Analysis

Huang LiChuang of Wie-Biotech

Contents

List of Figures

List of Tables

1 摘要

分析客户数据差异基因，查看是否存在胆固醇膜受体 GPBAR/TGR5 及其相关信号通路。结果见 3。

- 胆固醇结石 (Cholesterol gallstones, CGS)[1]

- 成岩饮食 (LD) 诱导的 CGS 动物模型
- 猪胆酸 (Hyocholic acid, HCA)
- TGR5 (GPBAR)[2]

2 材料和方法

2.1 方法

Mainly used method:

- Package biomaRt used for gene annotation.[3]
- ClusterProfiler used for gene enrichment analysis.[4]
- Fastp used for Fastq data preprocessing.[5]
- ClusterProfiler used for GSEA enrichment .
- Kallisto used for RNA-seq mapping and quantification.[6]
- Limma and edgeR used for differential expression analysis.[7,8]
- Other R packages used for statistic analysis or data visualization.

3 分析结果

- 对原始数据的处理分质控和比对参考基因组，结果见 5.1 和对应文件。
- 肝脏组织和回肠组织的数据分别进行了差异分析和通路富集分析 (Fig. 1 和 Fig. 3)
- 肝脏组织：
  - Model vs Control (Fig. 1a, d, e), 'Steroid biosynthesis' 为首要富集通路，且总体表达呈现下调趋势
  - Treatment vs Model (Fig. 1b, f, g), 'Steroid biosynthesis' 同样为首要富集通路，总体表达呈现上调趋势
  - 结合两组对比，药物干预改变了 'Steroid biosynthesis' 通路基因的表达水平 (使其上调)。药物的治疗作用可能和其下游 'steroid degradation' 有关 (Fig. 2a)。
  - 将药物有调控的基因 (详情见 5.4.3) 进行富集分析 (Fig. 2b, c) ，发现富集于 'Steroid hormone biosynthesis' (KEGG), 'fatty acid metabolic process' (GO), 'steroid metabolic process' (GO)。
- 回肠组织完成了类似上述肝脏组织的分析模式。
- 关于 TGR5 (TGR5 也就是 Gpbar1)，由于该基因表达量过低 (可能是该基因本身表达量低，且测序深度不足)，因此难以判断 TGR5 的显著性 (尤其是肝脏) (在 5.6 做了 TGR5 和另一个相关蛋白 S1pr2 的讨论)。

Figure 1 (下方图) 为图 Main fig1 概览。

(对应文件为 `./Figure+Table/fig1.pdf`)

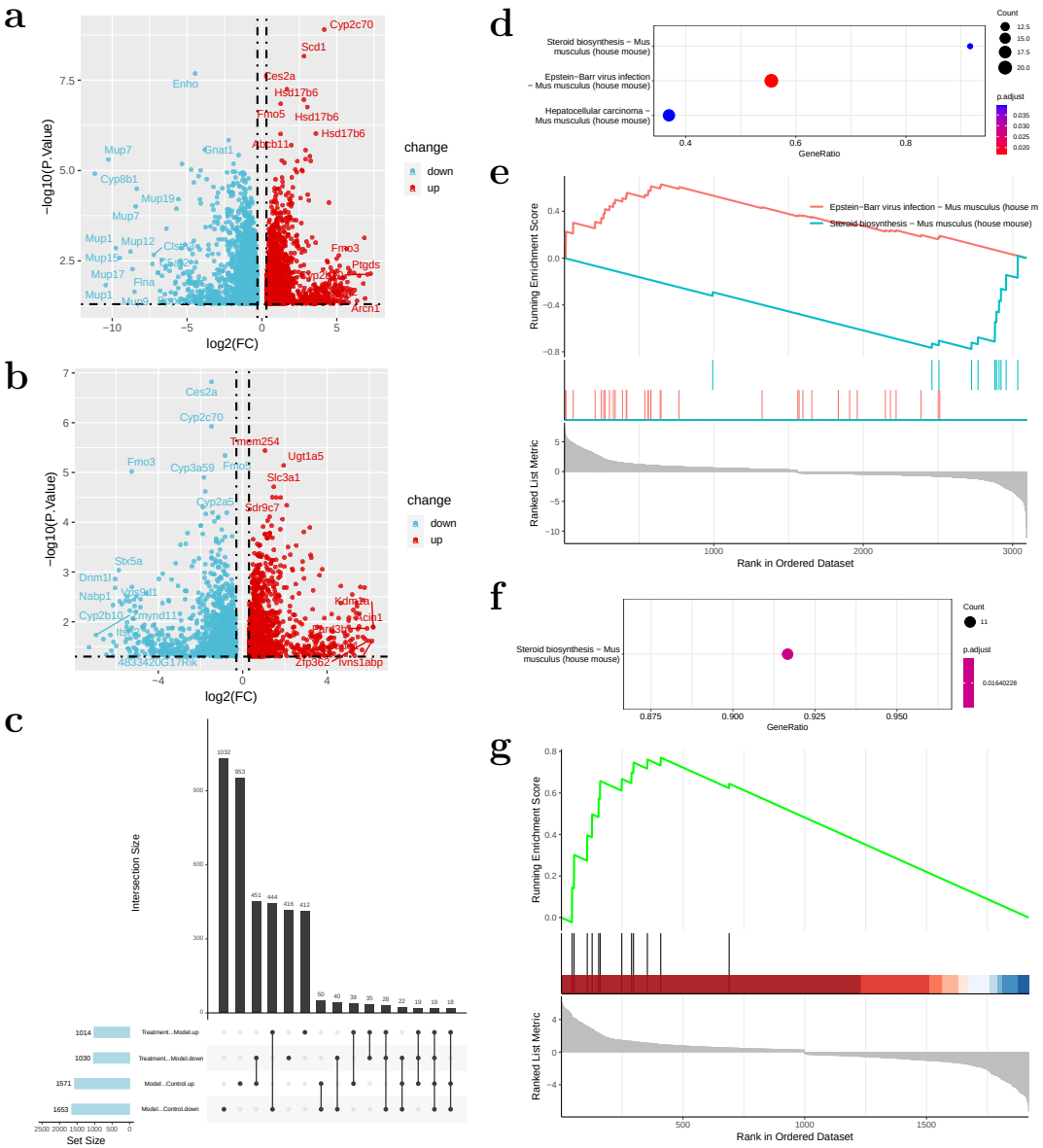

Figure 1: Main fig1

Figure 2 (下方图) 为图 Main fig2 概览。

(对应文件为 `./Figure+Table/fig2.pdf`)

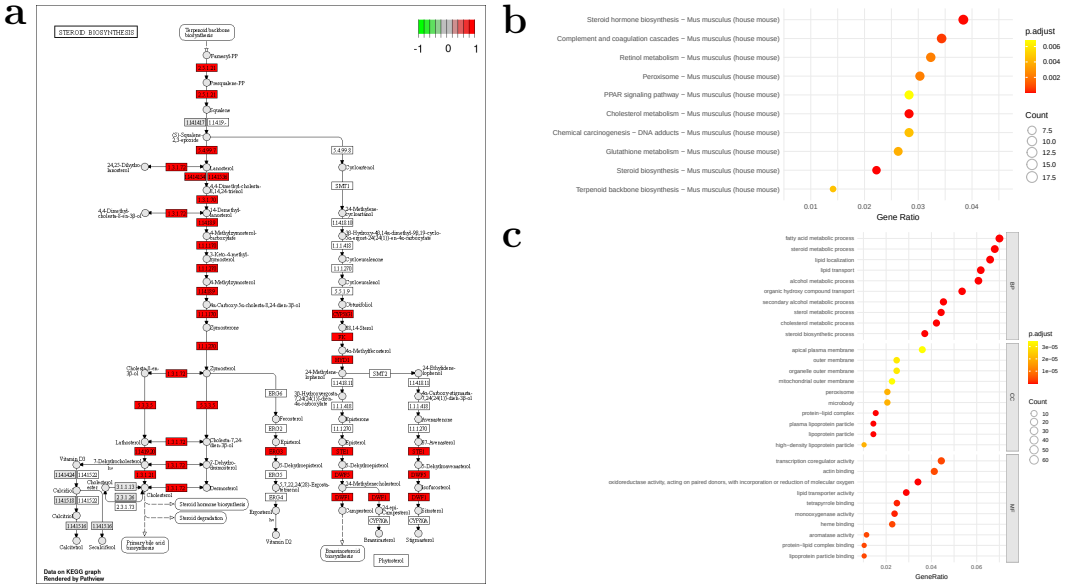

Figure 2: Main fig2

Figure 3 (下方图) 为图 Main fig3 概览。

(对应文件为 `./Figure+Table/fig3.pdf`)

