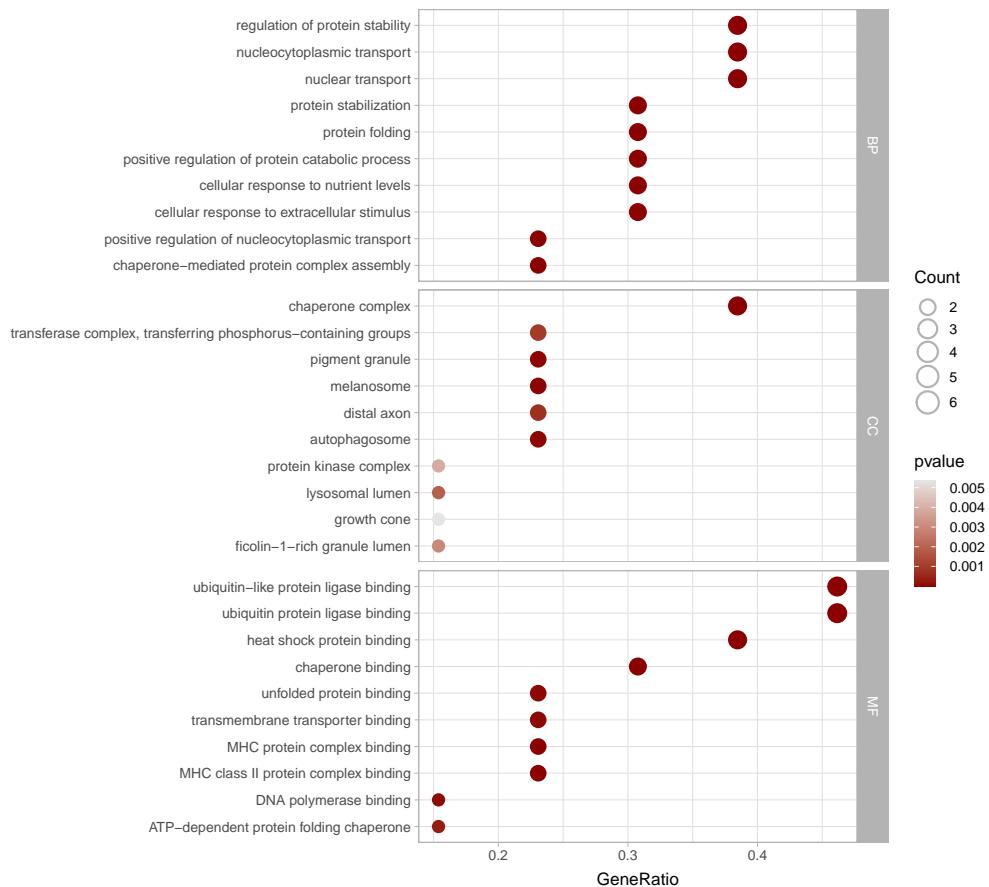


Figure 21 (下方图) 为图 TCGA TSC1 in hsa04151 visualization 概览。

(对应文件为 Figure+Table/hsa04151.pathview.png)

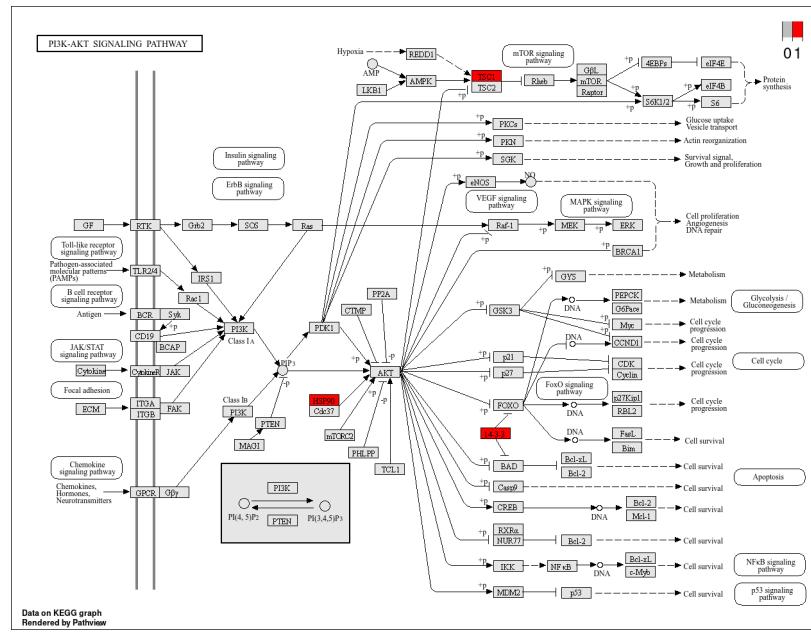


Figure 21: TCGA TSC1 in hsa04151 visualization

#### 7.5.4 通过关联分析筛选负相关性互作蛋白

Figure 22 (下方图) 为图 TCGA RNA correlation heatmap 概览。

(对应文件为 Figure+Table/TCGA-RNA-correlation-heatmap.pdf)

