# Analysis

Huang LiChuang of Wie-Biotech

## Contents



## List of Figures

## List of Tables



## 1 摘要

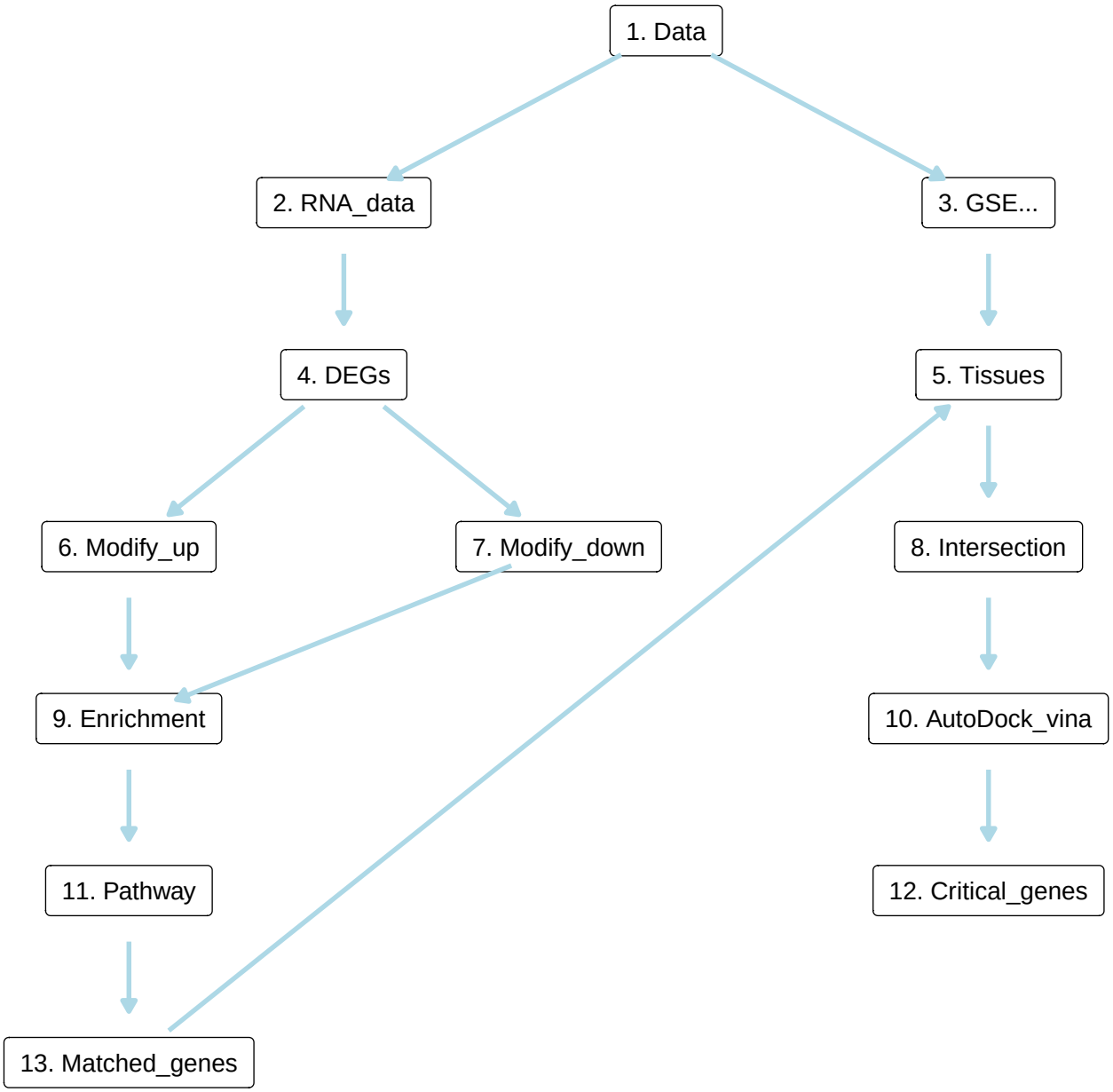
筛选丹参酮治疗脓毒症（sepsis）的关键差异表达基因及相关信号通路。

## 2 材料和方法

测序数据：Caco-2 细胞系，对照组 con，脂多糖组 LPS，丹参酮组 TNA（LPS+TNA）。

GEO 数据：GSE237861

## 3 研究设计流程图




## 4 分析结果

单以测序数据集筛选到 1797 个靶点，富集分析聚焦到 Hippo 通路（Fig. 7）。

以 GEO 数据 GSE237861 分析发现，6 种不同组织的 sepsis 病例存在 51 个共同的差异表达基因（Disease vs control）。进一步分析发现：无同时存在于 6 或 5 种组织的 Hippo 通路基因（同时也是 Tanshinone 的作用靶点）；*BIRC3*、*ID1* 在 4 种组织中差异表达；*DLG4* 在 3 种组织中差异表达（Fig. 10）。分子对接显示，SMAD7, SOX2, TGFBR2, DLG4, DLG2 具有良好亲和度（Fig. 11）。综上，*DLG4* 在 3 种 sepsis 组织中差异表达，且 DLG4 可与 Tanshinone I 结合，因此，*DLG4* 可能是 TNA 治疗 sepsis 的关键靶点之一，对应信号通路为 Hippo。