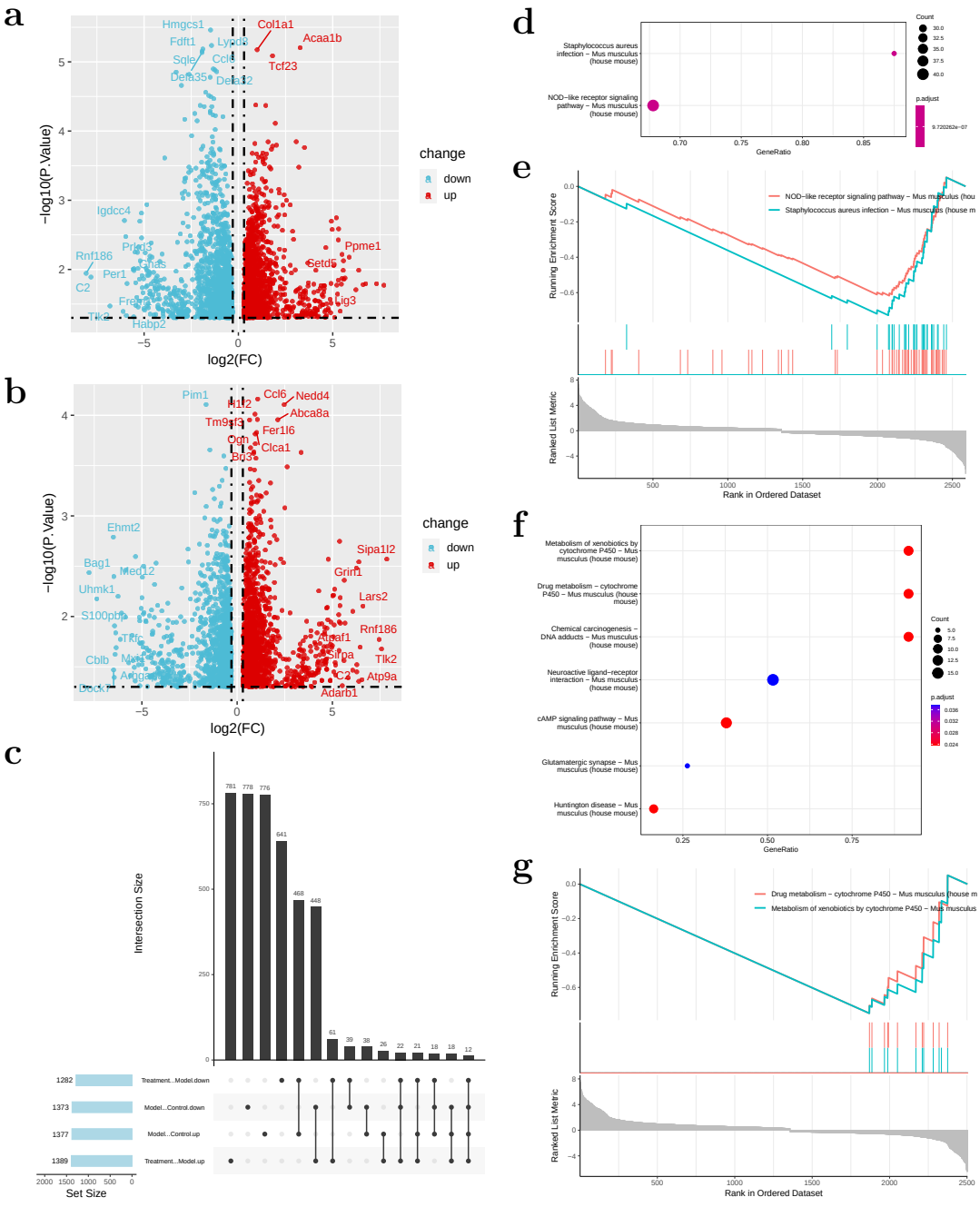

Figure 3: Main fig3

Figure 4 (下方图) 为图 Main fig4 概览。

(对应文件为 `./Figure+Table/fig4.pdf`)

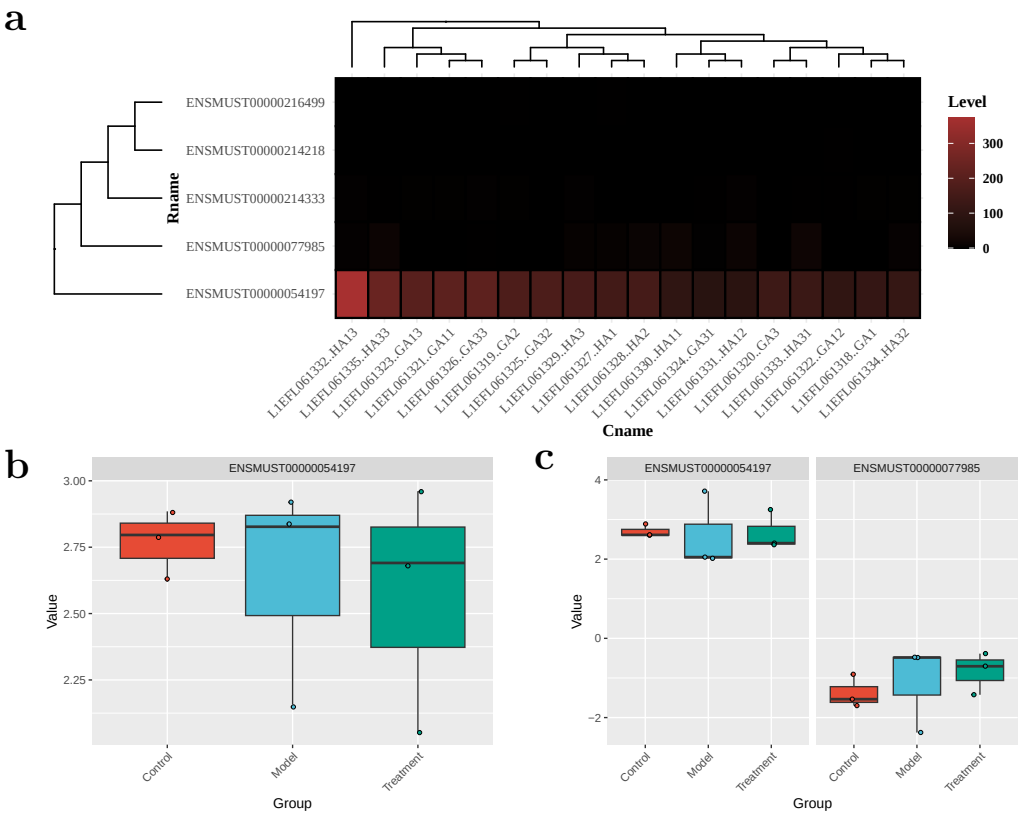

Figure 4: Main fig4

4 结论

5 附：分析流程

5.1 RNA 定量

5.1.1 FastQ 文件质控

'Fastp QC reports' 数据已全部提供。

(对应文件为 `./fastp_report/`)

注：文件夹./fastp_report/共包含 18 个文件。

1. L1EFL061318–GA1..html
2. L1EFL061319–GA2..html
3. L1EFL061320–GA3..html
4. L1EFL061321–GA11..html
5. L1EFL061322–GA12..html
6. ...

5.1.2 获取参考基因注释

下载 cDNA 参考基因注释 (使用的是 mus musculus 的参考基因) 。https://ftp.ensembl.org/pub/release-110/fasta/mus_musculus/

5.1.3 使用 Kallisto 定量

所有样本的 RNA 表达定量数据已提供。

'Kallisto quantification' 数据已全部提供。

(对应文件为 `./kallisto_quantification`)

注：文件夹./kallisto_quantification 共包含 18 个文件。

1. L1EFL061318–GA1
2. L1EFL061319–GA2
3. L1EFL061320–GA3
4. L1EFL061321–GA11
5. L1EFL061322–GA12
6. ...

5.2 样本元数据

根据客户提供的样本信息，整理的元数据：

Table 1 (下方表格) 为表格 metadata of all samples 概览。

(对应文件为 `Figure+Table/metadata-of-all-samples.csv`)

注: 表格共有 18 行 6 列, 以下预览的表格可能省略部分数据; 表格含有 18 个唯一 'sample'。

Table 1: Metadata of all samples

sample	group	file	direc...	ident	tissue
L1EFL...	Control	kalli...	kalli...	GA1	Liver
L1EFL...	Control	kalli...	kalli...	GA2	Liver
L1EFL...	Control	kalli...	kalli...	GA3	Liver

sample	group	file	direc…	ident	tissue
L1EFL…	Model	kalli…	kalli…	GA11	Liver
L1EFL…	Model	kalli…	kalli…	GA12	Liver
L1EFL…	Model	kalli…	kalli…	GA13	Liver
L1EFL…	Treat…	kalli…	kalli…	GA31	Liver
L1EFL…	Treat…	kalli…	kalli…	GA32	Liver
L1EFL…	Treat…	kalli…	kalli…	GA33	Liver
L1EFL…	Control	kalli…	kalli…	HA1	Theileum
L1EFL…	Control	kalli…	kalli…	HA2	Theileum
L1EFL…	Control	kalli…	kalli…	HA3	Theileum
L1EFL…	Model	kalli…	kalli…	HA11	Theileum
L1EFL…	Model	kalli…	kalli…	HA12	Theileum
L1EFL…	Model	kalli…	kalli…	HA13	Theileum
…	…	…	…	…	…

5.3 RNA 注释

5.3.1 使用 Biomart 注释

由于实验数据来源于小暑，因此首要关注的基因名称是 `mgi_symbol`

Table 2 (下方表格) 为表格 Gene annotation 概览。

(对应文件为 `Figure+Table/Gene-annotation.tsv`)

> 注：表格共有 116873 行 5 列，以下预览的表格可能省略部分数据；表格含有 35815 个唯一 'mgi_symbol'。

Table 2: Gene annotation

ensem…	mgi_s…	entre…	hgnc_…	descr…
ENSMU…	mt-Nd1	17716	NA	mitoc…
ENSMU…	mt-Nd2	17717	NA	mitoc…
ENSMU…	mt-Co1	17708	NA	mitoc…
ENSMU…	mt-Co2	17709	NA	mitoc…
ENSMU…	mt-Atp8	17706	NA	mitoc…
ENSMU…	mt-Atp6	17705	NA	mitoc…
ENSMU…	mt-Co3	17710	NA	mitoc…
ENSMU…	mt-Nd3	17718	NA	mitoc…
ENSMU…	mt-Nd4l	17720	NA	mitoc…
ENSMU…	mt-Nd4	17719	NA	mitoc…
ENSMU…	mt-Nd5	17721	NA	mitoc…
ENSMU…	mt-Nd6	17722	NA	mitoc…

ensem...	mgi_s...	entre...	hgnc_...	descr...
ENSMU...	mt-Cytb	17711	NA	mitoc...
ENSMU...		NA	NA	
ENSMU...		671917	NA	
...	...	...	...	...

