

# Analysis

Huang LiChuang of Wie-Biotech

## Contents

<b>1 题目</b>	<b>2</b>
<b>2 摘要</b>	<b>2</b>
<b>3 研究设计流程图</b>	<b>3</b>
<b>4 材料和方法</b>	<b>3</b>
4.1 Methods . . . . .	3
4.2 materials . . . . .	3
<b>5 分析结果</b>	<b>3</b>
5.1 结直肠癌细胞的特定药物化疗差异分析 . . . . .	3
5.2 TCGA 临床样本的泛型化疗差异分析 . . . . .	3
5.3 肌少症的基因共表达分析 . . . . .	3
5.4 结直肠癌与肌少症的综合分析 . . . . .	3
<b>6 结论</b>	<b>3</b>
<b>7 附：分析流程</b>	<b>3</b>
7.1 相关文献 . . . . .	3
7.2 GEO 结直肠癌（细胞样本） . . . . .	4
7.2.1 GSE142340: six CRC cell lines treated with CTRL and optimized drug combinations (ODCs) . . . . .	4
7.2.2 GSE153412: radio-chemoresistance in colorectal cancer cell lines . . . . .	5
7.3 TCGA 直肠癌（TCGA-READ） . . . . .	7
7.4 TCGA 结肠癌（TCGA-COAD） . . . . .	9
7.5 GEO 肌少症 . . . . .	11
7.5.1 GSE167186: transcriptome profiling on lower limb muscle biopsies from 72 young, old and sarcopenic subjects . . . . .	11
7.6 综合：结直肠癌和肌少症 . . . . .	18
7.6.1 （结直肠癌）数据整合 . . . . .	18
7.6.2 （结直肠癌与肌少症）交集基因 . . . . .	18
7.6.3 富集分析 . . . . .	19

7.6.4 通路可视化 . . . . .	20
<b>Reference</b>	<b>22</b>

## List of Figures

1	DEGs in different cell types with chemotherapy or not GSE142340 . . . . .	5
2	DEGs in different cell types with chemotherapy or not GSE153412 . . . . .	7
3	READ whether with chemotherapy . . . . .	8
4	READ difference expressed genes . . . . .	9
5	COAD whether with chemotherapy . . . . .	10
6	COAD difference expressed genes . . . . .	11
7	Filtering of Sarcopenia datasets . . . . .	12
8	Normalization of Sarcopenia datasets . . . . .	12
9	Whether with Sarcopenia . . . . .	14
10	Soft threshold . . . . .	15
11	Clustering of gene modules . . . . .	16
12	Correlation of gene modules and traits data . . . . .	17
13	Intersection of genes significant and module memberships . . . . .	18
14	All colorectal DEGs . . . . .	18
15	Intersection of colorectal DEGs with Sarcopenia significant genes . . . . .	19
16	Go enrichment . . . . .	19
17	Kegg enrichment . . . . .	20
18	Hits in autophagy . . . . .	21
19	Hits in Mitophagy . . . . .	21

## List of Tables

1	READ clinical data . . . . .	7
2	COAD clinical data . . . . .	9
3	Metadata of samples used in GEO Sarcopenia data . . . . .	13

## 1 题目

## 2 摘要

关于转录组数据库筛选肌少症（Sarcopenia）、结直肠癌（colorectal cancer）、化疗（Chemotherapy）共同的通路：

### 3 研究设计流程图

### 4 材料和方法

#### 4.1 Methods

- GEOquery
- TCGAbiolinks
- limma
- edgeR
- clusterProfiler
- R
- ...

#### 4.2 materials

使用的数据集见 7

### 5 分析结果

#### 5.1 结直肠癌细胞的特定药物化疗差异分析

#### 5.2 TCGA 临床样本的泛型化疗差异分析

#### 5.3 肌少症的基因共表达分析

#### 5.4 结直肠癌与肌少症的综合分析

### 6 结论

### 7 附：分析流程

#### 7.1 相关文献

- Skeletal muscle-specific Keap1 disruption modulates fatty acid utilization and enhances exercise capacity in female mice<sup>1</sup>
- Mitochondrial dysfunction and oxidative stress in aging and cancer<sup>2</sup>
- Danger signals: Chemotherapy enhancers?<sup>3</sup>
- Current Targeted Therapy for Metastatic Colorectal Cancer<sup>4</sup>
- Sarcopenia in the Older Adult With Cancer<sup>5</sup>
- The role of aging in cancer<sup>6</sup>

## 7.2 GEO 结直肠癌 (细胞样本)

### 7.2.1 GSE142340: six CRC cell lines treated with CTRL and optimized drug combinations (ODCs)

- RNA sequencing was conducted for six CRC cell lines treated with CTRL and optimized drug combinations (ODCs), providing samples in duplicate
  - GSE142340

#### **data\_\_processing :**

Illumina Casava2.2 software used for basecalling.

#### **data\_\_processing.1 :**

Sequenced reads were mapped to rn6 whole genome using STAR v2.5.3a with default parameters

#### **data\_\_processing.2 :**

Raw counts are produced by htseq-count (HTSeq v.0.9.1)

#### **data\_\_processing.3 :**

Normalization and differential expression analysis were performed with edgeR v.3.24.3

#### **data\_\_processing.4 :**

Genome\_build: hg38

#### **data\_\_processing.5 :**

Supplementary\_files\_format\_and\_content: tab-delimited text files include rawcount values for each Sample

#### **data\_\_processing.6 :**

Supplementary\_files\_format\_and\_content: tab-delimited text files include normalized expression values(cpm = count per millions) for each Sample

Figure 1为图 DEGs in different cell types with chemotherapy or not GSE142340 概览。

(对应文件为 **Figure+Table/DEGs-in-different-cell-types-with-chemotherapy-or-not-GSE142340.pdf**)