# 5 结论

*DLG4* 可能是 TNA 治疗 sepsis 的关键靶点，相关信号通路为 Hippo。

# 6 附：分析流程

## 6.1 测序数据

### 6.1.1 差异分析

Figure 1为图 Low expression filtering 概览。

**(对应文件为 `Figure+Table/Low-expression-filtering.pdf`)**

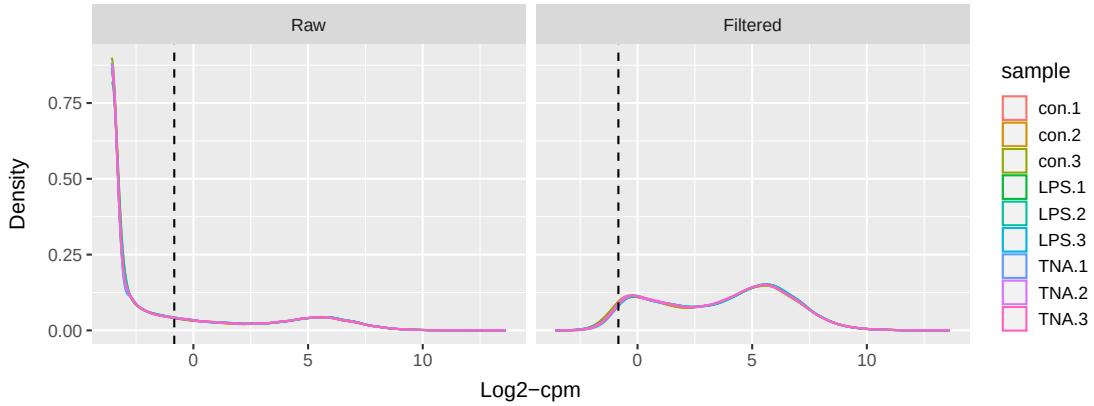



Figure 1: Low expression filtering

Figure 2为图 expression normalization 概览。

