

乙肝病毒 HBx 利用泛素化系统降解 XXX 上调 YYY 诱导肝癌线粒体自噬

2024-02-27

LiChuang Huang



@ 立效研究院

Contents

1 摘要	1
1.1 需求	1
1.2 结果	1
2 前言	2
3 材料和方法	2
3.1 材料	2
3.2 方法	2
4 分析结果	3
5 结论	3
6 附：分析流程	3
6.1 乙肝病毒 HBx 处理 DEGs	3
6.1.1 数据来源	3
6.1.2 DEGs	4
6.1.3 富集分析 (尝试)	5
6.2 线粒体自噬	7
6.2.1 GeneCards	7
6.3 DEGs 与线粒体自噬	8
6.3.1 交集 (Inter-DEGs-Mito)	8
6.3.2 PPI	9
6.4 泛素化	12
6.4.1 GeneCards	12
6.5 泛素化基因集与筛选基因集 (DEGs-down) 的相关性	13
6.5.1 关联热图	13
6.5.2 构建网络	14
6.6 整合: 泛素化 -> DEGs-down -> DEGs-up-Mitophagy	15
6.7 富集分析	16
6.7.1 KEGG	16
6.7.2 pathway visualization	18
6.7.3 富集于 hsa04120 (Ubiquitination) 与 hsa04137 (Mitophagy) 的基因	19
Reference	20

List of Figures

1 L02 Model vs Control DEGs	4
2 L02 KEGG enrichment with enriched genes	6

3	L02 GSEA plot of the pathways	7
4	Venn Intersection DEGs with Mitophagy related	8
5	UpSet Intersection DEGs with Mitophagy related	9
6	Raw PPI network	10
7	Filtered and formated PPI network	11
8	Top MCC score	12
9	L02 correlation heatmap	14
10	Correlation filtered	15
11	Integrated relationship	16
12	INTE KEGG enrichment	17
13	INTE GO enrichment	17
14	INTE hsa04120 visualization	18
15	INTE hsa04137 visualization	18
16	Co Exists in integrated relationship	19

List of Tables

1	L02 metadata	4
2	L02 data Model vs Control DEGs	5
3	MIT related targets from GeneCards	8
4	UBI related targets from GeneCards	13
5	Co Exists in integrated relationship data	19

1 摘要

1.1 需求

乙肝病毒 HBx 利用泛素化系统降解 XXX 上调 YYY 诱导肝癌线粒体自噬

筛选建议：

- 1、筛选乙肝病毒 HBx (乙型肝炎病毒的外壳蛋白) 处理诱导肝癌细胞差异表达基因集 A；
- 2、基因集 A 与线粒体自噬相关基因 B 的相关性 (PPI)；
- 3、筛选最佳相关性组合 XXX 和 YYY。

1.2 结果

注：与上述建议有不同之处，考虑了与泛素化相关基因的关联。

- 以 GSE186862 数据集差异分析获得基因集 DEGs (Fig. 1, Tab. 2)
- 获取自噬相关基因集 Mitophagy (Tab. 3)
- 分析 DEGs 中的上调、下调组与 Mitophagy 的交集 (Fig. 5)。
- DEGs 构建 PPI 网络 (Fig. 6):
 - 预计泛素化会导致基因的表达量下降¹，因此这里推断，受泛素化的 XXX 基因主要存在于 DEGs-down；随后，挖掘 DEGs-up-with-Mitophagy (DEGs-up 与 Mitophagy 交集) 与 DEGs-down 的关联 (Fig. 7)。
 - 根据 DEGs-down 的 MCC score 筛选 Top 10 (Fig. 8)。
- 获取泛素化相关基因集 (Tab. 4)
- 泛素化相关的筛选：
 - 将 Tab. 4 和 Fig. 8 中的 Top 10 DEGs-down 关联分析 (GSE186862 数据集)，获得关联热图 (Fig. 9)。
 - 以 P-value < 0.001 筛选 Fig. 9，得到 Fig. 10。
- 整合上述过程的数据：泛素化 -> DEGs-down -> DEGs-up-Mitophagy，Fig. 11
- 将整合后的所有基因富集分析，Fig. 12, Fig. 13:
 - 主要关注两条通路 (分别与泛素化和自噬相关)：Fig. 14, Fig. 15
 - 两条通路有交错的基因 (6.7.3): HUWE1, RPS27A
 - 根据交错基因重新整理 Fig. 11，得到 Fig. 16, Tab. 5
- 最终筛选：
 - 建议：结合通路 Mitophagy (Fig. 15)，和 Fig. 16，可发现：HUWE1 -> RPS27A (UB) -> ULK1 之间存在关联。
 - 额外：可根据 Tab. ?? 筛选其他可能。

2 前言

3 材料和方法

3.1 材料

All used GEO expression data and their design:

- **GSE186862**: mRNA profiles of L02-vector and L02-HBx cells

3.2 方法

Mainly used method:

- R package `ClusterProfiler` used for gene enrichment analysis².
- The Human Gene Database `GeneCards` used for disease related genes prediction³.
- GEO <https://www.ncbi.nlm.nih.gov/geo/> used for expression dataset acquisition.
- R package `ClusterProfiler` used for GSEA enrichment².
- R package `Limma` and `edgeR` used for differential expression analysis^{4,5}.
- R package `STEINGdb` used for PPI network construction^{6,7}.
- The MCC score was calculated referring to algorithm of `CytoHubba`⁷.
- 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

6.1 乙肝病毒 HBx 处理 DEGs

6.1.1 数据来源

Data Source ID :

GSE186862

data_processing :

Illumina Casava1.7 software used for basecalling.

data_processing.1 :

Sequenced reads were trimmed for adaptor sequence, and masked for low-complexity or low-quality sequence, then mapped to mm8 whole genome using bowtie v0.12.2 with parameters -q -p 4 -e 100 -y -a -m 10 –best –strata

data_processing.2 :

Reads Per Kilobase of exon per Megabase of library size (RPKM) were calculated using a protocol from Chepelev et al., Nucleic Acids Research, 2009. In short, exons from all isoforms of a gene were merged to create one meta-transcript. The number of reads falling in the exons of this meta-transcri...

data_processing.3 :

Genome_build: HG19

(Others) :

...

Table 1 (下方表格) 为表格 L02 metadata 概览。

(对应文件为 Figure+Table/L02-metadata.csv)

注: 表格共有 6 行 9 列, 以下预览的表格可能省略部分数据; 表格含有 6 个唯一 ‘rownames’。

1. sample: 样品名称
2. group: 分组名称

Table 1: L02 metadata

rownames	sample	group	lib.size	norm.f...	title	cell.l...	cell.t...	genoty...
A	A	Control	16890967	1	L02-ve...	L02	liver ...	control
C	C	Control	13575225	1	L02-ve...	L02	liver ...	control
E	E	Control	14827232	1	L02-ve...	L02	liver ...	control
B	B	Model	21985666	1	L02-HBx1	L02	liver ...	HBx ex...
D	D	Model	16595110	1	L02-HBx2	L02	liver ...	HBx ex...
F	F	Model	19786946	1	L02-HBx3	L02	liver ...	HBx ex...