5.4 肝脏组织 (Liver)

5.4.1 差异分析

Figure 5 (下方图) 为图 Liver Model vs Control 概览。

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

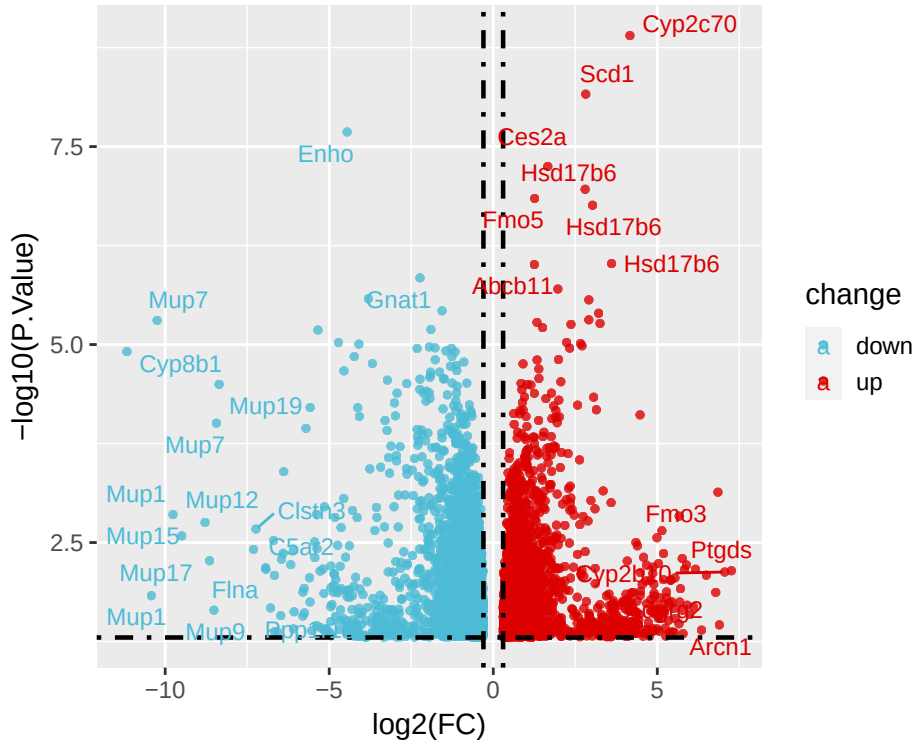

Figure 5: Liver Model vs Control

Figure 6 (下方图) 为图 Liver Treatment vs Model 概览。

(对应文件为 `Figure+Table/Liver-Treatment-vs-Model.pdf`)

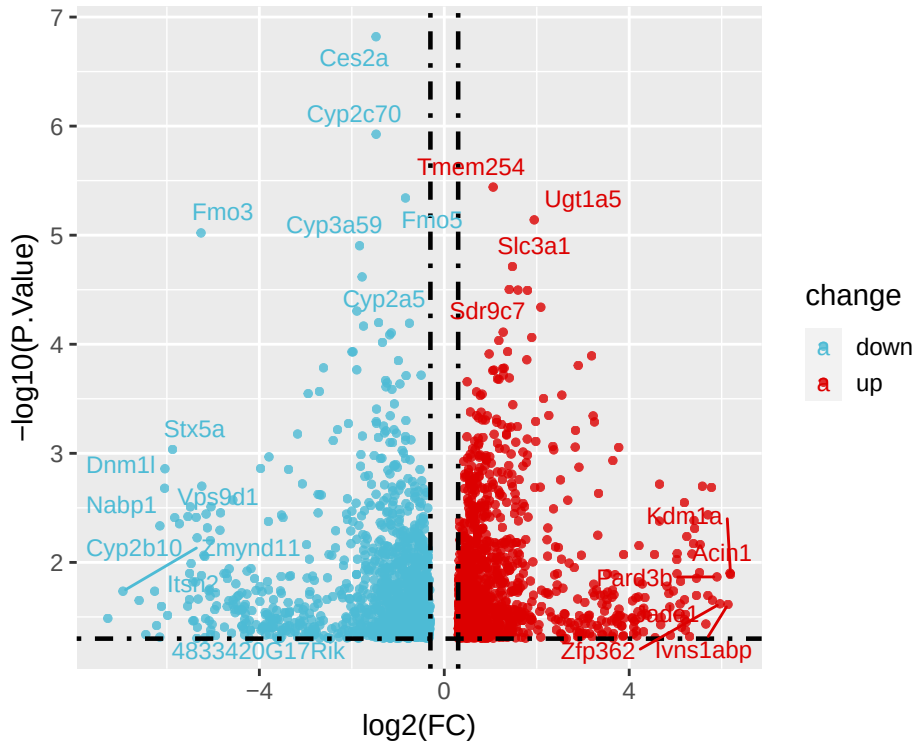

Figure 6: Liver Treatment vs Model

Figure 7 (下方图) 为图 Liver DEGs intersection 概览。

(对应文件为 `Figure+Table/Liver-DEGs-intersection.pdf`)

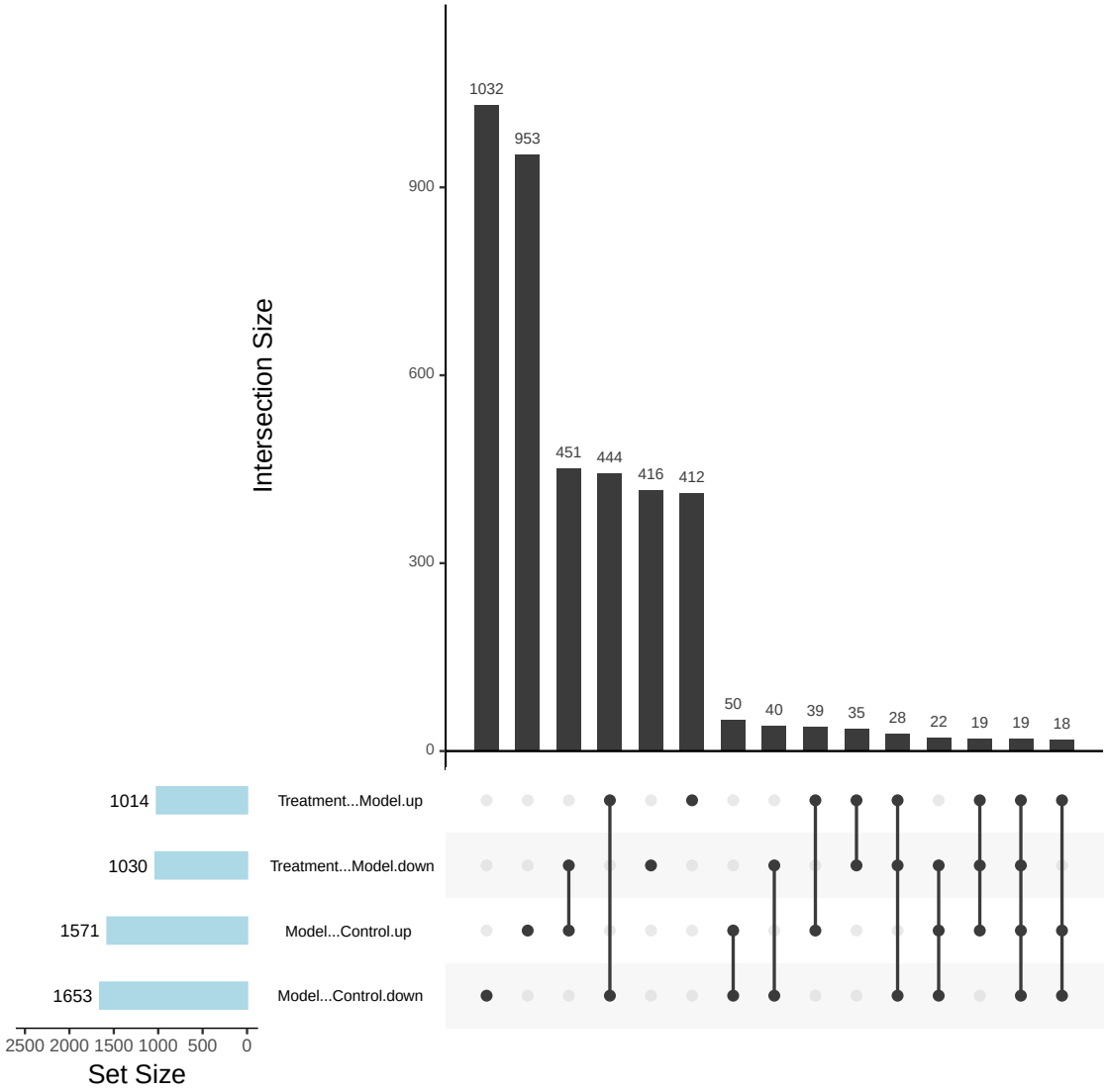

Figure 7: Liver DEGs intersection

'Data of DEGs of Liver' 数据已全部提供。

(对应文件为 `Figure+Table/data-of-DEGs-of-Liver`)

> 注：文件夹 Figure+Table/data-of-DEGs-of-Liver 共包含 2 个文件。
>
> 1. 1_Model - Control.csv
> 2. 2_Treatment - Model.csv