**(对应文件为 `Figure+Table/expression-normalization.pdf`)**

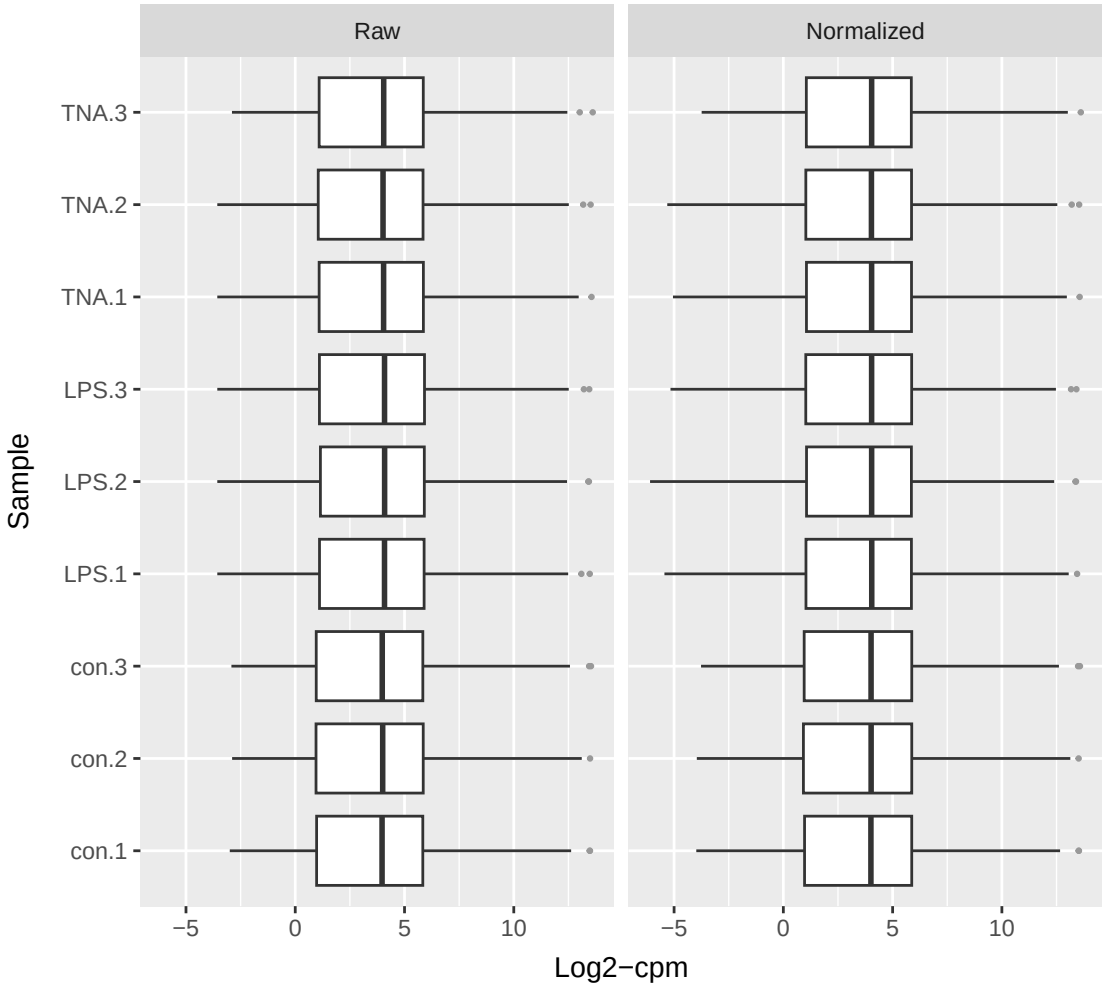



Figure 2: Expression normalization

Figure 3为图 DEGs of model versus control 概览。

**(对应文件为 `Figure+Table/DEGs-of-model-versus-control.pdf`)**

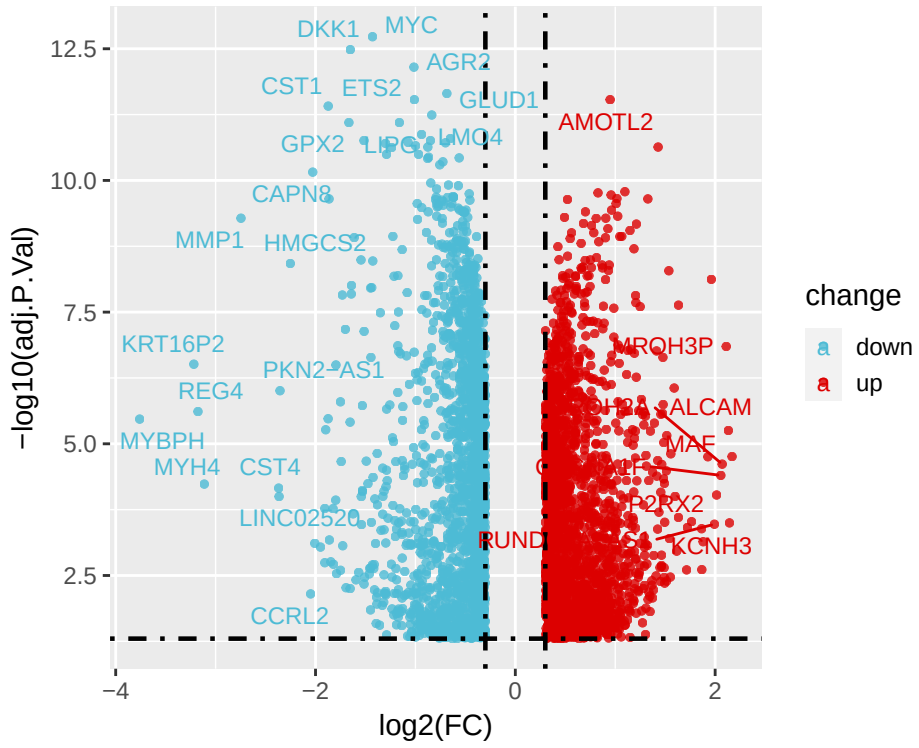



Figure 3: DEGs of model versus control

Figure 4为图 DEGs of treatment versus model 概览。