注: 该 GSE 数据集的补充材料没有注明样品分组 (即, A、B、C.....等样品是属于哪个组别), 元数据表格中的分组信息, 是我根据原文 Figure 和 LogFC 数值推断的⁸。

6.1.2 DEGs

Figure 1 (下方图) 为图 L02 Model vs Control DEGs 概览。

(对应文件为 Figure+Table/L02-Model-vs-Control-DEGs.pdf)

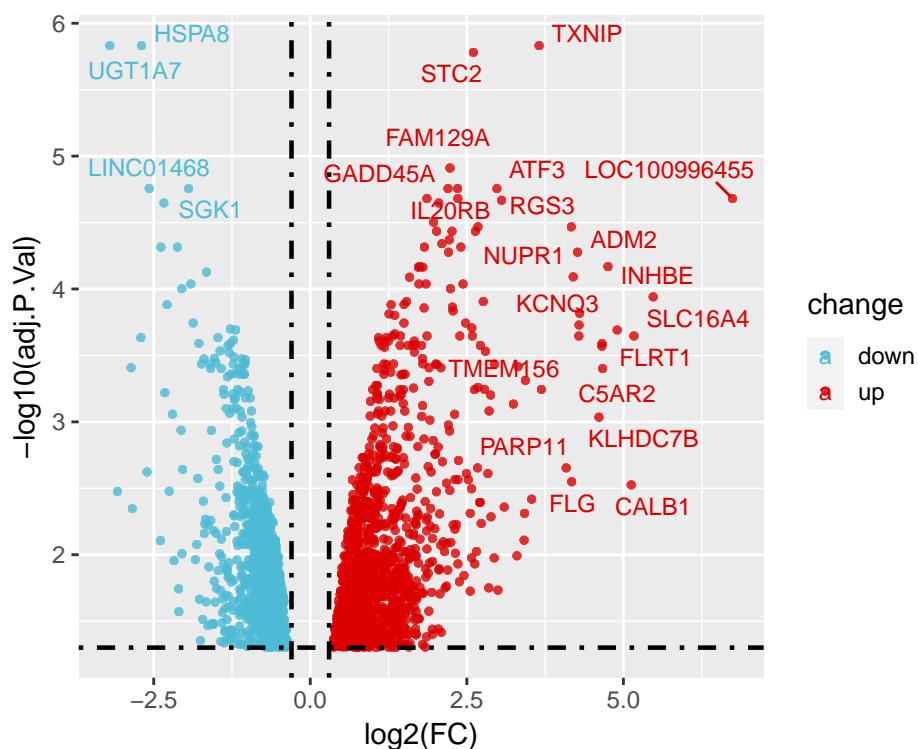


Figure 1: L02 Model vs Control DEGs

Table 2 (下方表格) 为表格 L02 data Model vs Control DEGs 概览。

(对应文件为 [Figure+Table/L02-data-Model-vs-Control-DEGs.xlsx](#))

注：表格共有 2352 行 12 列，以下预览的表格可能省略部分数据；表格含有 2205 个唯一‘Symbol’。

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: L02 data Model vs Control DEGs

rownames	AccID	AccID.1	Symbol	Strand	KeggID	logFC	AveExpr	t	P.Value
38934	ENSG00...	NM_019...	UGT1A7	hsa:54577	UDP gl...	-3.200...	5.1819...	-24.69...	2.2577...
12949	ENSG00...	NM_153...	HSPA8	hsa:3312	heat s...	-2.694...	10.198...	-23.85...	3.1834...
15559	ENSG00...	NM_006472	TXNIP	hsa:10628	thiore...	3.6547...	5.2722...	24.094...	2.8880...
22453	ENSG00...	NM_003...	STC2	hsa:8614	stanni...	2.6030...	7.9409...	22.909...	4.7679...
56468	ENSG00...	NM_052...	FAM129A	hsa:11...	family...	2.2312...	7.2870...	18.291...	4.4197...
52927	ENSG00...	NM_144...	IL20RB	hsa:53833	interl...	2.3520...	4.4217...	16.820...	1.0066...
22482	ENSG00...	NM_004...	ATF3	hsa:467	activa...	2.9807...	4.1438...	16.708...	1.0745...
15033	ENSG00...	NM_005...	SGK1	hsa:6446	serum/...	-1.942...	6.7155...	-16.43...	1.2627...
49697	ENSG00...	NM_001...	GADD45A	hsa:1647	growth...	2.1993...	5.0430...	16.435...	1.2622...
23878	ENSG00...	NR_120...	LINC01468	hsa:10...	long i...	-2.572...	3.8734...	-16.50...	1.2116...
39631	ENSG00...	NM_014...	PPP1R15A	hsa:23645	protei...	1.8632...	6.8100...	15.752...	1.9102...
5042	ENSG00...	NM_004...	F2RL2	hsa:2151	coagul...	2.3576...	3.9660...	15.718...	1.9511...
7150	ENSG00...	NM_002...	PTX3	hsa:5806	pentra...	3.0542...	3.7001...	15.552...	2.1645...
3587	ENSG00...	NM_002...	CXCL2	hsa:2920	chemok...	2.0506...	4.7692...	15.311...	2.5193...
58151	ENSG00...	NM_005...	HSPA1A	hsa:3303	heat s...	-2.338...	6.8642...	-15.26...	2.5935...
...

6.1.3 富集分析 (尝试)

Figure 2 (下方图) 为图 L02 KEGG enrichment with enriched genes 概览。

(对应文件为 [Figure+Table/L02-KEGG-enrichment-with-enriched-genes.pdf](#))

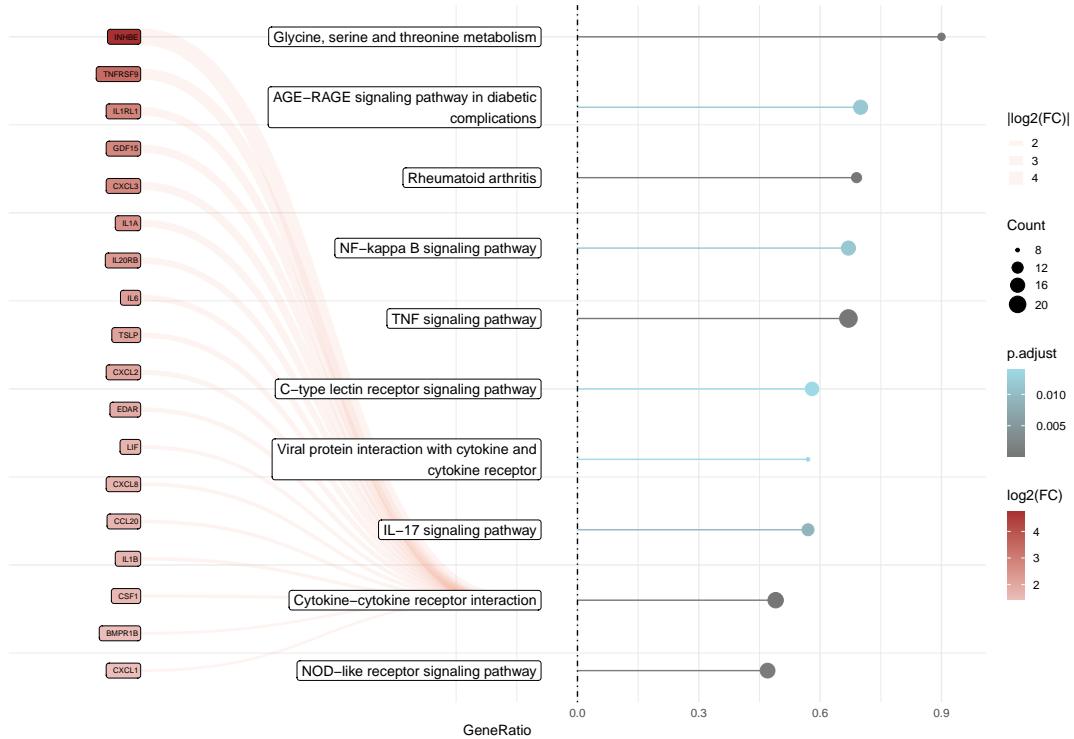


Figure 2: L02 KEGG enrichment with enriched genes

Figure 3 (下方图) 为图 L02 GSEA plot of the pathways 概览。

(对应文件为 Figure+Table/L02-GSEA-plot-of-the-pathways.pdf)