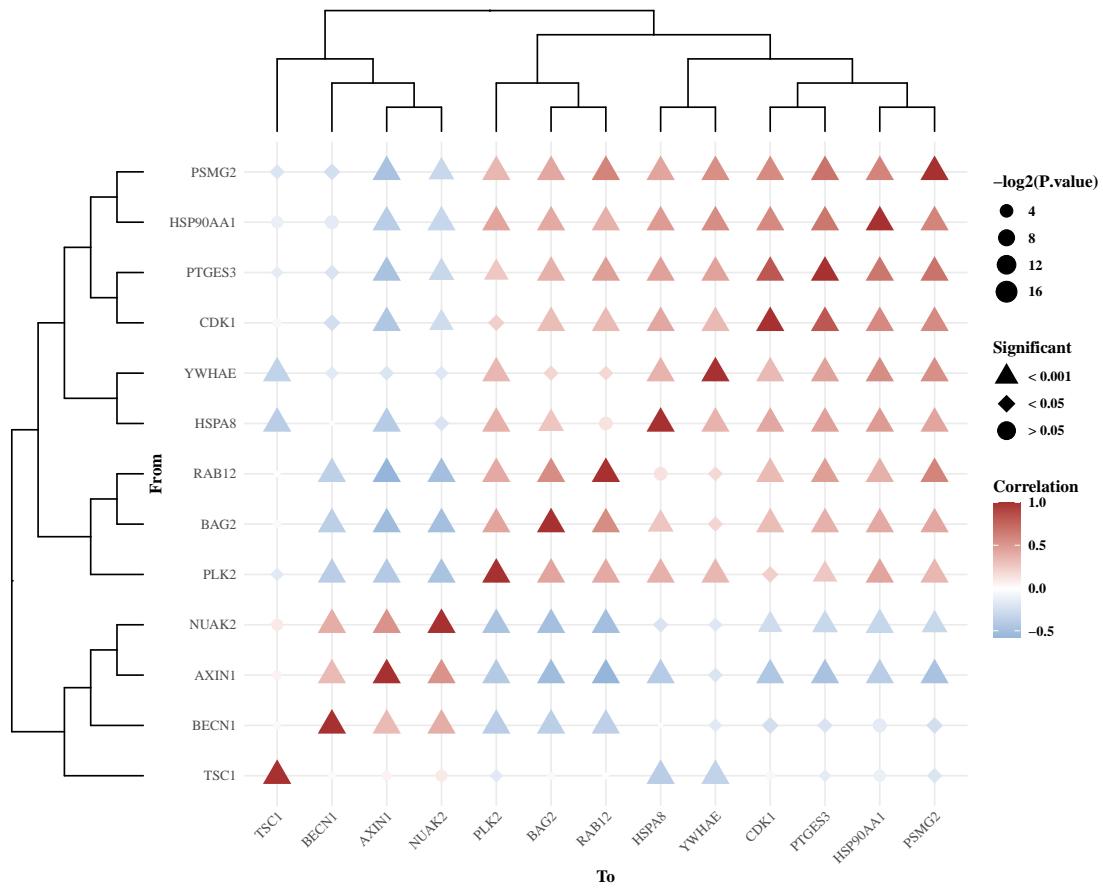


Figure 22: TCGA RNA correlation heatmap

Table 11 (下方表格) 为表格 TCGA RNA TSC1 negtive correlated 概览。

(对应文件为 [Figure+Table/TCGA-RNA-TSC1-negtive-correlated.csv](#))

注：表格共有 5 行 7 列，以下预览的表格可能省略部分数据；表格含有 5 个唯一 ‘From’。

1. cor: 皮尔逊关联系数，正关联或负关联。
2. pvalue: 显著性 P。
3. -log2(P.value): P 的对数转化。
4. significant: 显著性。
5. sign: 人为赋予的符号，参考 significant。

Table 11: TCGA RNA TSC1 negtive correlated

From	To	cor	pvalue	-log2(P.va...)	significant	sign
YWHAE	TSC1	-0.34	0	16.6096404...	< 0.001	**
HSPA8	TSC1	-0.38	0	16.6096404...	< 0.001	**
PTGES3	TSC1	-0.15	0.0497	4.33061033...	< 0.05	*
PSMG2	TSC1	-0.2	0.0061	7.35697504...	< 0.05	*

From	To	cor	pvalue	-log2(P.va...)	significant	sign
PLK2	TSC1	-0.17	0.0265	5.23786383...	< 0.05	*

## Reference

- Shao, Y. *et al.* Proteomics profiling of colorectal cancer progression identifies plod2 as a potential therapeutic target. *Cancer communications (London, England)* **42**, 164–169 (2022).
- Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
- Ritchie, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).
- Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data user's guide. 119.
- Geeleher, P., Cox, N. & Huang, R. S. PRRProphetic: An r package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PloS one* **9**, (2014).
- Szklarczyk, D. *et al.* The string database in 2021: Customizable proteinprotein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**, D605–D612 (2021).
- Chin, C.-H. *et al.* CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**, S11 (2014).
- Colaprico, A. *et al.* TCGAbiolinks: An r/bioconductor package for integrative analysis of tcga data. *Nucleic Acids Research* **44**, (2015).
- Dobson, L. *et al.* UniTmp: Unified resources for transmembrane proteins. *Nucleic acids research* **52**, D572–D578 (2024).
- Dobson, L., Reményi, I. & Tusnády, G. E. The human transmembrane proteome. *Biology direct* **10**, (2015).
- Yan, J., Wan, P., Choksi, S. & Liu, Z.-G. Necroptosis and tumor progression. *Trends in cancer* **8**, 21–27 (2022).
- Albakova, Z., Armeev, G. A., Kanevskiy, L. M., Kovalenko, E. I. & Sapozhnikov, A. M. HSP70 multi-functionality in cancer. *Cells* **9**, (2020).