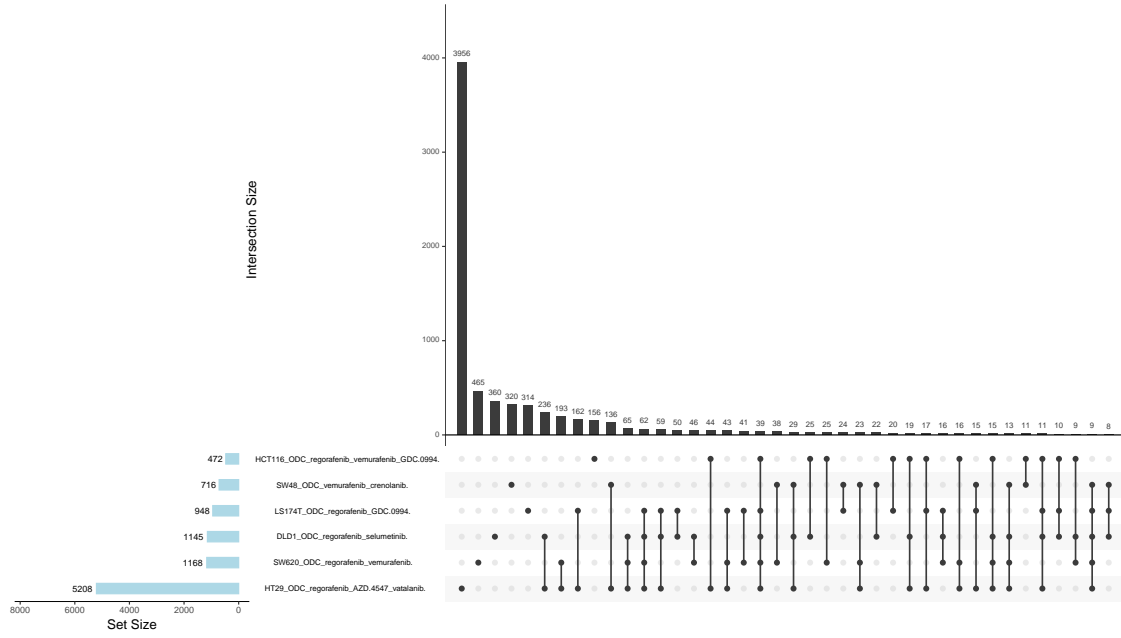


Figure 1: DEGs in different cell types with chemotherapy or not GSE142340

## 7.2.2 GSE153412: radio-chemoresistance in colorectal cancer cell lines

- RNAseq analysis of radio-chemoresistance in colorectal cancer cell lines
  - GSE153412

**data\_processing :**

FASTQ files were generated using bcl2fastq version v2.20.0.422

**data\_processing.1 :**

FASTQ files were filtered using Trimmomatic 0.36 (PE ILLUMINA-CLIP:adapters.fa:2:30:10 LEADING:5 TRAILING:5 MINLEN:45).

**data\_processing.2 :**

The quality was assessed with FastQC 0.11.8.

**data\_processing.3 :**

The transcriptome was generated with gffread, using the GRCh38.p13 genome and the latest Ensembl annotation (Homo sapiens version 98).

**data\_processing.4 :**

Complete read pairs were aligned and quantified on a human transcriptome using Kallisto 0.44.0 (index built with -kmer-size=31).

**data\_processing.5 :**

Genome\_build: GRCh38.p13

**data\_processing.6 :**

Supplementary\_files\_format\_and\_content: Kallisto output tsv file with transcripts abundance (raw counts and tpm).

**data\_processing.7 :**

Supplementary\_files\_format\_and\_content: Matrix table with raw gene counts for every gene and every sample

**data\_processing.8 :**

Supplementary\_files\_format\_and\_content: Matrix table with tpm values for every gene and every sample

Figure 2为图 DEGs in different cell types with chemotherapy or not GSE153412 概览。

(对应文件为 Figure+Table/DEGs-in-different-cell-types-with-chemotherapy-or-not-GSE153412.pdf)

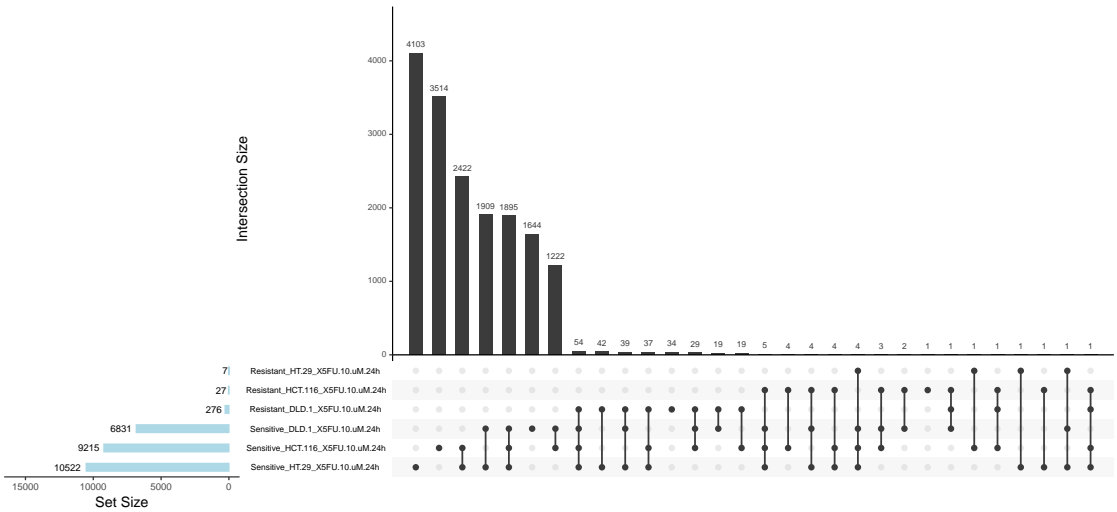


Figure 2: DEGs in different cell types with chemotherapy or not GSE153412

### 7.3 TCGA 直肠癌 (TCGA-READ)

Table 1为表格 READ clinical data 概览。

(对应文件为 Figure+Table/READ-clinical-data.xlsx)

注：表格共有 135 行 111 列，以下预览的表格可能省略部分数据；表格含有 135 个唯一 ‘sample’。

Table 1: READ clinical data

group	lib.size	norm....	sample	barcode	patient	short.....7	defin...	sampl.....9	sampl.....10	...
yes	60935...	1.250...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	52344...	0.910...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
no	48571...	1.031...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	50713...	0.857...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
yes	60107...	0.945...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
no	47905...	0.993...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	19924...	0.898...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	63999...	1.069...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	21927...	1.060...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	51616...	0.974...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	33491...	0.871...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
yes	33640...	1.124...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	38780...	1.124...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	41384...	1.106...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	53539...	1.104...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...

group	lib.size	norm....	sample	barcode	patient	short.....7	defin...	sampl.....9	sampl.....10	...
...	...	...	...	...	...	...	...	...	...	...