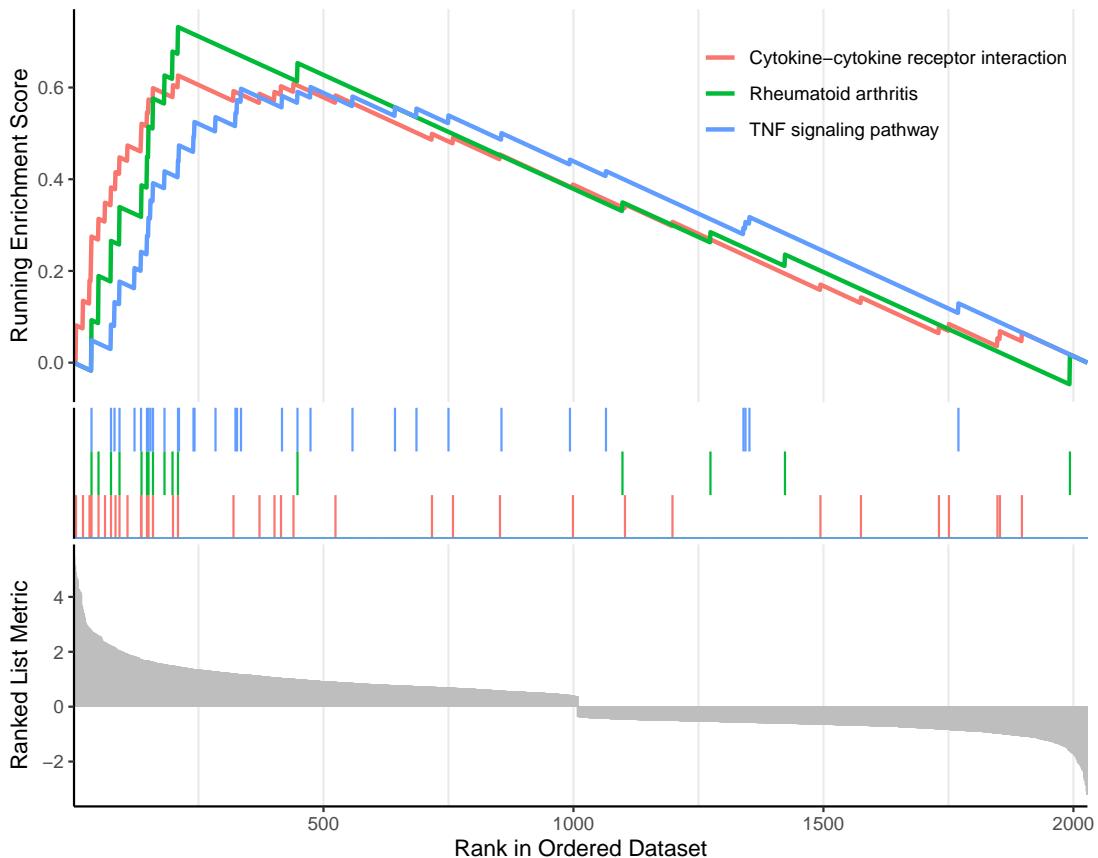


Figure 3: L02 GSEA plot of the pathways

6.2 线粒体自噬

6.2.1 GeneCards

The GeneCards data was obtained by filtering: :

Score > 1

Table 3 (下方表格) 为表格 MIT related targets from GeneCards 概览。

(对应文件为 Figure+Table/MIT-related-targets-from-GeneCards.xlsx)

注: 表格共有 1686 行 7 列, 以下预览的表格可能省略部分数据; 表格含有 1686 个唯一‘Symbol’。

Table 3: MIT related targets from GeneCards

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
PRKN	Parkin RBR...	Protein Co...	O60260	57	GC06M161348	19.14
PINK1	PTEN Induc...	Protein Co...	Q9BXM7	55	GC01P020634	18.01
MAP1LC3B	Microtubul...	Protein Co...	Q9GZQ8	50	GC16P087413	11.07
VDAC1	Voltage De...	Protein Co...	P21796	54	GC05M133975	10.11
FUNDC1	FUN14 Doma...	Protein Co...	Q8IVP5	37	GC0XM044523	9.21
MFN2	Mitofusin 2	Protein Co...	O95140	57	GC01P011980	9.05
SQSTM1	Sequestoso...	Protein Co...	Q13501	58	GC05P179806	8.40
MAP1LC3A	Microtubul...	Protein Co...	Q9H492	48	GC20P034546	7.92
ULK1	Unc-51 Lik...	Protein Co...	O75385	54	GC12P131894	7.22
UBC	Ubiquitin C	Protein Co...	P0CG48	51	GC12M124911	6.81
PHB2	Prohibitin 2	Protein Co...	Q99623	49	GC12M006965	6.75
ATG13	Autophagy ...	Protein Co...	O75143	49	GC11P047383	6.54
SOD2-OT1	SOD2 Overl...	RNA Gene		18	GC06M159772	6.47
TOMM20	Translocas...	Protein Co...	Q15388	48	GC01M235109	6.45
AMBRA1	Autophagy ...	Protein Co...	Q9C0C7	46	GC11M120823	6.26
...

6.3 DEGs 与线粒体自噬

6.3.1 交集 (Inter-DEGs-Mito)

Figure 4 (下方图) 为图 Venn Intersection DEGs with Mitophagy related 概览。

(对应文件为 Figure+Table/Venn-Intersection-DEGs-with-Mitophagy-related.pdf)

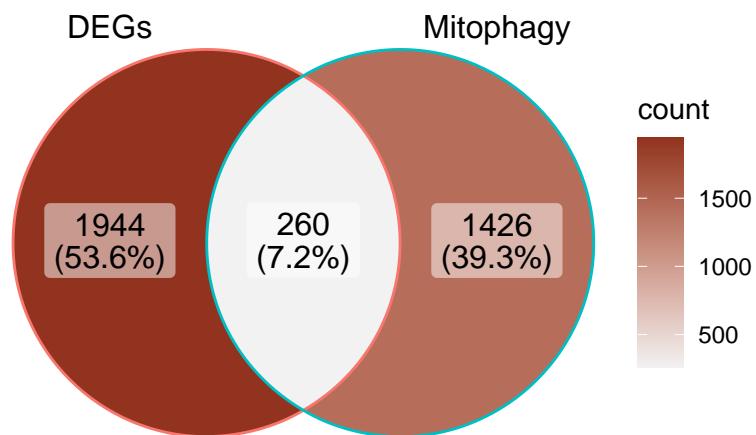


Figure 4: Venn Intersection DEGs with Mitophagy related

Intersection :

HSPA8, HSPA1A, SESN2, HMGCS1, RNF41, PSAT1, NDRG1, DNAJA1, PCK2, ZC3HAV1, SLFN11, RCAN1, KPNA2, BNIP3, UGP2, SHMT2, LDHA, VDAC1, PLOD2, ANLN, BNIP3L, NSDHL, PCYOX1, PHGDH, TRIM25, PDP1, SQSTM1, HSPH1, PLSCR1, SLC3A2, GABARAPL1, HK2, HSP90AA1, APAF1, LMO7, ARHGEF2, GPCPD1, NFKB1, CUL3, SMAD3, NFKB...