**(对应文件为 `Figure+Table/DEGs-of-treatment-versus-model.pdf`)**

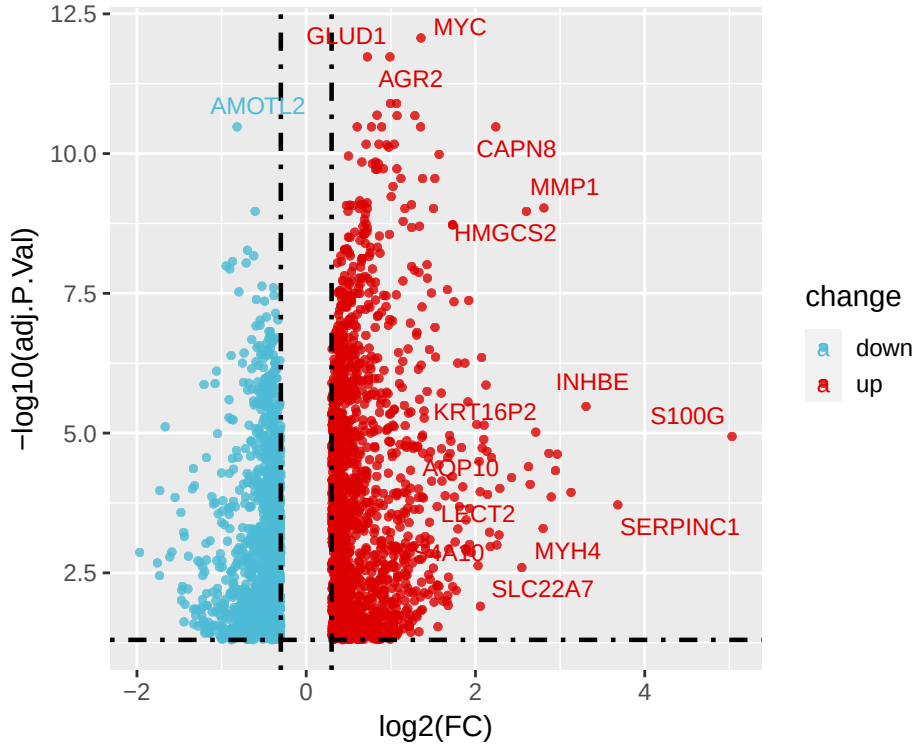


Figure 4: DEGs of treatment versus model

丹参酮的疗效有两种情况：

- 模型组相比对照组，基因上调；而以丹参酮处理后，基因下调（相比于模型组）。
- 模型组相比对照组，基因下调；而以丹参酮处理后，基因上调（相比于模型组）。

Figure 5为图 intersection of disease genes expression and treatment effect of TNA 概览。