统计是否化疗：

Figure 3为图 READ whether with chemotherapy 概览。

(对应文件为 `Figure+Table/READ-whether-with-chemotherapy.pdf`)

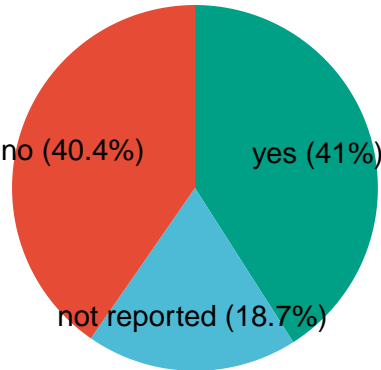


Figure 3: READ whether with chemotherapy

Figure 4为图 READ difference expressed genes 概览。

(对应文件为 `Figure+Table/READ-difference-expressed-genes.pdf`)



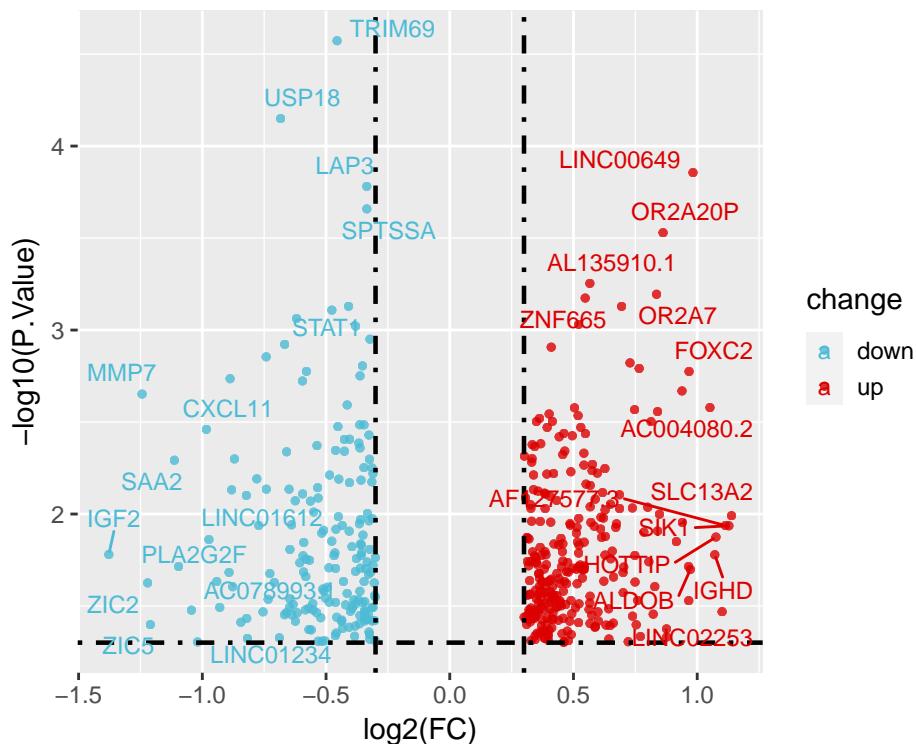


Figure 4: READ difference expressed genes

## 7.4 TCGA 结肠癌 (TCGA-COAD)

Table 2为表格 COAD clinical data 概览。

(对应文件为 `Figure+Table/COAD-clinical-data.xlsx`)

注：表格共有 388 行 112 列，以下预览的表格可能省略部分数据；表格含有 388 个唯一 ‘sample’。

Table 2: COAD clinical data

group	lib.size	norm....	sample	barcode	patient	short.....7	defin...	sampl.....9	sampl.....10	...
no	50725...	1.167...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	46391...	1.030...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	29081...	0.988...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	51150...	1.097...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	33677...	1.080...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	31679...	1.073...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	44720...	0.781...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
no	21922...	1.073...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	22321...	1.326...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
no	35916...	0.872...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...

group	lib.size	norm....	sample	barcode	patient	short.....7	defin...	sampl.....9	sampl.....10	...
yes	19466...	0.969...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	30802...	1.147...	TCGA-...	TCGA-...	TCGA-...	TP	Prima...	TCGA-...	01	...
yes	52195...	0.909...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
no	40206...	0.831...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
no	46895...	0.751...	TCGA-...	TCGA-...	TCGA-...	NT	Solid...	TCGA-...	11	...
...	...	...	...	...	...	...	...	...	...	...

统计是否化疗：

Figure 5为图 COAD whether with chemotherapy 概览。

(对应文件为 `Figure+Table/COAD-whether-with-chemotherapy.pdf`)

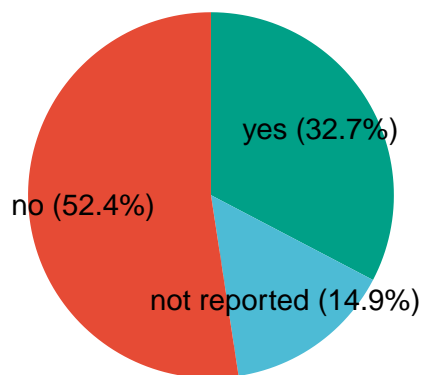


Figure 5: COAD whether with chemotherapy

使用 Limma 计算差异表达基因。

Figure 6为图 COAD difference expressed genes 概览。

(对应文件为 `Figure+Table/COAD-difference-expressed-genes.pdf`)

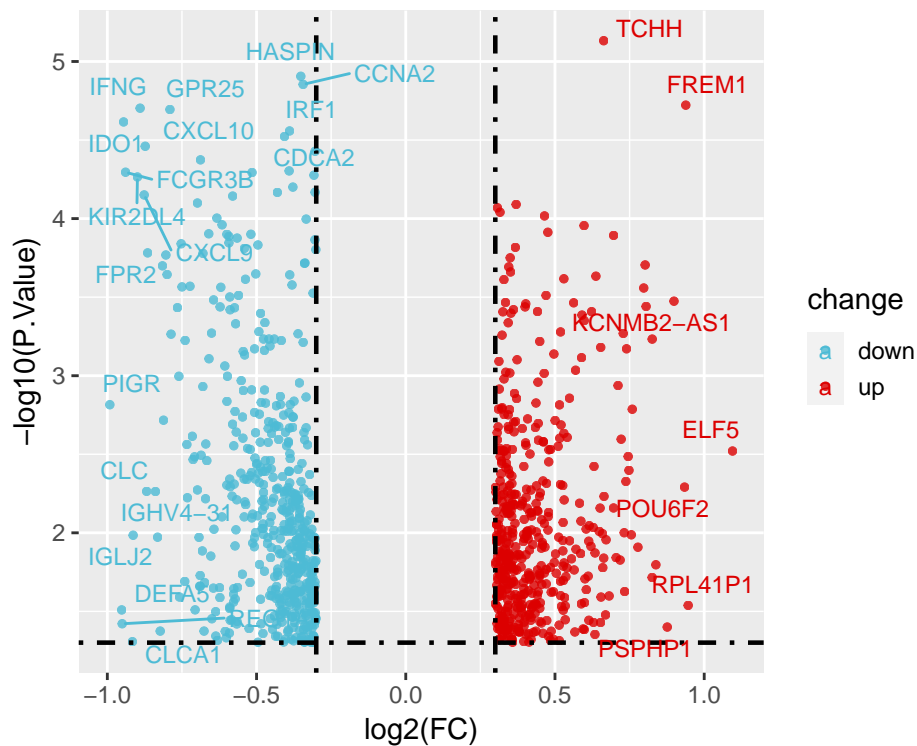


Figure 6: COAD difference expressed genes

## 7.5 GEO 肌少症

### 7.5.1 GSE167186: transcriptome profiling on lower limb muscle biopsies from 72 young, old and sarcopenic subjects

- We performed transcriptome profiling on lower limb muscle biopsies from 72 young, old and sarcopenic subjects using bulk RNA-seq (N = 72) and single-nuclei RNA-seq (N = 17).
  - GSE167186