(上述信息框内容已保存至 Figure+Table/Venn-Intersection-DEGs-with-Mitophagy-related-content)

Figure 5 (下方图) 为图 UpSet Intersection DEGs with Mitophagy related 概览。

(对应文件为 Figure+Table/UpSet-Intersection-DEGs-with-Mitophagy-related.pdf)

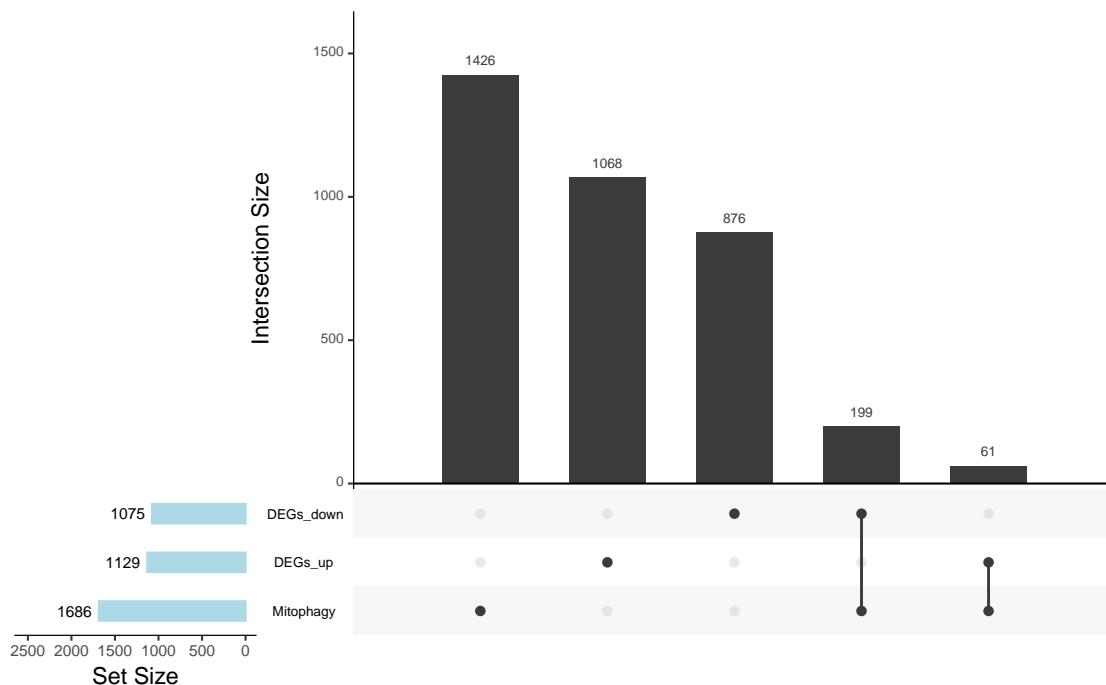


Figure 5: UpSet Intersection DEGs with Mitophagy related

All_intersection :

(上述信息框内容已保存至 Figure+Table/UpSet-Intersection-DEGs-with-Mitophagy-related-content)

6.3.2 PPI

构建 DEGs 的 PPI 网络。

Figure 6 (下方图) 为图 Raw PPI network 概览。

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

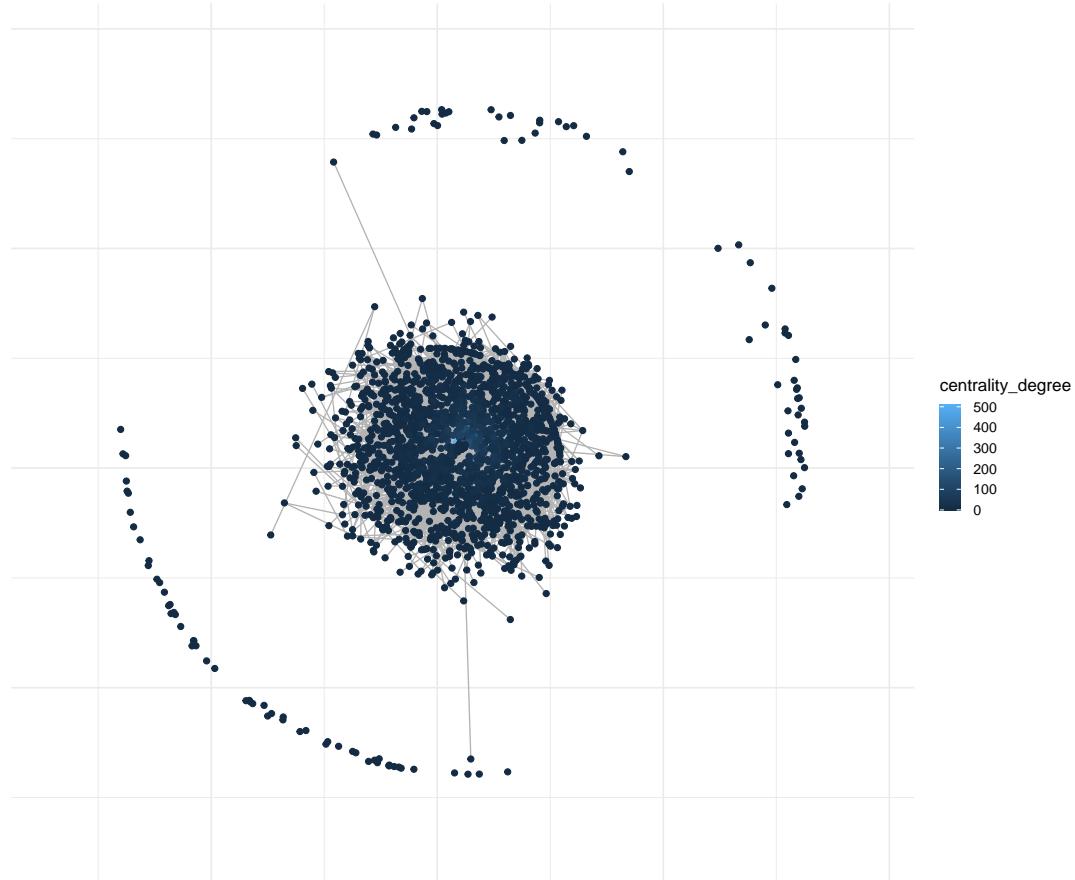


Figure 6: Raw PPI network

预计泛素化会导致基因的表达量下降¹，因此这里可以推断，受泛素化的 XXX 基因主要存在于 DEGs-down。挖掘 DEGs-up-with-Mitophagy (DEGs-up 与 Mitophagy 交集) 与 DEGs-down 的关联。