5.4.2 GSEA 富集分析

5.4.2.1 Model vs Control Figure 8 (下方图) 为图 Liver KEGG Model vs Control 概览。

(对应文件为 `Figure+Table/Liver-KEGG-Model-vs-Control.pdf`)

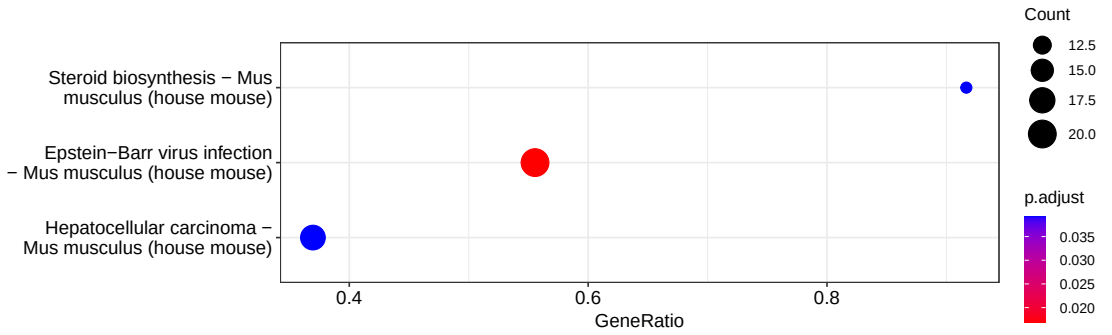

Figure 8: Liver KEGG Model vs Control

Figure 9 (下方图) 为图 Liver GSEA plot of pathways Model vs Control 概览。

(对应文件为 `Figure+Table/Liver-GSEA-plot-of-pathways-Model-vs-Control.pdf`)

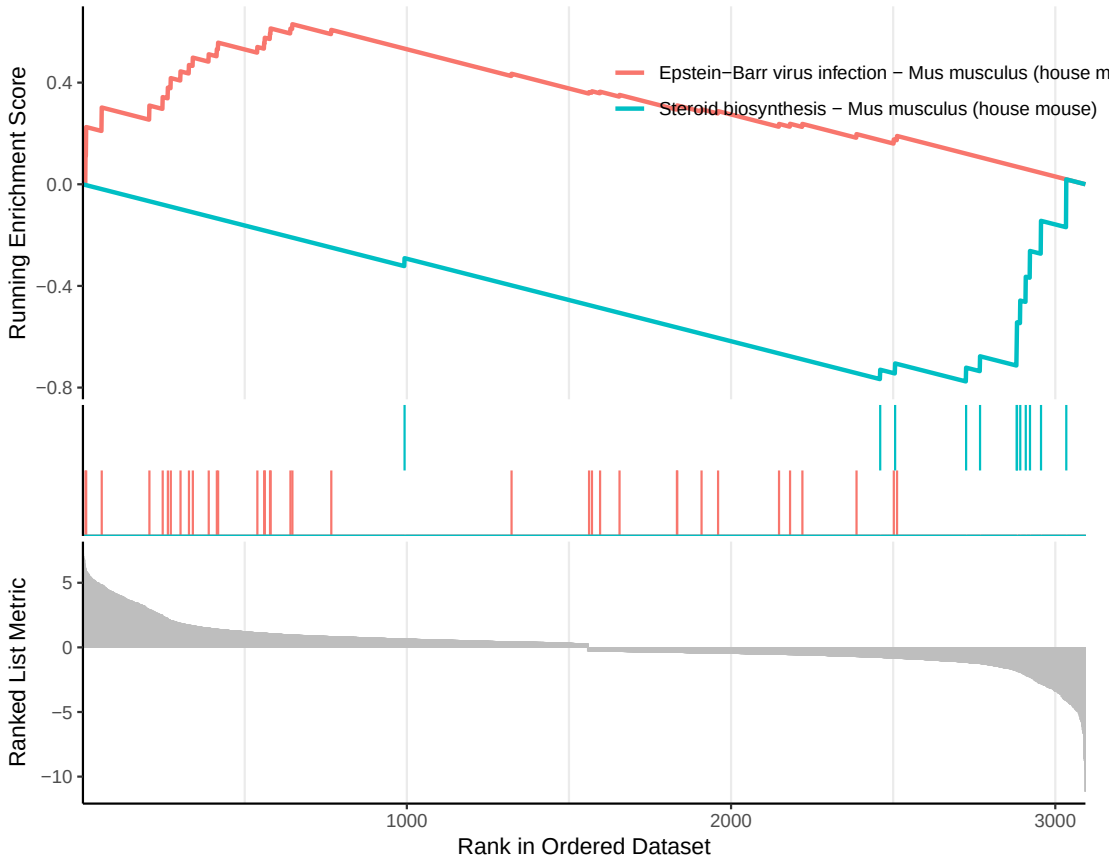

Figure 9: Liver GSEA plot of pathways Model vs Control

5.4.2.2 Treatment vs Model Figure 10 (下方图) 为图 Liver KEGG Treatment vs Model 概览。

(对应文件为 `Figure+Table/Liver-KEGG-Treatment-vs-Model.pdf`)

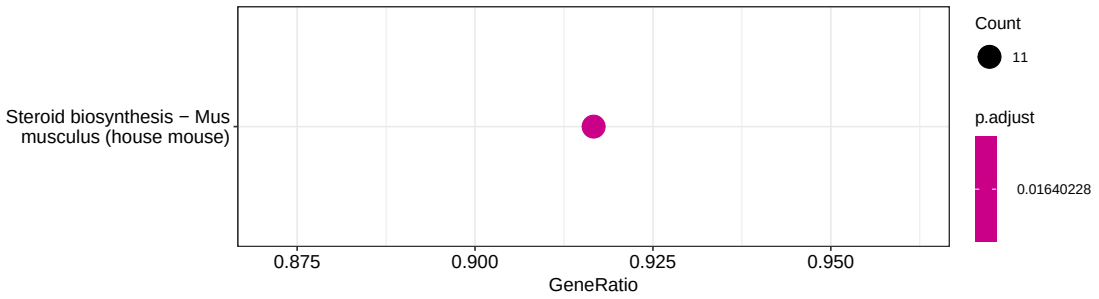

Figure 10: Liver KEGG Treatment vs Model

Figure 11 (下方图) 为图 Liver GSEA plot of pathways Treatment vs Model 概览。

(对应文件为 `Figure+Table/Liver-GSEA-plot-of-pathways-Treatment-vs-Model.pdf`)

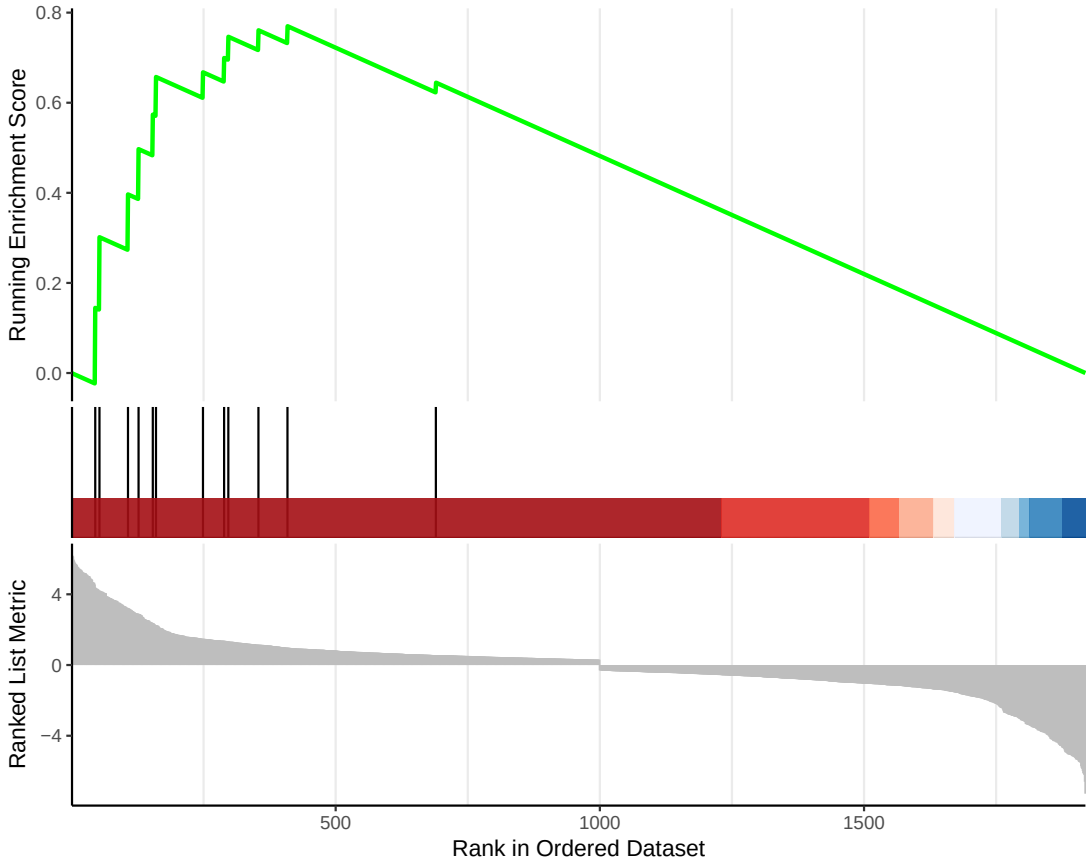

Figure 11: Liver GSEA plot of pathways Treatment vs Model

Figure 12 (下方图) 为图 Liver pathway of Steroid biosynthesis Treatment vs Model 概览。

(对应文件为 `Figure+Table/mmu00100.pathview.png`)