#### 7.5.1.1 edgeR 对数据进行标准化处理。

Figure 7为图 filtering of Sarcopenia datasets 概览。

(对应文件为 [Figure+Table/filtering-of-Sarcopenia-datasets.pdf](#))

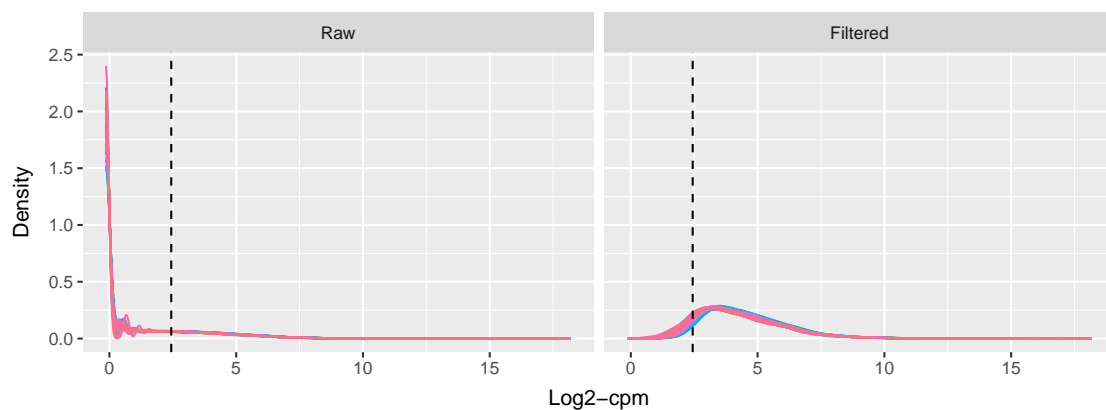


Figure 7: Filtering of Sarcopenia datasets

Figure 8为图 nomalization of Sarcopenia datasets 概览。

(对应文件为 `Figure+Table/nomalization-of-Sarcopenia-datasets.pdf`)

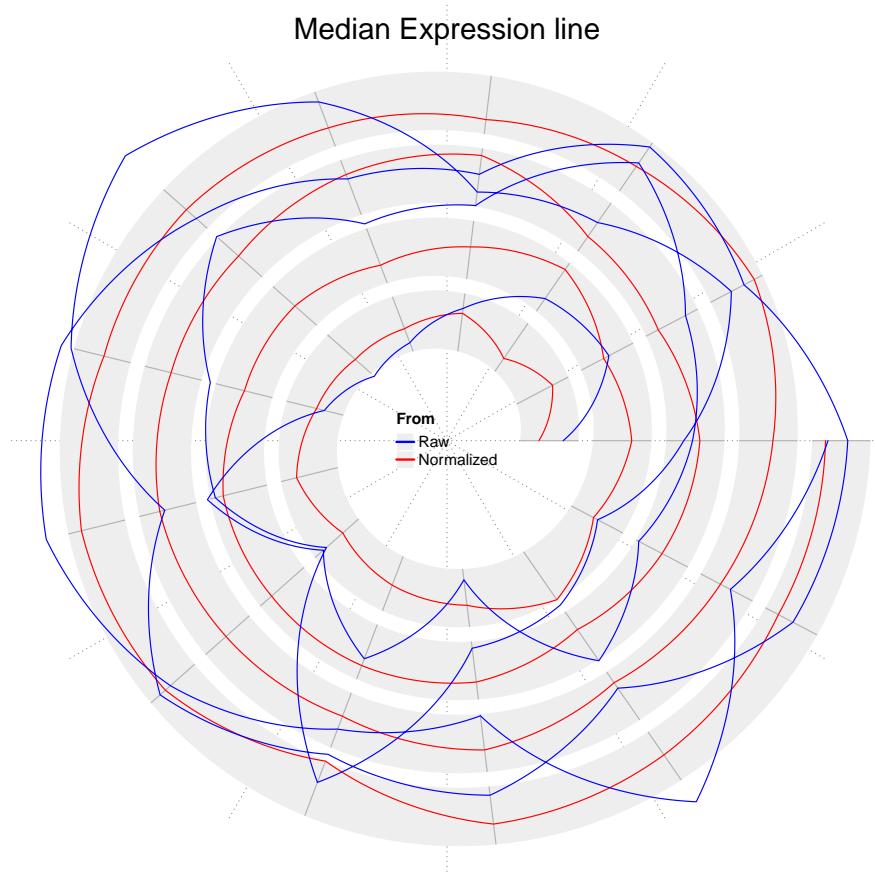


Figure 8: Nomalization of Sarcopenia datasets

**7.5.1.2 WGCNA** Table 3为表格 metadata of samples used in GEO Sarcopenia data 概览。

(对应文件为 `Figure+Table/metadata-of-samples-used-in-GEO-Sarcopenia-data.csv`)

注：表格共有 53 行 16 列，以下预览的表格可能省略部分数据；表格含有 53 个唯一 ‘sample’。

Table 3: Metadata of samples used in GEO Sarcopenia data

rownames	group	lib.size	norm....	sample	X6.mi...	age.ch1	biode...	grip....	group...	leg.p...	...
X_10	OLD_S...	23299...	1.064...	X_10	1.010...	68	166.2	41.8	Sarco...	130.5	...
X_11	OLD_H...	24242...	1.078...	X_11	1.088...	87	136.4	24.2	Old H...	129.55	...
X_13	OLD_S...	17915...	0.968...	X_13	1.425...	83	143.1	40	Sarco...	127.27	...
X_14	OLD_H...	22048...	0.932...	X_14	0.918...	77	222.1	35.5	Old H...	229.5	...
X_15	OLD_H...	21527...	0.916...	X_15	NA	82	127.9	35.1	Old H...	94.5	...
X_16	OLD_H...	15754...	0.949...	X_16	NA	73	202.7	49.1	Old H...	145.5	...
X_17	OLD_S...	23226...	1.116...	X_17	0.886...	77	171.7	45.5	Sarco...	87.5	...
X_18	OLD_H...	22781...	0.998...	X_18	1.086...	64	203.9	42.7	Old H...	112.5	...
X_19	OLD_S...	21202...	1.092...	X_19	1.041...	77	143.8	33.7	Sarco...	112.5	...
X_1	OLD_S...	25372...	0.994...	X_1	NA	67	148.7	41.3	Sarco...	NA	...
X_20	OLD_S...	21469...	1.085...	X_20	1.136...	78	152.3	54.7	Sarco...	148.5	...
X_21	OLD_S...	23263...	1.041...	X_21	0.970...	83	100.6	25.8	Sarco...	67.5	...
X_22	OLD_H...	22763...	1.059...	X_22	1.025...	78	207.2	45.1	Old H...	139.5	...
X_23	OLD_H...	26356...	1.114...	X_23	1	76	236.8	42.4	Old H...	112.5	...
X_24	OLD_H...	30075...	1.185...	X_24	0.745...	65	120.8	23.6	Old H...	103.5	...
...	...	...	...	...	...	...	...	...	...	...	...