Figure 7 (下方图) 为图 Filtered and formated PPI network 概览。

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

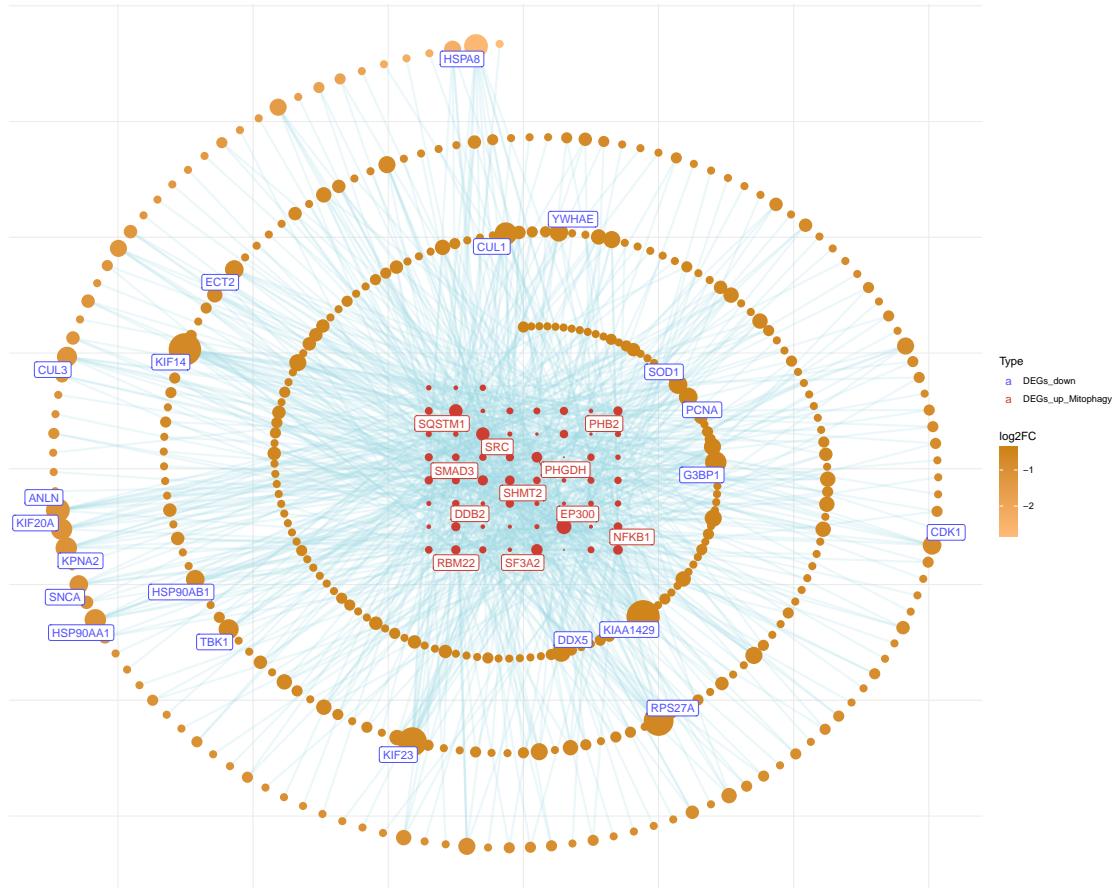


Figure 7: Filtered and formated PPI network

根据 DEGs-down 的 MCC score 筛选 Top 10。

Figure 8 (下方图) 为图 Top MCC score 概览。

(对应文件为 Figure+Table/Top-MCC-score.pdf)

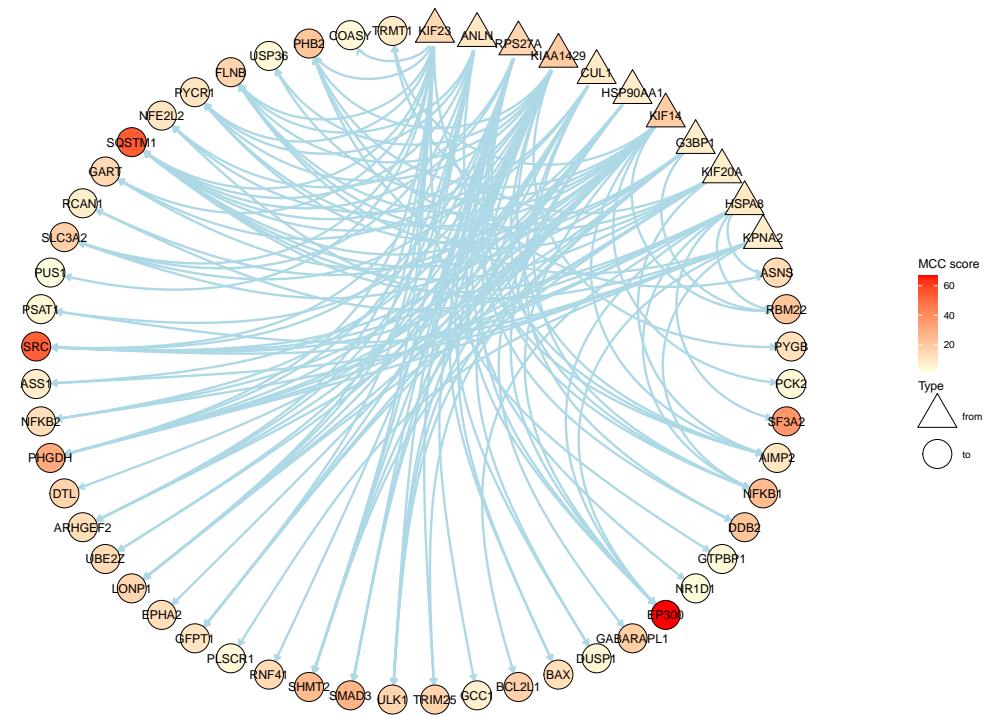


Figure 8: Top MCC score

6.4 泛素化

6.4.1 GeneCards

The GeneCards data was obtained by filtering: :

Score > 15

Table 4 (下方表格) 为表格 UBI 相关靶标从 GeneCards 概览。

(对应文件为 Figure+Table/UBI-related-targets-from-GeneCards.xlsx)

注：表格共有 161 行 7 列，以下预览的表格可能省略部分数据；表格含有 161 个唯一‘Symbol’。

Table 4: UBI related targets from GeneCards

Symbol	Description	Category	UniProt_ID	GIFtS	GC_id	Score
RPS27A	Ribosomal ...	Protein Co...	P62979	51	GC02P055231	41.57
PRKN	Parkin RBR...	Protein Co...	O60260	57	GC06M161348	40.64
UBC	Ubiquitin C	Protein Co...	P0CG48	51	GC12M124911	37.54
UBE2D1	Ubiquitin ...	Protein Co...	P51668	52	GC10P058334	37.22
UBE2D3	Ubiquitin ...	Protein Co...	P61077	52	GC04M102794	35.79
UBE2D2	Ubiquitin ...	Protein Co...	P62837	51	GC05P139526	35.33
UBE2L3	Ubiquitin ...	Protein Co...	P68036	54	GC22P021549	33.3
UBE2N	Ubiquitin ...	Protein Co...	P61088	55	GC12M093406	32.84
RBX1	Ring-Box 1	Protein Co...	P62877	51	GC22P040951	30.81
USP7	Ubiquitin ...	Protein Co...	Q93009	57	GC16M008892	30.55
VCP	Valosin Co...	Protein Co...	P55072	58	GC09M035056	30.55
UBE3A	Ubiquitin ...	Protein Co...	Q05086	56	GC15M025333	30.26
MDM2	MDM2 Proto...	Protein Co...	Q00987	62	GC12P068808	30.23
STUB1	STIP1 Homo...	Protein Co...	Q9UNE7	54	GC16P064961	30.01
UBE4B	Ubiquitina...	Protein Co...	O95155	49	GC01P010032	29.68
...

6.5 泛素化基因集与筛选基因集 (DEGs-down) 的相关性

6.5.1 关联热图

将 Tab. 4 和 Fig. 8 中的 Top 10 DEGs-down 关联分析 (GSE186862 数据集)。

Figure 9 (下方图) 为图 L02 correlation heatmap 概览。

(对应文件为 [Figure+Table/L02-correlation-heatmap.pdf](#))