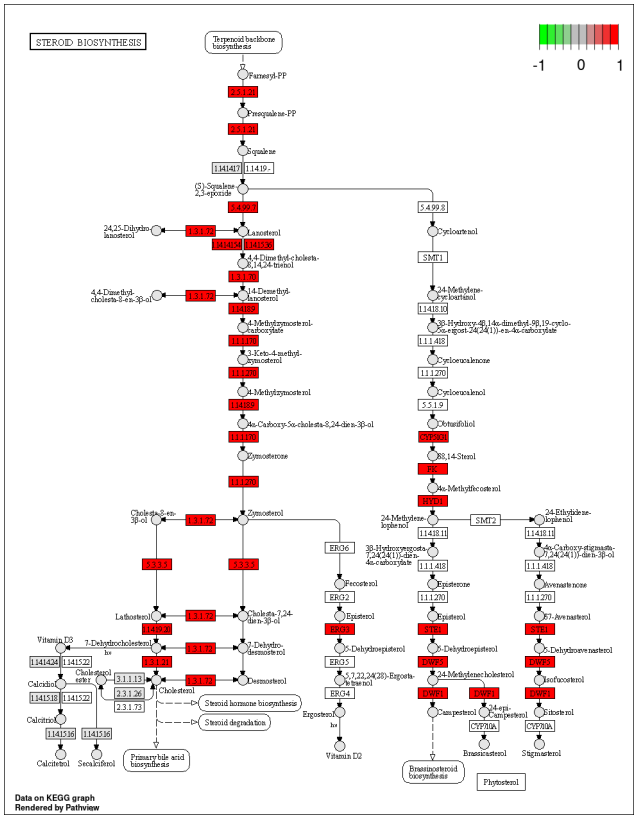

Figure 12: Liver pathway of Steroid biosynthesis Treatment vs Model

5.4.3 Classical 富集分析

将有药物调控作用的 DEGs 以传统的富集方式分析。

- 模型组相比对照组，基因上调；而以药物处理后，基因下调（相比于模型组）。
- 模型组相比对照组，基因下调；而以药物处理后，基因上调（相比于模型组）。

Figure 13 (下方图) 为图 Liver drug regulated DEGs KEGG enrichment 概览。

(对应文件为 `Figure+Table/Liver-drug-regulated-DEGs-KEGG-enrichment.pdf`)

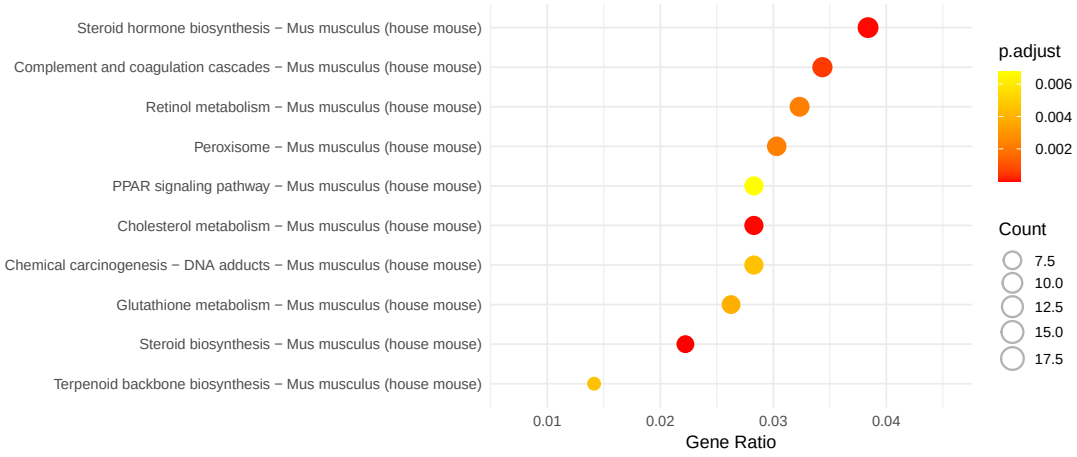

Figure 13: Liver drug regulated DEGs KEGG enrichment

Figure 14 (下方图) 为图 Liver drug regulated DEGs GO enrichment 概览。

(对应文件为 `Figure+Table/Liver-drug-regulated-DEGs-GO-enrichment.pdf`)

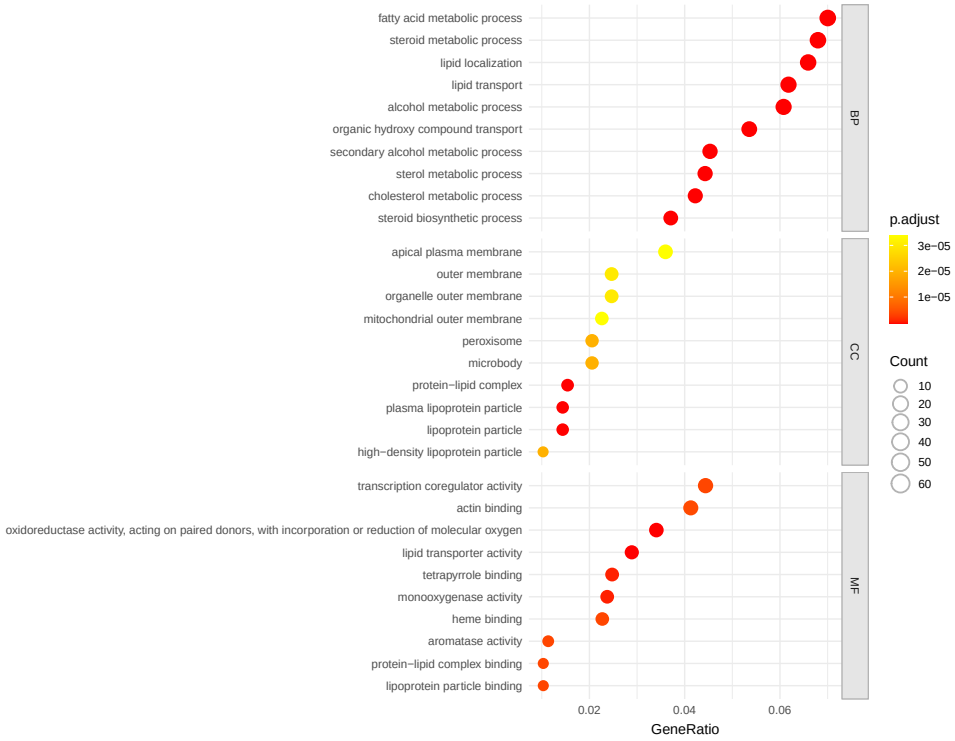

Figure 14: Liver drug regulated DEGs GO enrichment

5.5 回肠组织 (Theileum)

5.5.1 差异分析

Figure 15 (下方图) 为图 Theileum Model vs Control 概览。

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

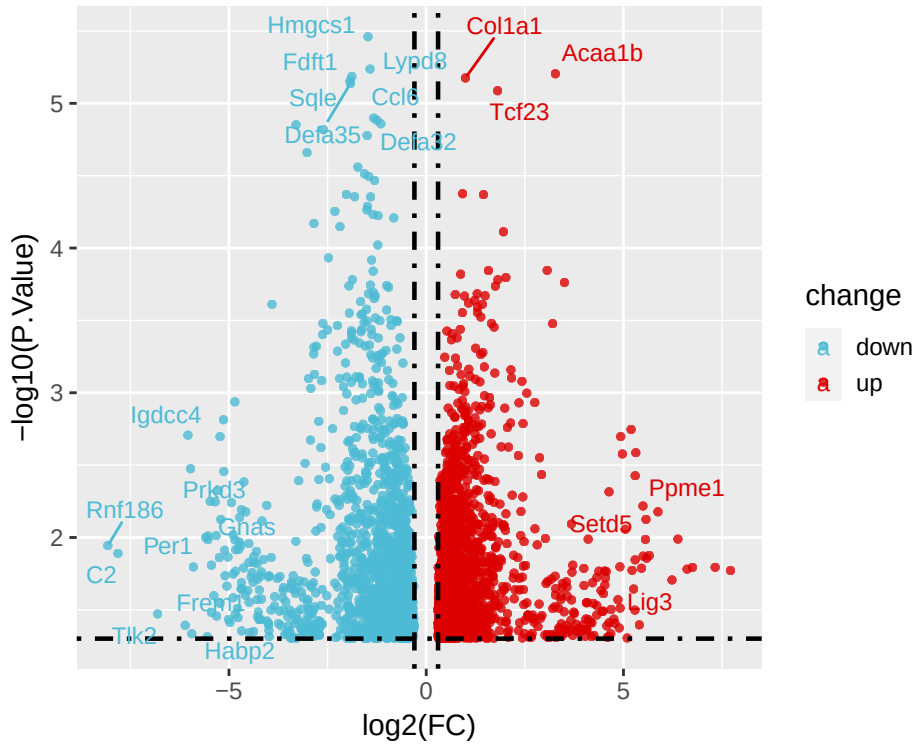

Figure 15: Theileum Model vs Control

Figure 16 (下方图) 为图 Theileum Treatment vs Model 概览。

(对应文件为 `Figure+Table/Theileum-Treatment-vs-Model.pdf`)