Figure 9为图 whether with Sarcopenia 概览。

(对应文件为 **Figure+Table/whether-with-Sarcopenia.pdf**)

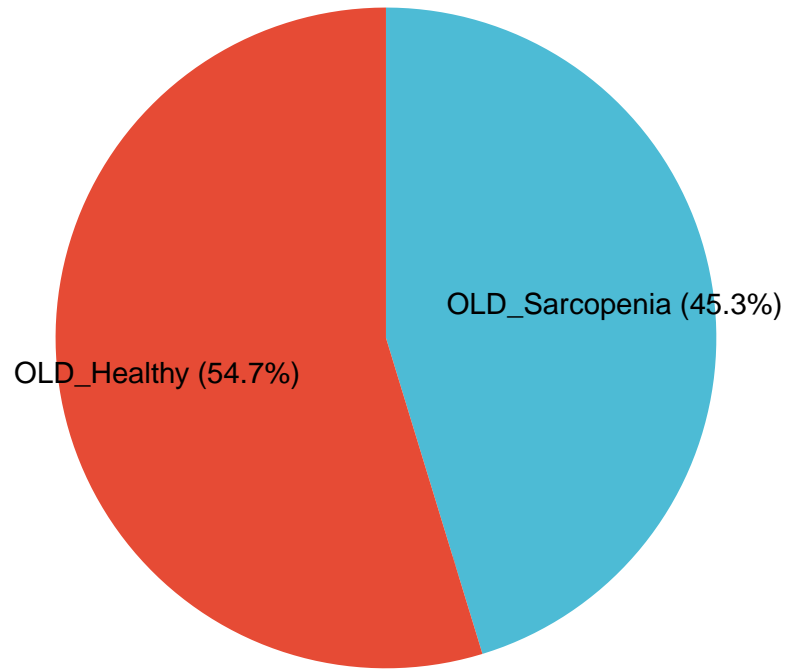


Figure 9: Whether with Sarcopenia

**data\_\_processing :**

Bulk RNA-seq was aligned using the STAR software, and counted using featureCounts.

**data\_\_processing.1 :**

Single-nuclei RNA-seq was aligned using the CellRanger software (10x).

**data\_\_processing.2 :**

Genome\_build: Homo\_sapiens.GRCh38

**data\_\_processing.3 :**

Supplementary\_files\_format\_and\_content: counts.csv: Counts data for bulk RNA-seq.

**data\_\_processing.4 :**

Supplementary\_files\_format\_and\_content: HM\*.csv: Counts for single-nuclei RNA-seq.

Figure 10为图 soft threshold 概览。

(对应文件为 `Figure+Table/soft-threshold.pdf`)

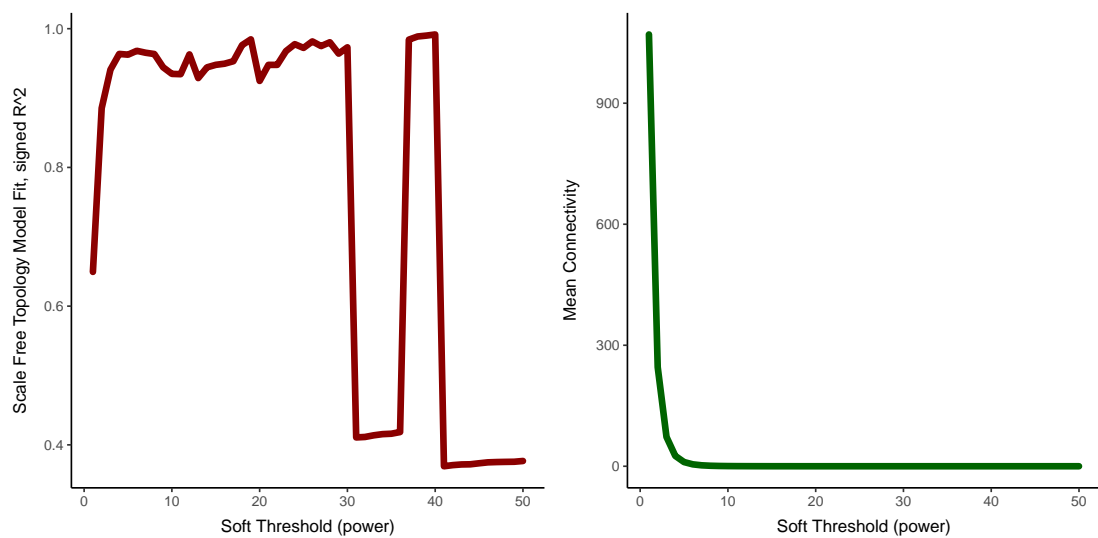


Figure 10: Soft threshold

Figure 11为图 clustering of gene modules 概览。

(对应文件为 `Figure+Table/clustering-of-gene-modules.pdf`)