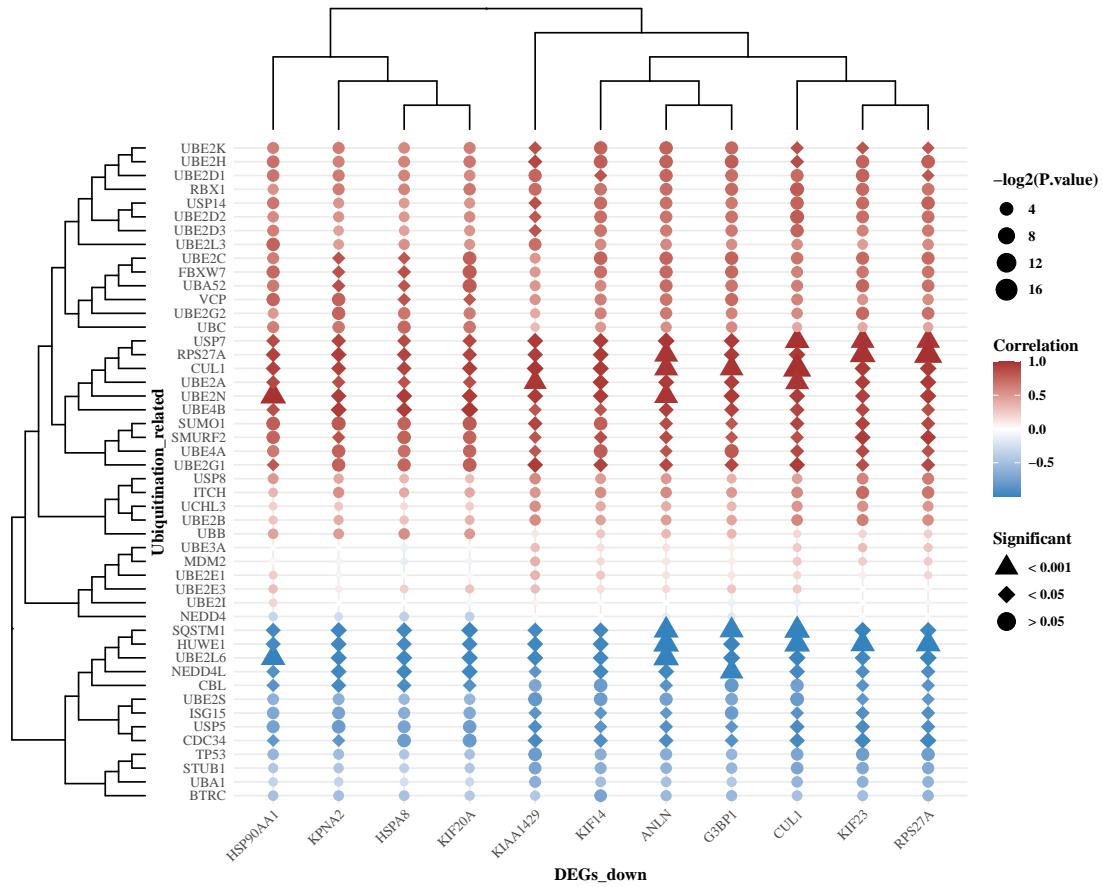


Figure 9: L02 correlation heatmap

6.5.2 构建网络

以 $P\text{-value} < 0.001$ 筛选 Fig. 9。

Figure 10 (下方图) 为图 Correlation filtered 概览。

(对应文件为 Figure+Table/Correlation-filtered.pdf)

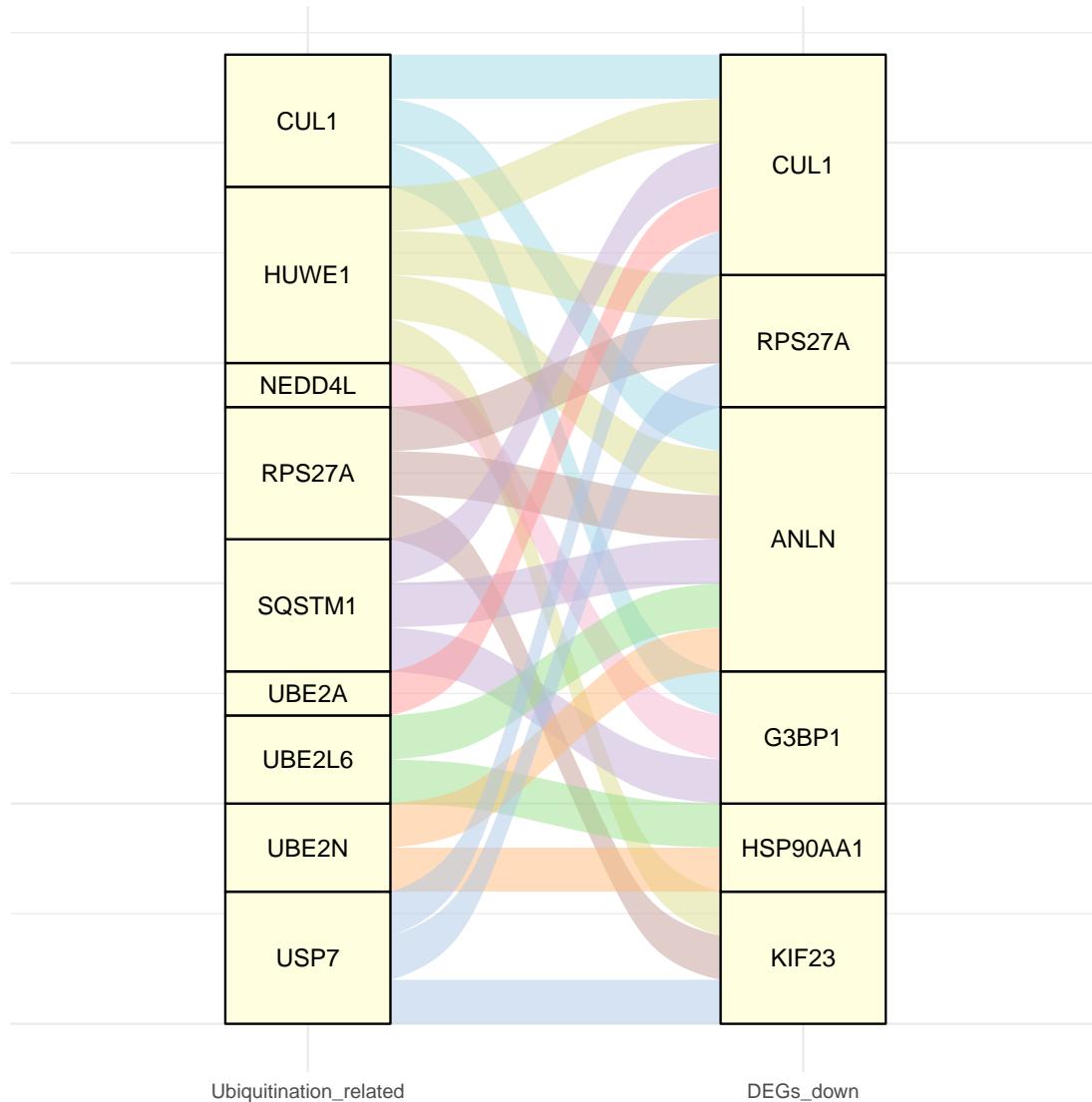


Figure 10: Correlation filtered

6.6 整合: 泛素化 -> DEGs-down -> DEGs-up-Mitophagy

Figure 11 (下方图) 为图 integrated relationship 概览。

(对应文件为 Figure+Table/integrated-relationship.pdf)