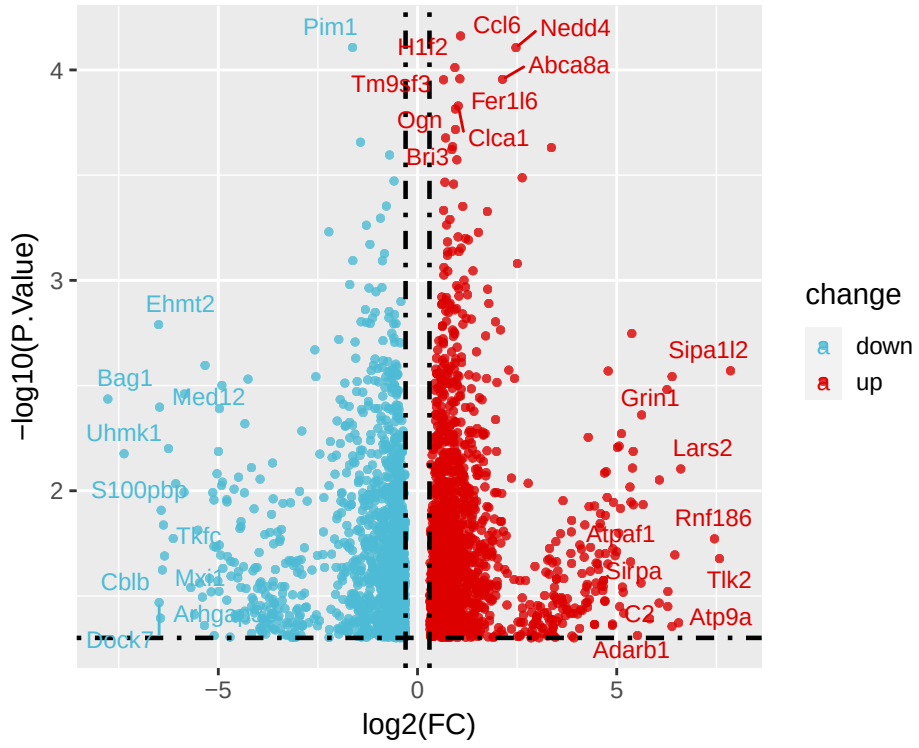

Figure 16: Theileum Treatment vs Model

Figure 17 (下方图) 为图 Theileum DEGs intersection 概览。

(对应文件为 `Figure+Table/Theileum-DEGs-intersection.pdf`)

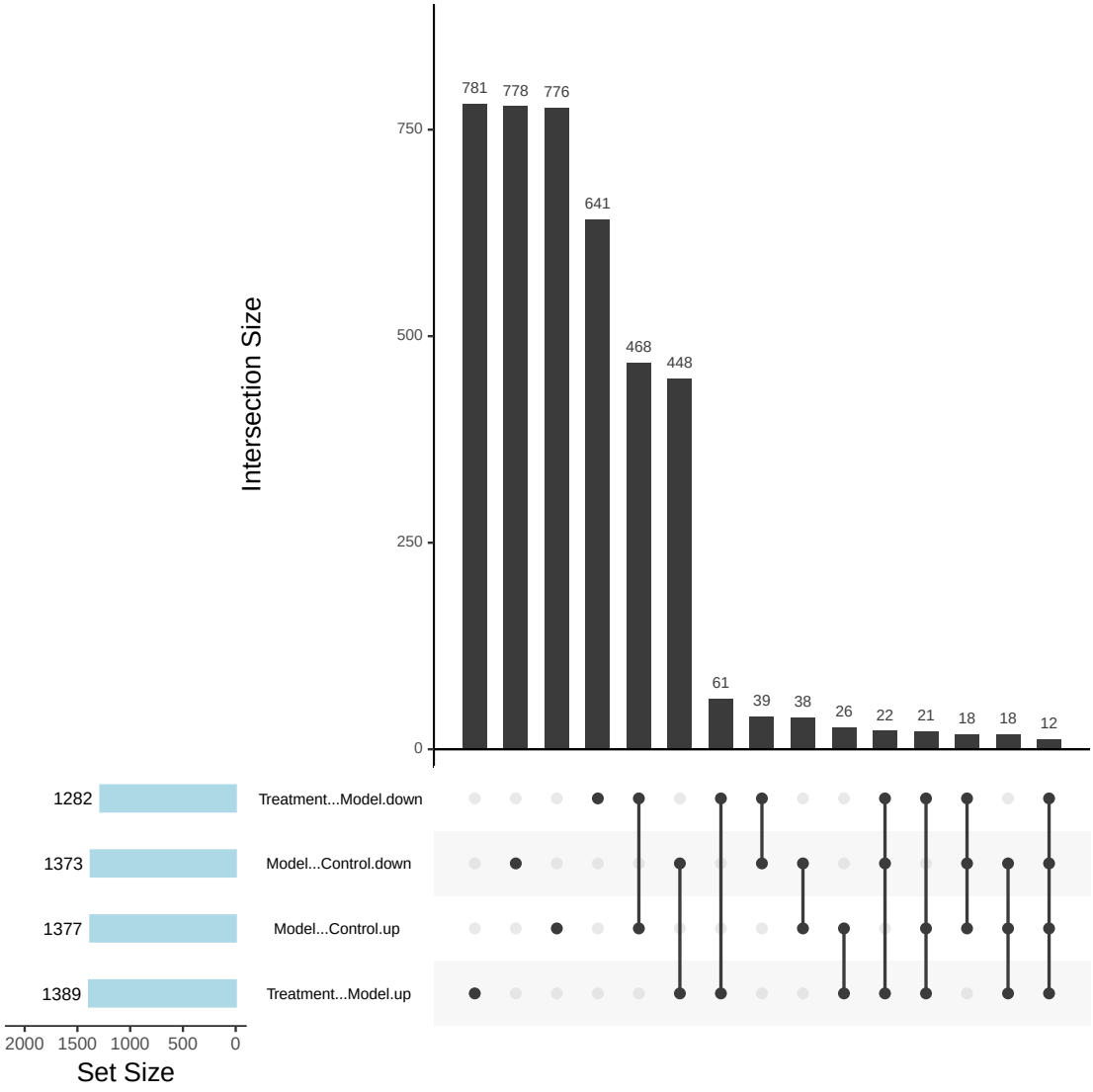

Figure 17: Theileum DEGs intersection

'Data of DEGs of Theileum' 数据已全部提供。

(对应文件为 `Figure+Table/data-of-DEGs-of-Theileum`)

> 注：文件夹 Figure+Table/data-of-DEGs-of-Theileum 共包含 2 个文件。
>
> 1. 1_Model - Control.csv
> 2. 2_Treatment - Model.csv

5.5.2 GSEA 富集分析

5.5.2.1 Model vs Control Figure 18 (下方图) 为图 Theileum KEGG Model vs Control 概览。

(对应文件为 `Figure+Table/Theileum-KEGG-Model-vs-Control.pdf`)

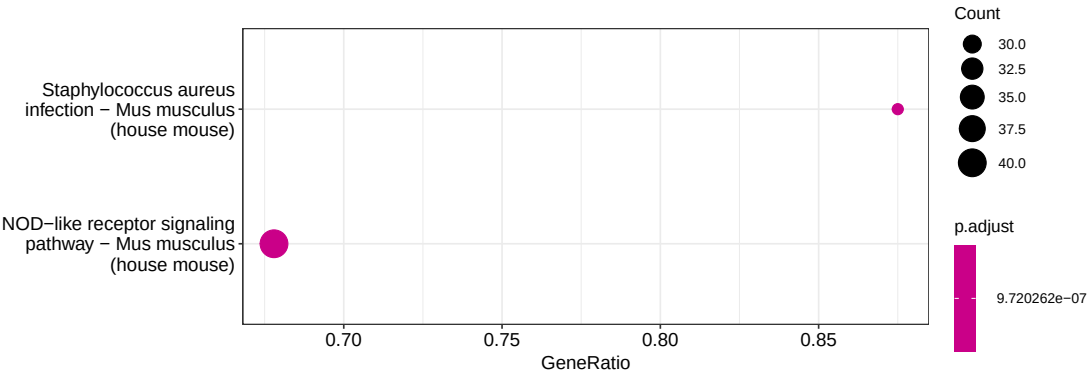

Figure 18: Theileum KEGG Model vs Control

Figure 19 (下方图) 为图 Theileum GSEA plot of pathways Model vs Control 概览。

(对应文件为 `Figure+Table/Theileum-GSEA-plot-of-pathways-Model-vs-Control.pdf`)