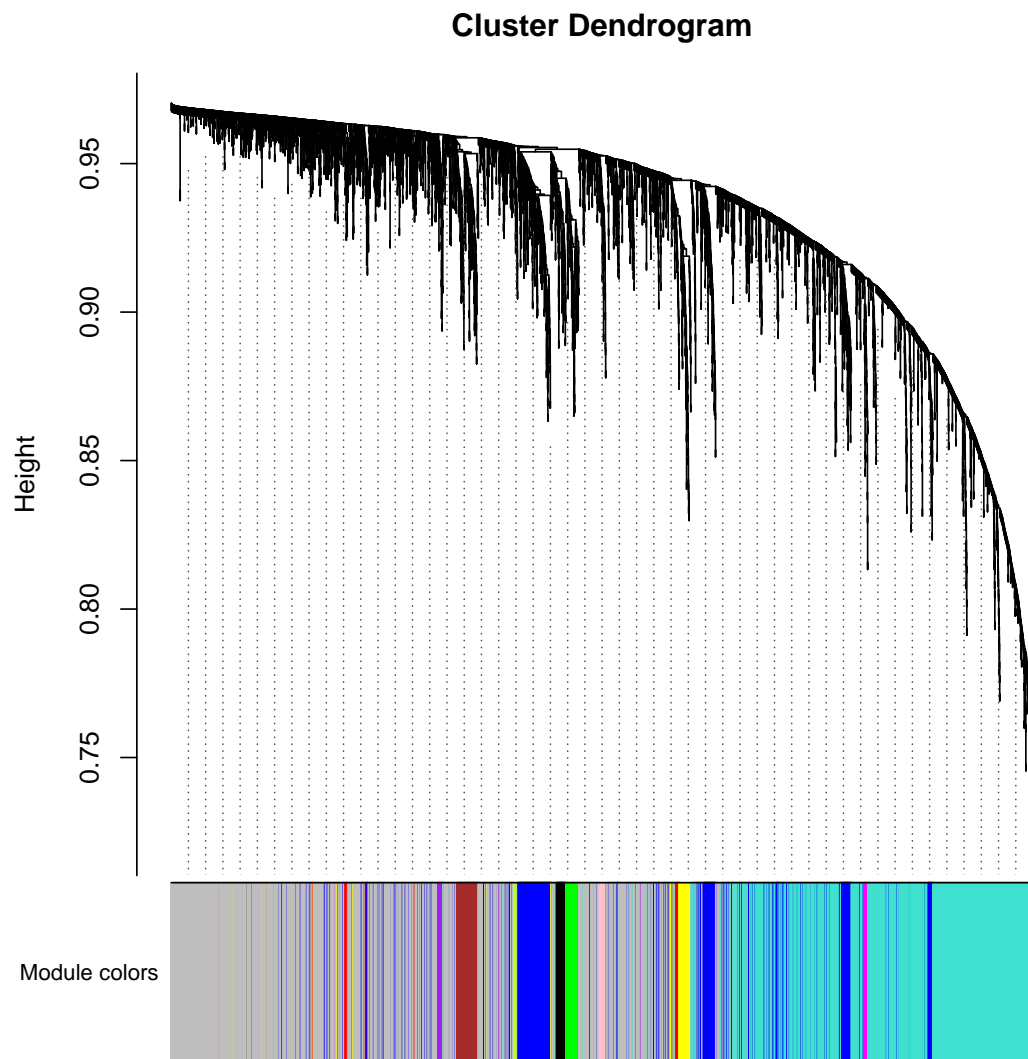


Figure 11: Clustering of gene modules

Figure 12为图 correlation of gene modules and traits data 概览。

(对应文件为 `Figure+Table/correlation-of-gene-modules-and-traits-data.pdf`)



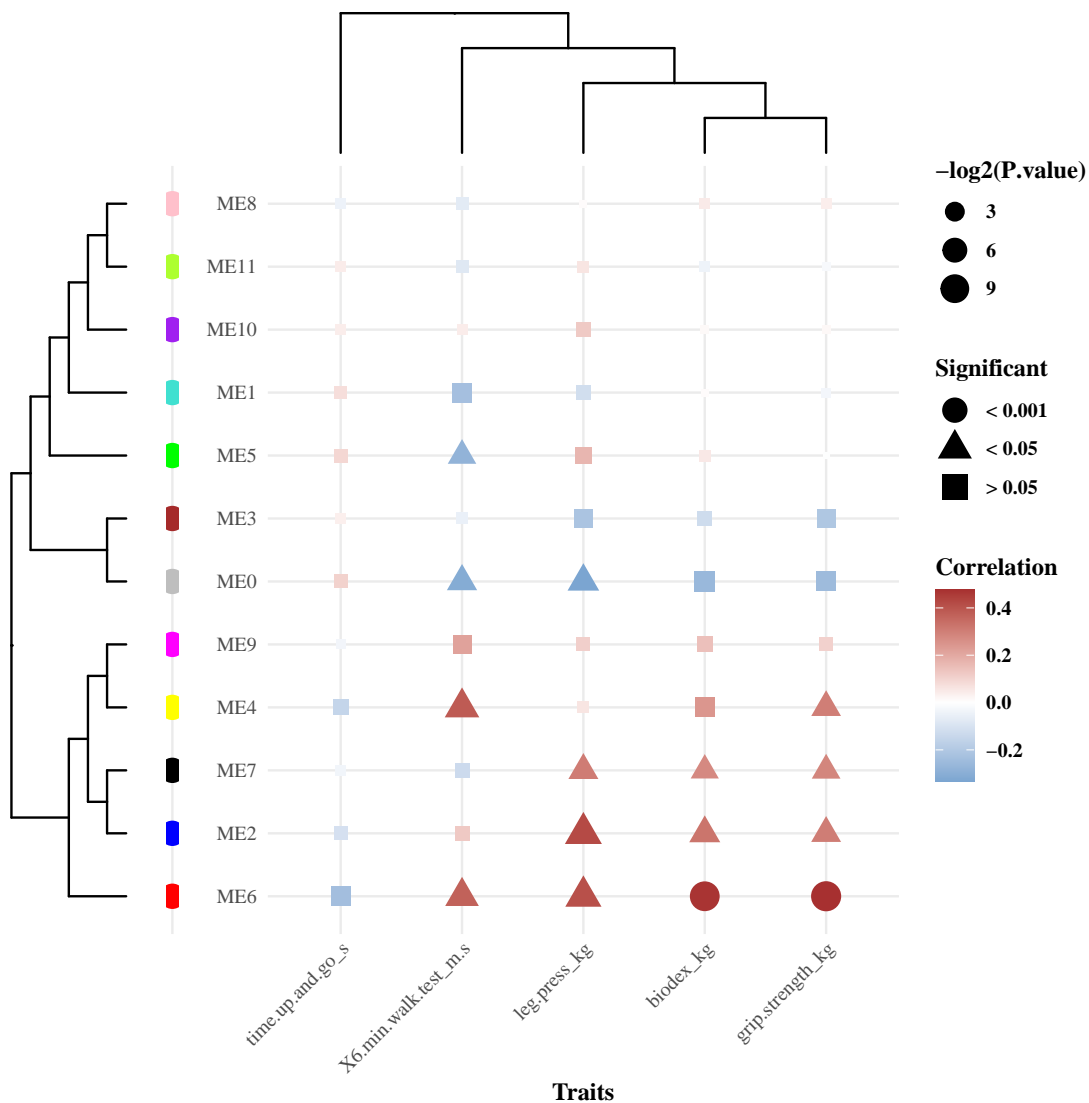


Figure 12: Correlation of gene modules and traits data

Figure 13为图 intersection of genes significant and module memberships 概览。

(对应文件为 Figure+Table/intersection-of-genes-significant-and-module-memberships.pdf)

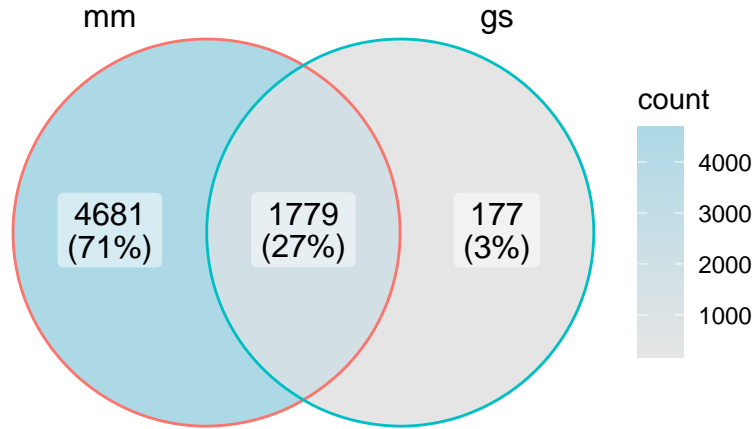


Figure 13: Intersection of genes significant and module memberships

## 7.6 综合：结直肠癌和肌少症

### 7.6.1 （结直肠癌）数据整合

Figure 14为图 all colorectal DEGs 概览。

(对应文件为 Figure+Table/all-colorectal-DEGs.pdf)