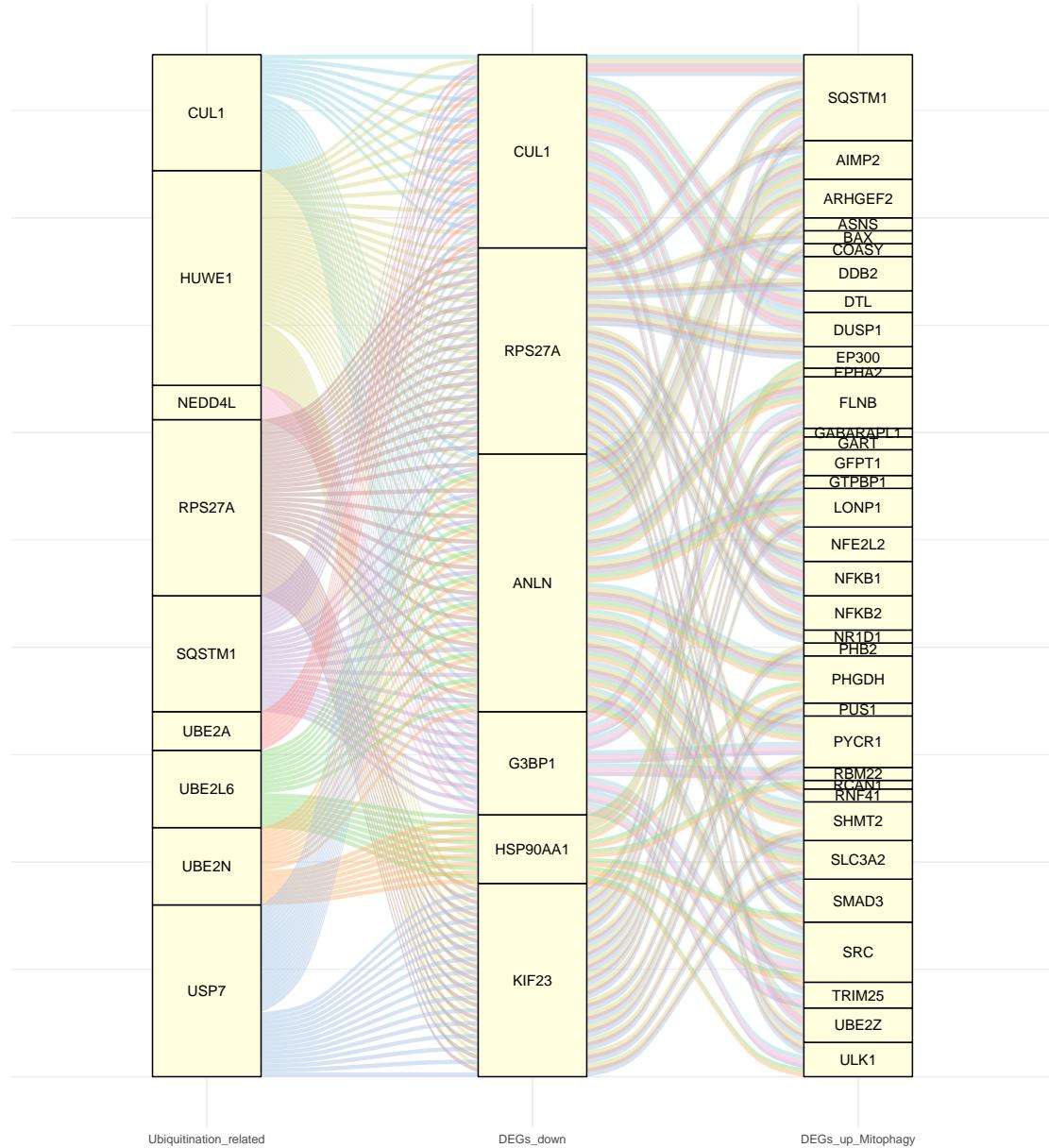


Figure 11: Integrated relationship

6.7 富集分析

6.7.1 KEGG

Figure 12 (下方图) 为图 INT-E KEGG enrichment 概览。

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

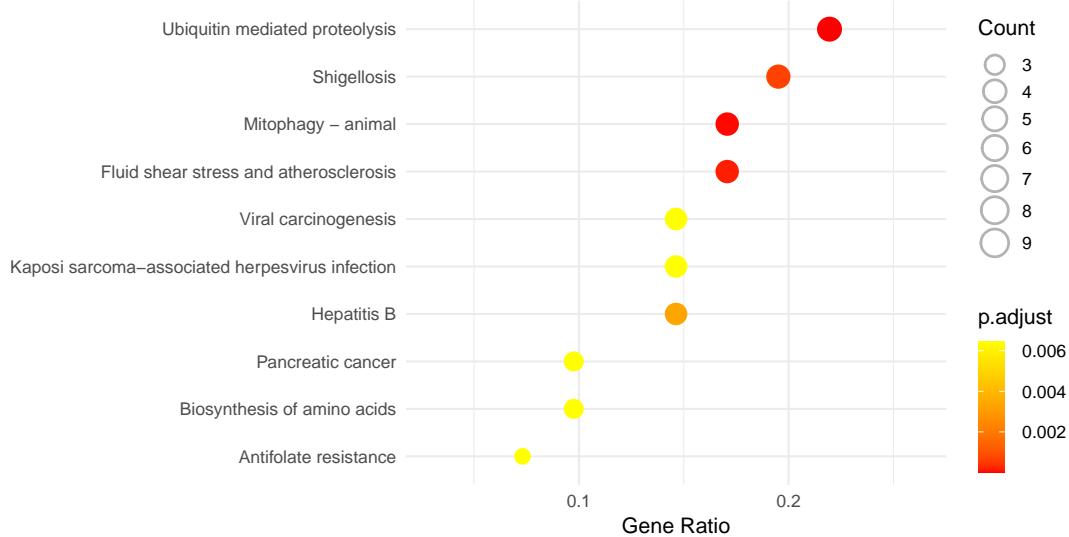


Figure 12: INTE KEGG enrichment

Figure 13 (下方图) 为图 INTE GO enrichment 概览。

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

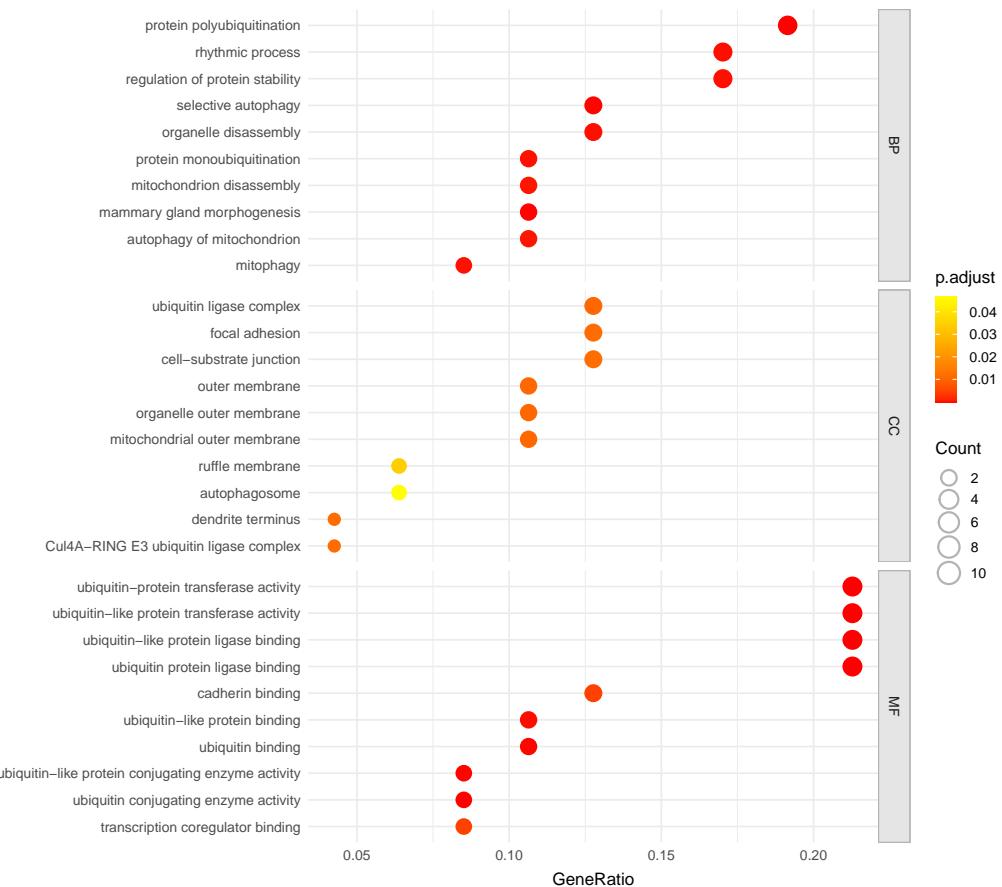


Figure 13: INTE GO enrichment

6.7.2 pathway visualization

Figure 14 (下方图) 为图 INT hsa04120 visualization 概览。

(对应文件为 Figure+Table/hsa04120.pathview.png)