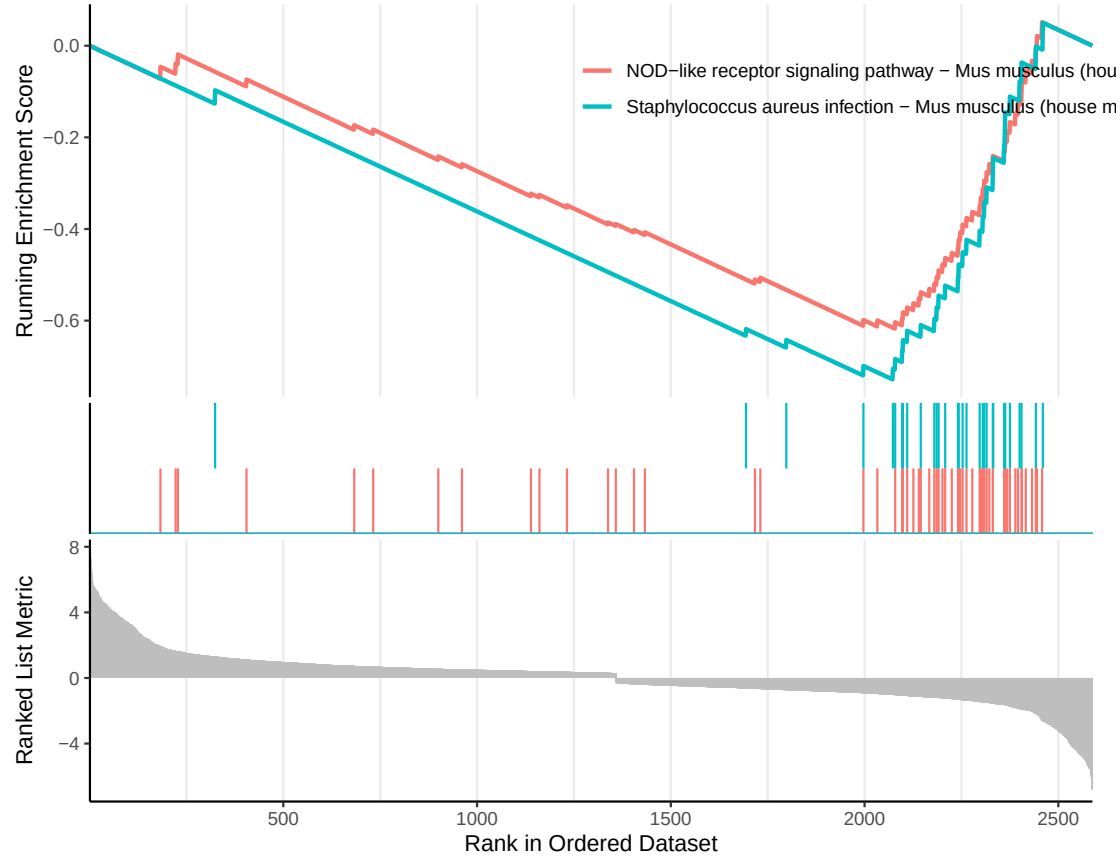

Figure 19: Theileum GSEA plot of pathways Model vs Control

5.5.2.2 Treatment vs Model

Figure 20 (下方图) 为图 Theileum KEGG Treatment vs Model 概览。

(对应文件为 `Figure+Table/Theileum-KEGG-Treatment-vs-Model.pdf`)

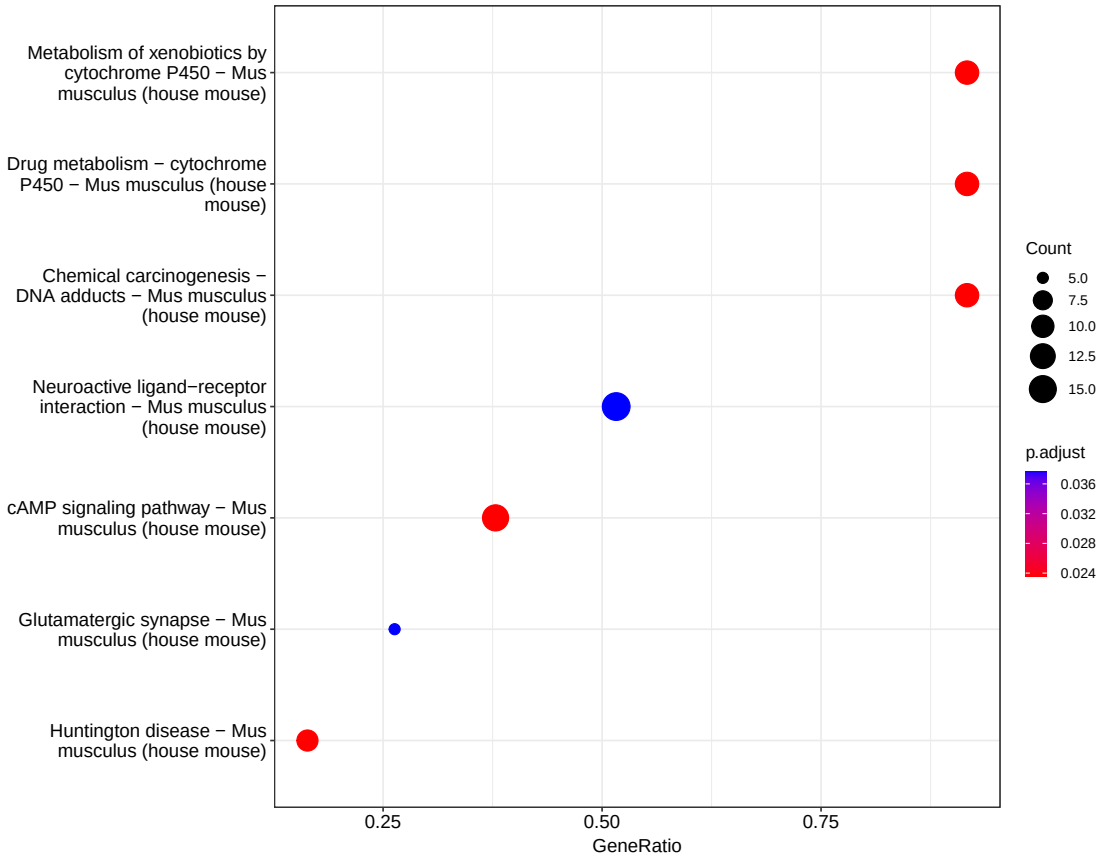

Figure 20: Theileum KEGG Treatment vs Model

Figure 21 (下方图) 为图 Theileum GSEA plot of pathways Treatment vs Model 概览。

(对应文件为 `Figure+Table/Theileum-GSEA-plot-of-pathways-Treatment-vs-Model.pdf`)

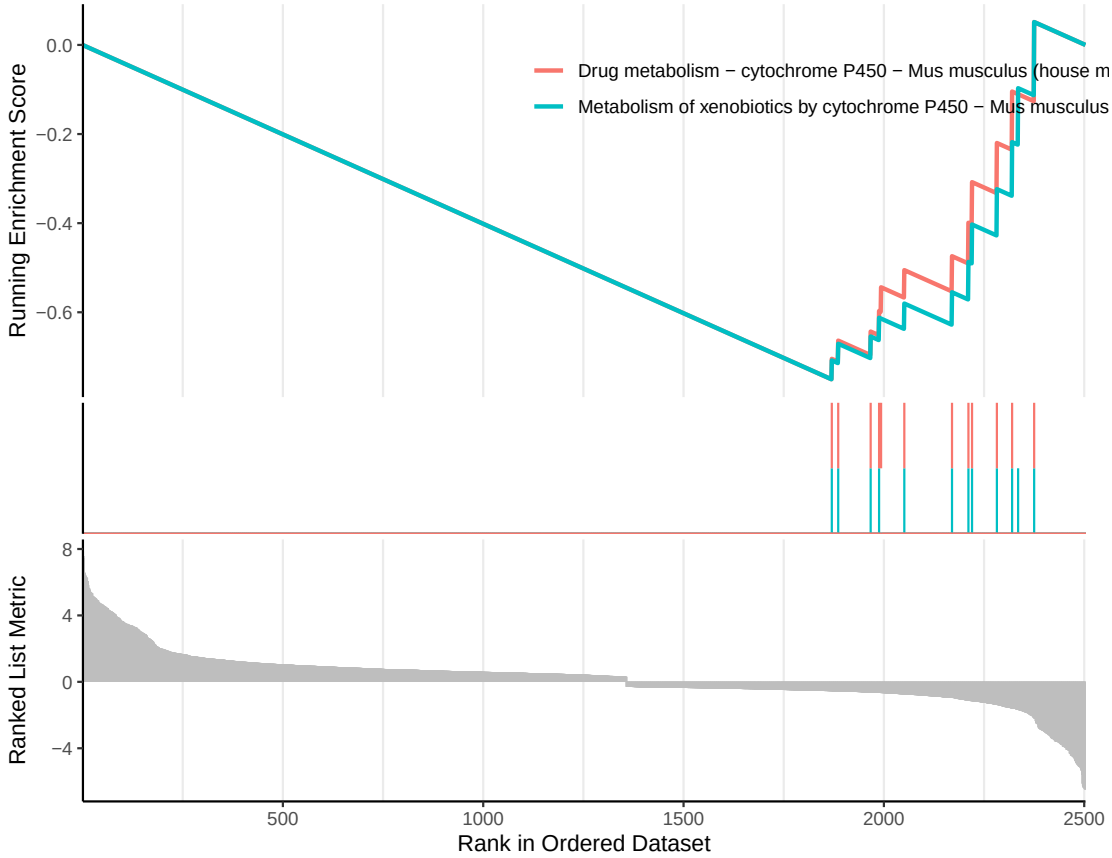

Figure 21: Theileum GSEA plot of pathways Treatment vs Model

5.6 TGR5 (Gpbar1) 相关信息

TGR5 Signaling in Hepatic Metabolic Health[2,9]

TGR5 Signaling in Hepatic Metabolic Health :

TGR5 is a transmembrane G-protein coupled receptor (GPCR) for bile acids that is ubiquitously expressed in mouse and human tissues ...

Bile Acid-Activated Receptors: GPBAR1 (TGR5) and Other G Protein-Coupled Receptors :

The BA-responsive GPCRs S1PR2 and TGR5 are almost ubiquitously expressed in human and rodent tissues. In the liver, S1PR2 is expressed in all cell types, while TGR5 is predominately found in non-parenchymal cells. In contrast to S1PR2, which is mainly activated by conjugated bile acids (BAs), all BAs serve as ligands for TGR5 irrespective of their conjugation state and substitution pattern.Mice with targeted deletion of either S1PR2 or TGR5 are viable and develop no overt phenotype ...