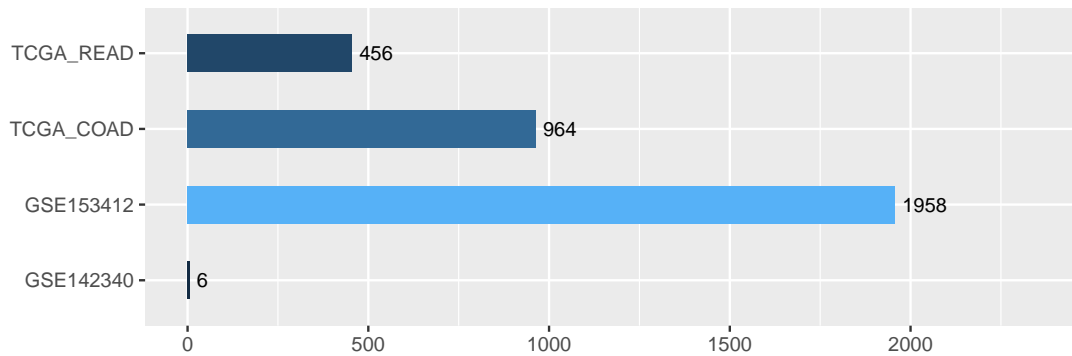


Figure 14: All colorectal DEGs

### 7.6.2 （结直肠癌与肌少症）交集基因

Figure 15为图 intersection of colorectal DEGs with Sarcopenia significant genes 概览。

(对应文件为 Figure+Table/intersection-of-colorectal-DEGs-with-Sarcopenia-significant-genes.pdf)

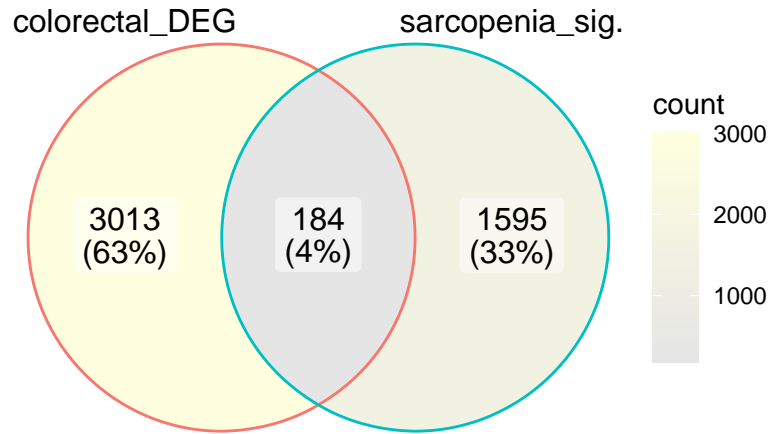


Figure 15: Intersection of colorectal DEGs with Sarcopenia significant genes

### 7.6.3 富集分析

The Cellular Component (CC), the Molecular Function (MF) and the Biological Process (BP).

Figure 16为图 go enrichment 概览。

(对应文件为 **Figure+Table/go-enrichment.pdf**)

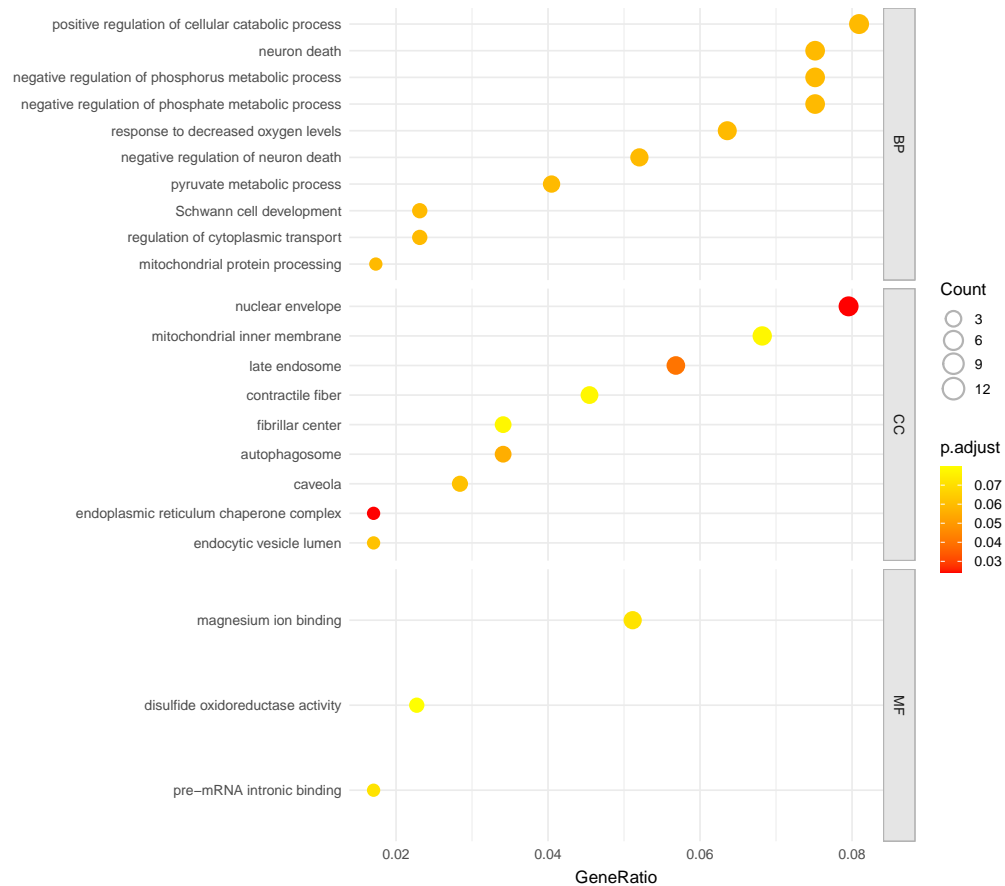


Figure 16: Go enrichment

Figure 17为图 kegg enrichment 概览。

(对应文件为 `Figure+Table/kegg-enrichment.pdf`)