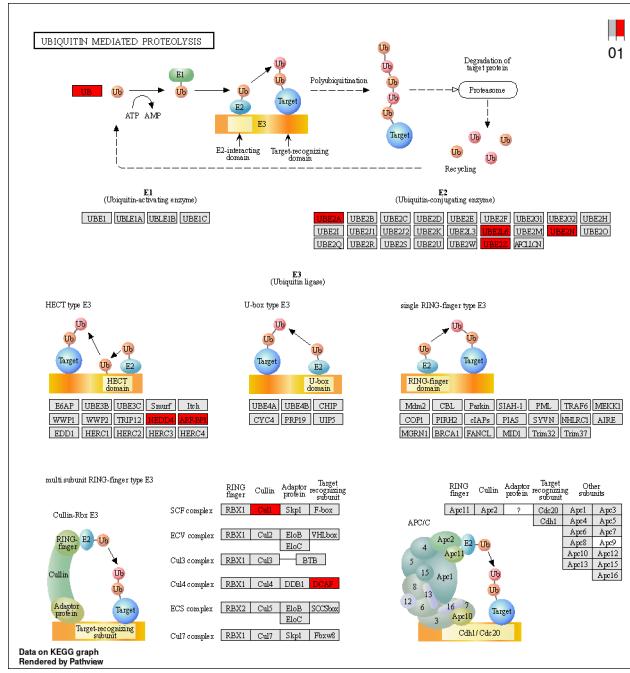


Figure 14: INTE hsa04120 visualization

Figure 15 (下方图) 为图 INTE hsa04137 visualization 概览。

(对应文件为 Figure+Table/hsa04137.pathview.png)

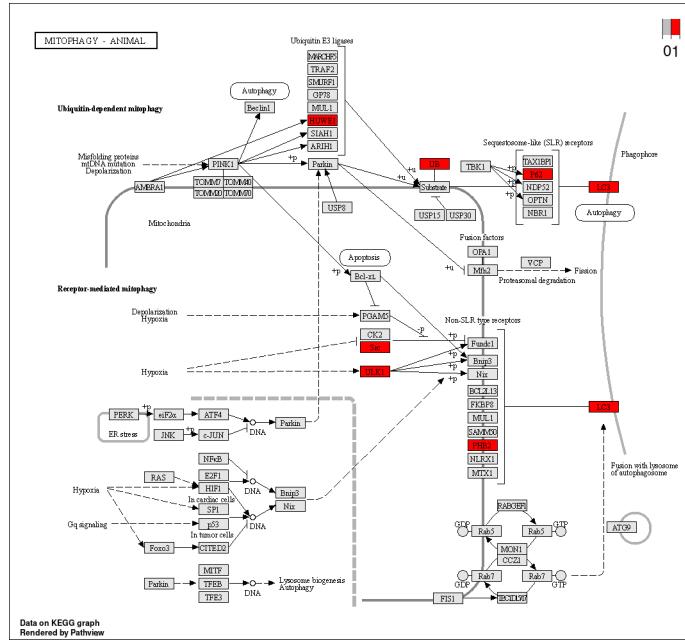


Figure 15: INTE hsa04137 visualization

6.7.3 富集于 hsa04120 (Ubiquitination) 与 hsa04137 (Mitophagy) 的基因

Content :

HUWE1, RPS27A

Figure 16 (下方图) 为图 co Exists in integrated relationship 概览。

(对应文件为 Figure+Table/co-Exists-in-integrated-relationship.pdf)

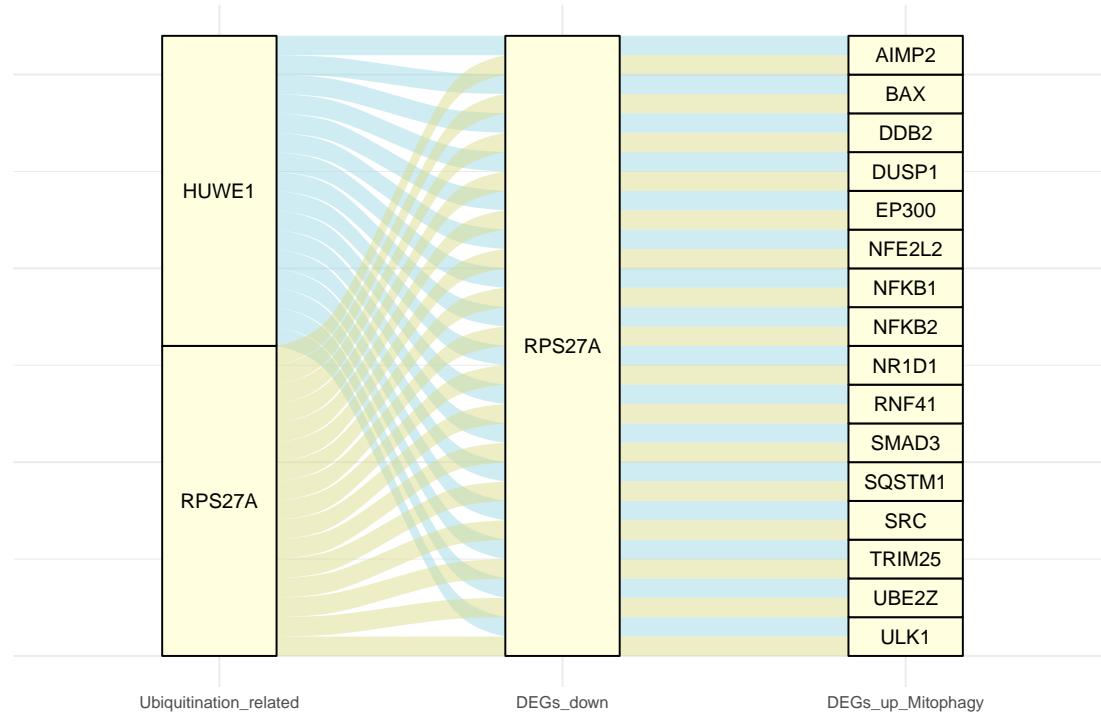


Figure 16: Co Exists in integrated relationship

Table 5 (下方表格) 为表格 co Exists in integrated relationship data 概览。

(对应文件为 Figure+Table/co-Exists-in-integrated-relationship-data.csv)

注：表格共有 32 行 3 列，以下预览的表格可能省略部分数据；表格含有 2 个唯一‘Ubiquitination_related’。

Table 5: Co Exists in integrated relationship data

Ubiquitination_related	DEGs_down	DEGs_up_Mitophagy
RPS27A	RPS27A	AIMP2
RPS27A	RPS27A	NFKB1
RPS27A	RPS27A	DUSP1
RPS27A	RPS27A	NR1D1

Ubiquitination_related	DEGs_down	DEGs_up_Mitophagy
RPS27A	RPS27A	DDB2
RPS27A	RPS27A	EP300
RPS27A	RPS27A	BAX
RPS27A	RPS27A	TRIM25
RPS27A	RPS27A	ULK1
RPS27A	RPS27A	SMAD3
RPS27A	RPS27A	RNF41
RPS27A	RPS27A	UBE2Z
RPS27A	RPS27A	NFKB2
RPS27A	RPS27A	SRC
RPS27A	RPS27A	SQSTM1
...

Reference

- Popovic, D., Vucic, D. & Dikic, I. Ubiquitination in disease pathogenesis and treatment. *Nature Medicine* **20**, (2014).
- Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
- Stelzer, G. *et al.* The genecards suite: From gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54**, 1.30.1–1.30.33 (2016).
- 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.
- 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).
- Zheng, C., Liu, M., Ge, Y., Qian, Y. & Fan, H. HBx increases chromatin accessibility and etv4 expression to regulate dishevelled-2 and promote hcc progression. *Cell death & disease* **13**, (2022).