5.6.1 Gpbar1 相关表达水平 (count)

TGR5 (GPBAR) 的 Mgi Symbol (对应小鼠的基因名称) 为 Gpbar1

此外，这里还测试了文献[9] 中提到的 S1pr2 基因 (和 Gpbar1 同属于 GPCR) 的表达。

Table 3 (下方表格) 为表格 Gpbar1 and S1pr2 of GPCRs 概览。

(对应文件为 `Figure+Table/Gpbar1-and-S1pr2-of-GPCRs.csv`)

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

Table 3: Gpbar1 and S1pr2 of GPCRs

ensem…	mgi_s…	entre…	hgnc_…	descr…
ENSMU…	S1pr2	14739	NA	sphin…
ENSMU…	S1pr2	14739	NA	sphin…
ENSMU…	S1pr2	14739	NA	sphin…
ENSMU…	S1pr2	14739	NA	sphin…
ENSMU…	Gpbar1	227289	NA	G pro…

在所有数据集中，仅转录因子 'ENSMUST00000054197' (对应 S1pr2) 有一定表达量 (Fig. 22)，而其它全部为零表达。这可能是测序深度不足，导致表达量较低的基因无法被涵盖。所以这里难以判断这些基因的表达水平在疾病或药物干预下的变化。

(一般 Count 水平可能在 100 ~ 1000000 以上不等)

Figure 22 (下方图) 为图 Counts level of the GPCRs 概览。

(对应文件为 `Figure+Table/Counts-level-of-the-GPCRs.pdf`)

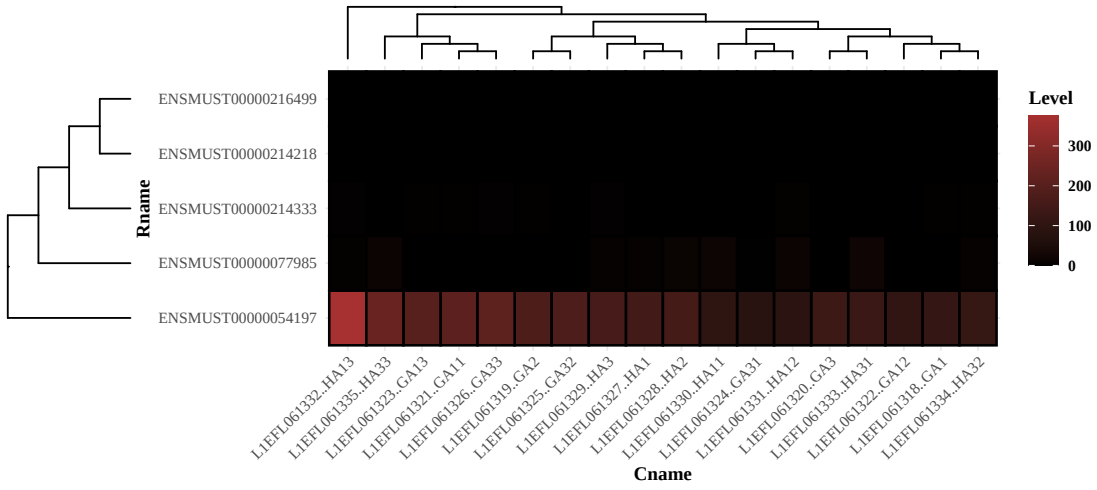

Figure 22: Counts level of the GPCRs

5.6.2 Gpbar1 或 S1pr2 表达水平 (normalized)

5.6.2.1 肝脏 Figure 23 (下方图) 为图 Normalized expression level of S1pr2 in Liver 概览。

(对应文件为 `Figure+Table/Normalized-expression-level-of-S1pr2-in-Liver.pdf`)

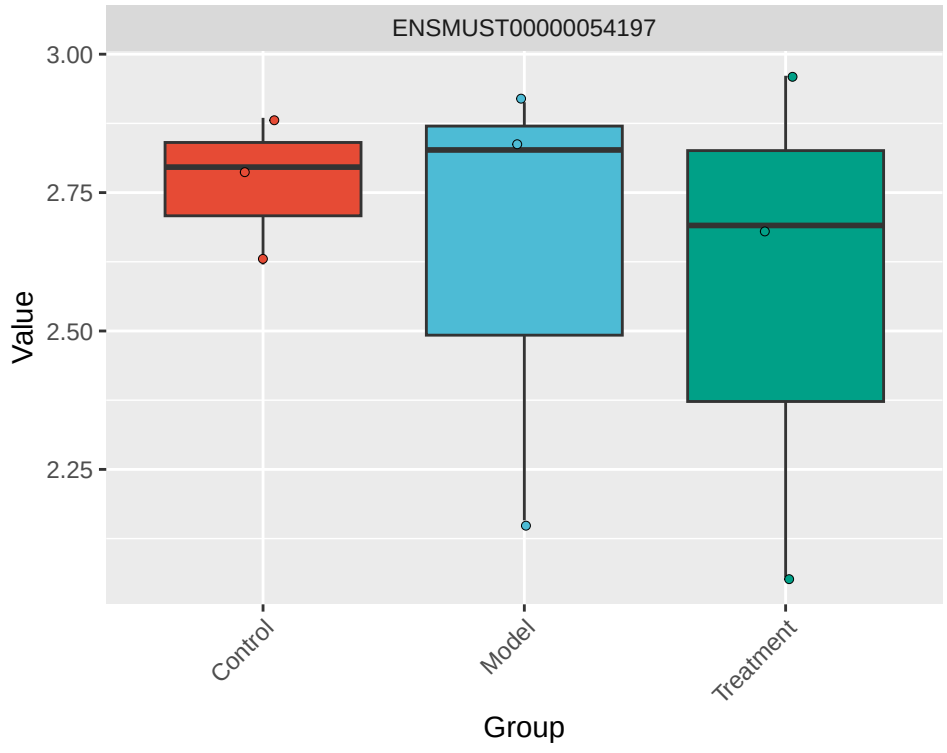

Figure 23: Normalized expression level of S1pr2 in Liver

S1pr2 不属于肝脏差异分析的显著 DEGs (5.4.1) 。

5.6.2.2 回肠 Figure 24 (下方图) 为图 Normalized expression level of S1pr2 in Theileum 概览。

(对应文件为 `Figure+Table/Normalized-expression-level-of-S1pr2-in-Theileum.pdf`)

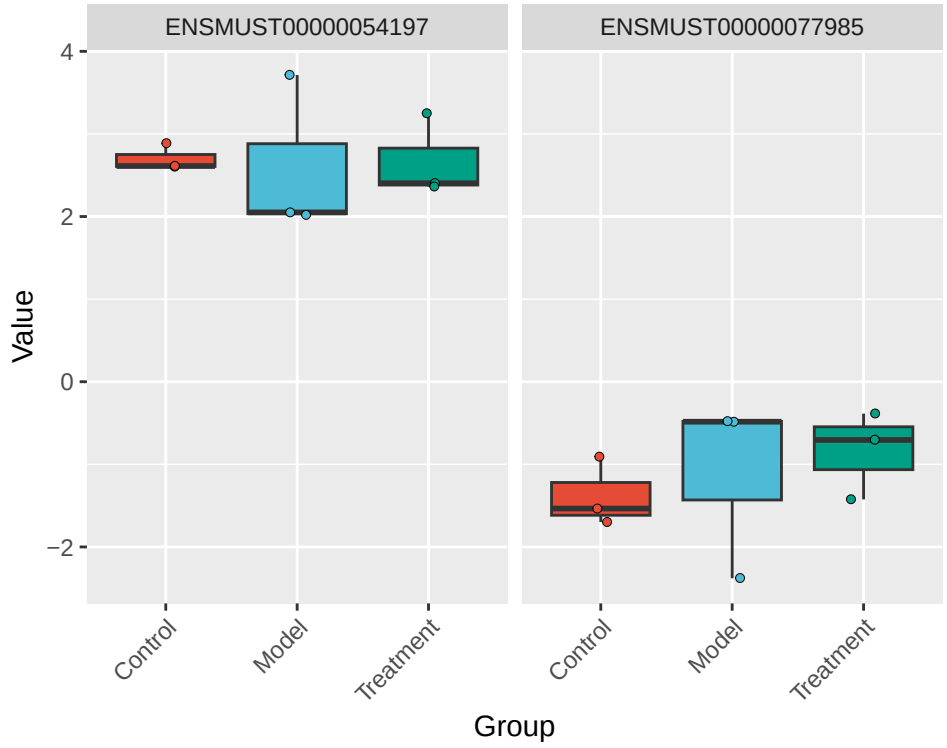

Figure 24: Normalized expression level of S1pr2 in Theileum

Gpbar1 和 S1pr2 不属于回肠差异分析的显著 DEGs (5.5.1) 。

Reference

1. Cole, T. J., Short, K. L. & Hooper, S. B. The science of steroids. *Seminars in Fetal and Neonatal Medicine* **24**, 170–175 (2019).

2. Holter, M. M., Chirikjian, M. K., Govani, V. N. & Cummings, B. P. TGR5 signaling in hepatic metabolic health. *Nutrients* **12**, (2020).

3. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomaRt. *Nature protocols* **4**, 1184–1191 (2009).

4. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).

5. Chen, S. Ultrafast one-pass fastq data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, (2023).

6. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic rna-seq quantification. *Nature Biotechnology* **34**, (2016).

7. Ritchie, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).

8. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data user's guide. 119.

9. Keitel, V., Stindt, J. & Häussinger, D. Bile acid-activated receptors: GPBAR1 (tgr5) and other g protein-coupled receptors. *Handbook of experimental pharmacology* **256**, 19–49 (2019).