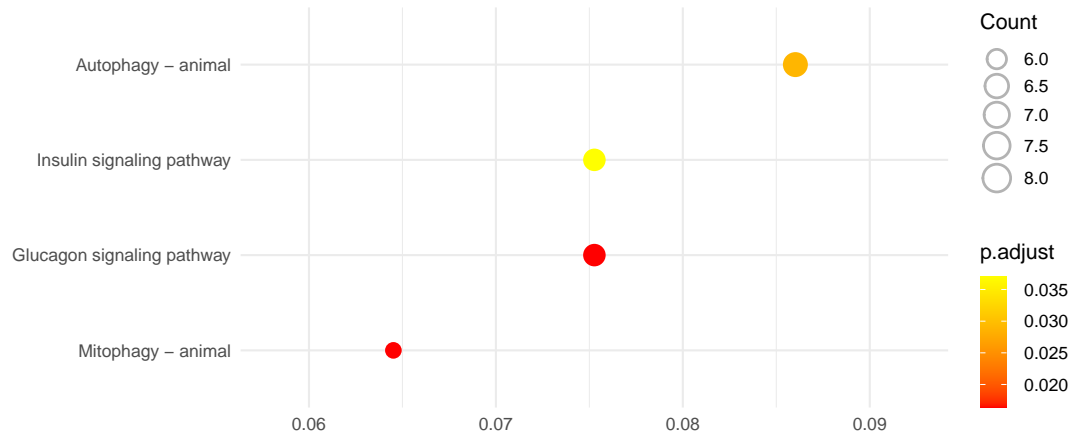


Figure 17: Kegg enrichment

#### 7.6.4 通路可视化

Figure 18为图 hits in autophagy 概览。

(对应文件为 `Figure+Table/hsa04140.pathview.png`)

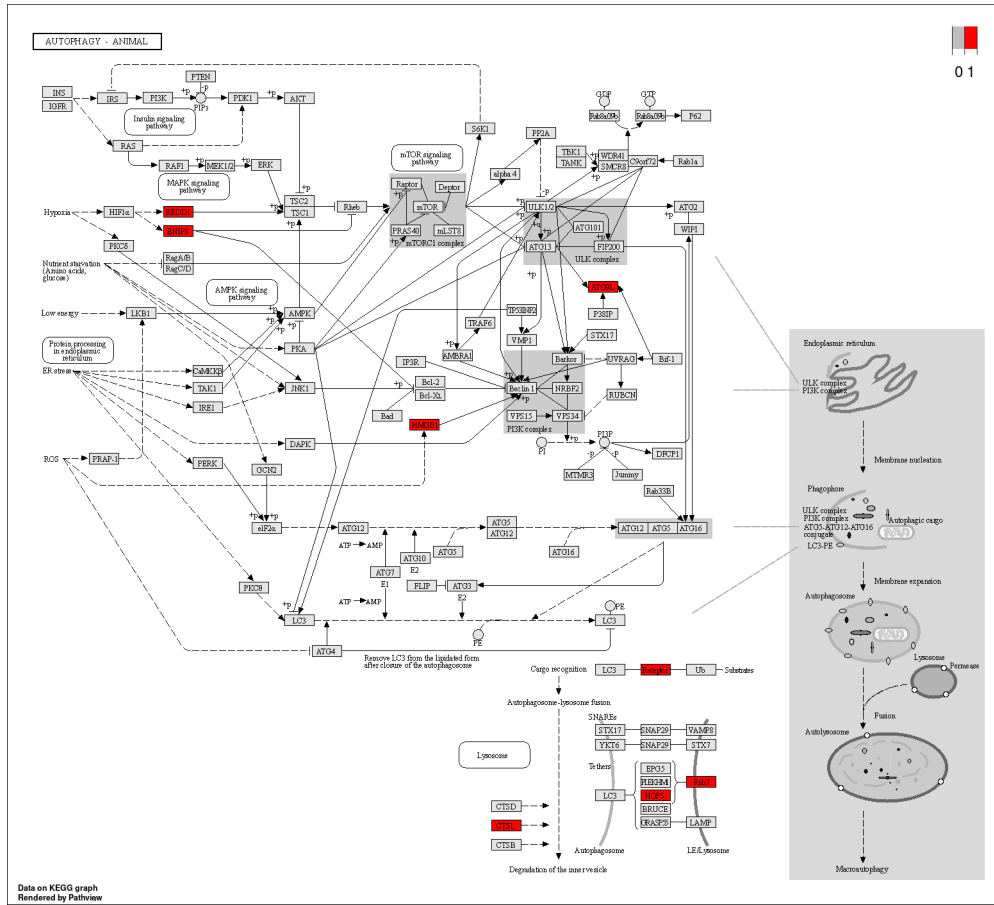


Figure 18: Hits in autophagy

Figure 19为图 hits in Mitophagy 概览。

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

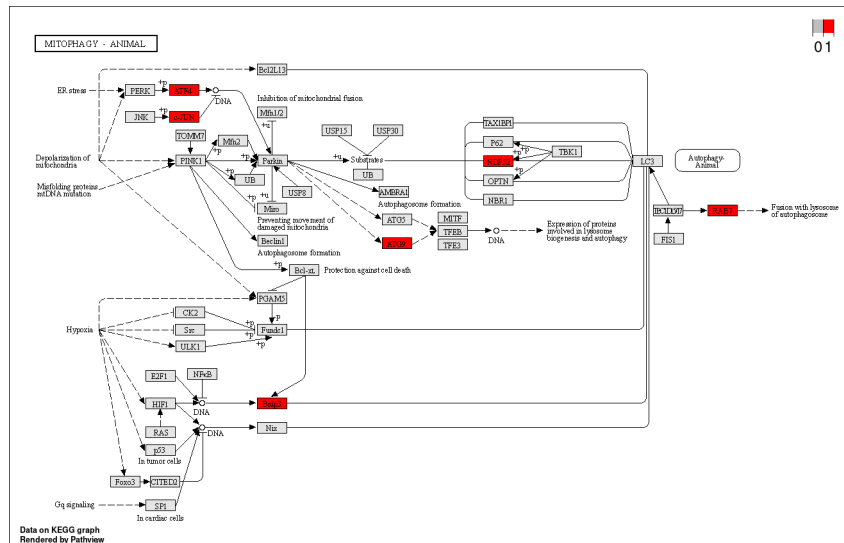


Figure 19: Hits in Mitophagy

## Reference

1. Onoki, T. *et al.* Skeletal muscle-specific keap1 disruption modulates fatty acid utilization and enhances exercise capacity in female mice. *Redox biology* **43**, (2021).
2. Kudryavtseva, A. V. *et al.* Mitochondrial dysfunction and oxidative stress in aging and cancer. *Oncotarget* **7**, 44879–44905 (2016).
3. Vargas, T. R. & Apetoh, L. Danger signals: Chemotherapy enhancers? *Immunological Reviews* **280**, (2017).
4. Ohishi, T. *et al.* Current targeted therapy for metastatic colorectal cancer. *International Journal of Molecular Sciences* **24**, (2023).
5. Williams, G. R., Dunne, R. F., Giri, S., Shachar, S. S. & Caan, B. J. Sarcopenia in the older adult with cancer. *Journal of Clinical Oncology* **39**, (2021).
6. Havas, A., Yin, S. & Adams, P. D. The role of aging in cancer. *Molecular Oncology* **16**, (2022).