**(对应文件为 `Figure+Table/intersection-of-disease-genes-expression-and-treatment-effect-of-TNA.pdf`)**

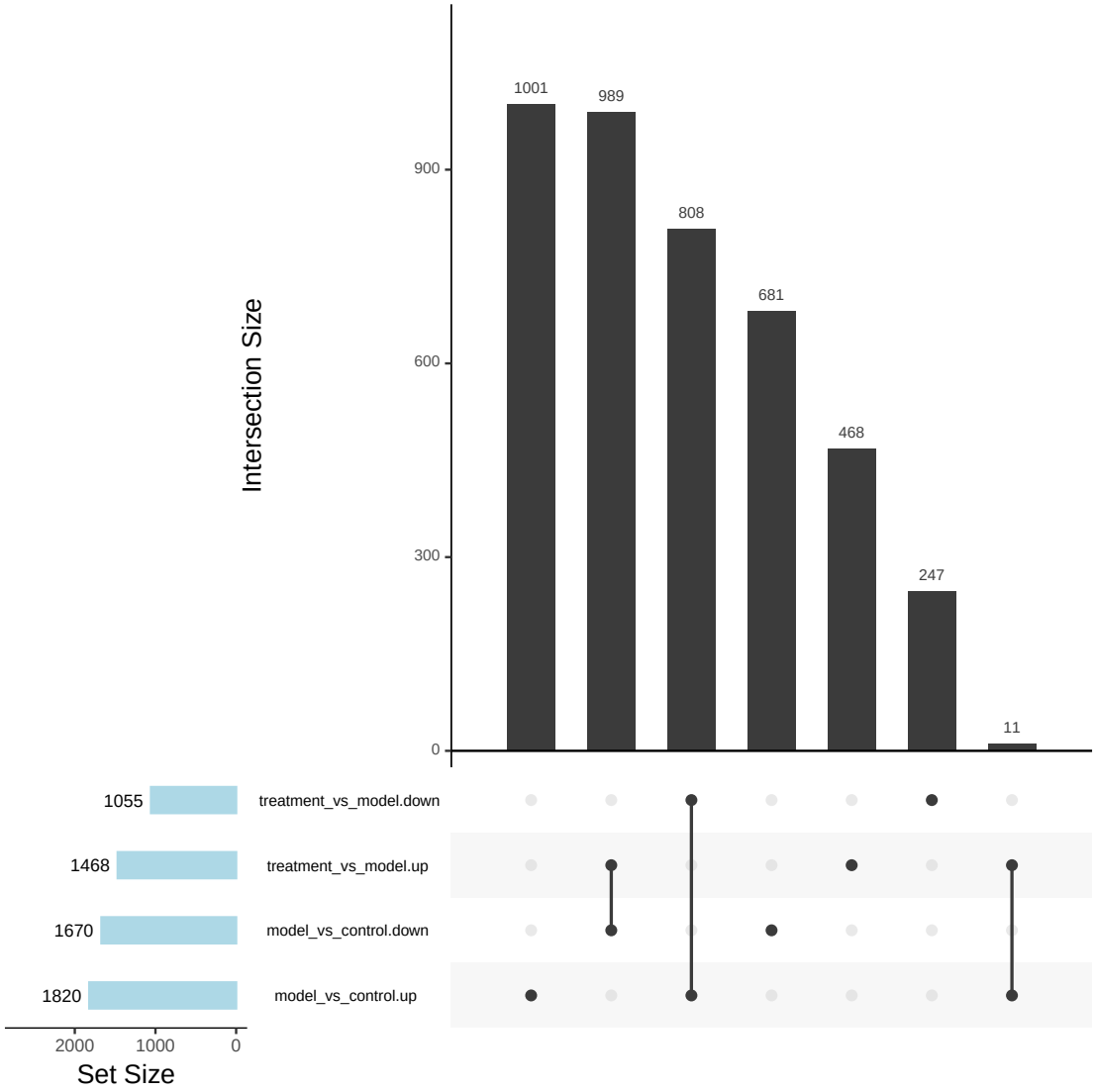



Figure 5: Intersection of disease genes expression and treatment effect of TNA

取 Fig. 5 的两组交集的合集（989 + 808），。

### 6.1.2 富集分析

以上述合集做富集分析。

Figure 6为图 KEGG enrichment 概览。

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

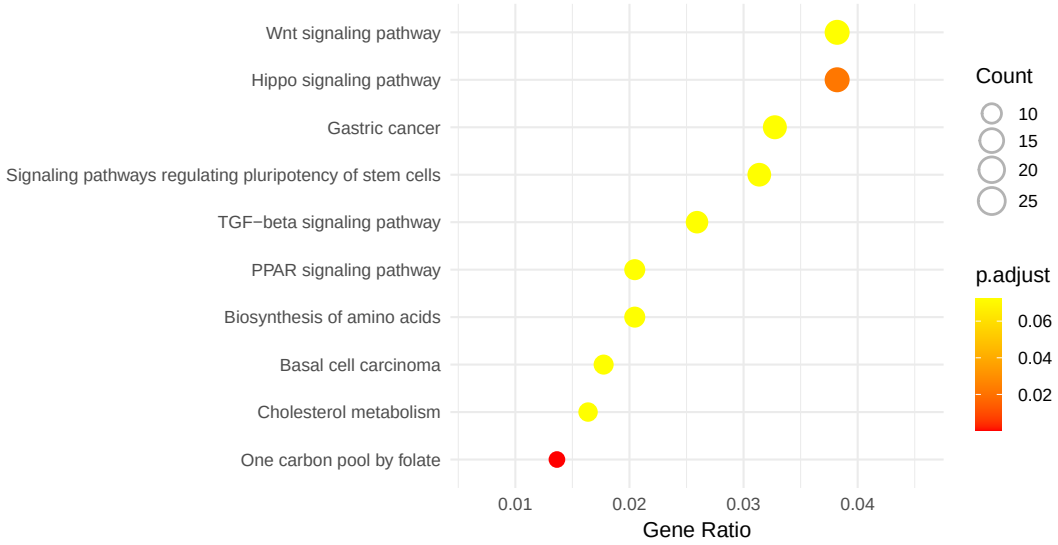



Figure 6: KEGG enrichment

Hippo 通路为显著富集通路。

Figure 7为图 genes enriched in hippo signiling pathway 概览。

**(对应文件为 `Figure+Table/hsa04390.pathview.png`)**

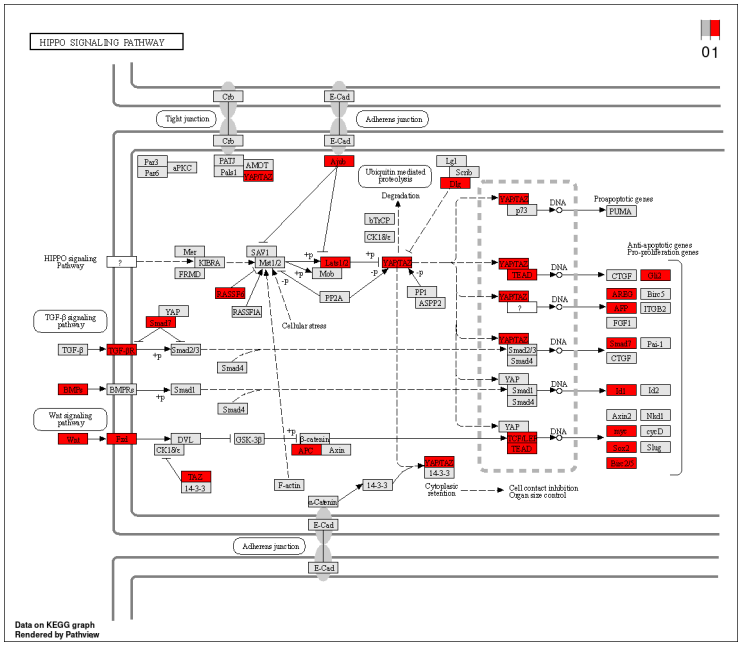



Figure 7: Genes enriched in hippo signiling pathway

## 6.2 GEO sepsis

### 6.2.1 GSE237861: Transcriptome analysis of six tissues obtained post mortem from sepsis patients

**data_processing :**

The libraries were quantified by Qubit dsDNA High Sensitivity Assay Kit (Life Technologies Corporation, Carlsbad, CA, United States) and the median sizes were determined by TapeStation 4200 (Agilent Technologies, USA), using the High Sensitivity D1000 ScreenTape assay, to form an equimolar pool.

**data_processing.1 :**

Sequencing was performed as a 75-bp single-read, single-index run on a NextSeq 500 next-generation sequencer (Illumina, San Diego, CA, United States) with High Output kit.

**data_processing.2 :**

Quality control analysis was performed using FastQC software, showing a Phred value superior to 30

**data_processing.3 :**

Trimmomatic software was used to trim low-quality reads and adapters. Raw reads were aligned to the hg38 reference through HISAT2 software. Quantification of the gene expression data was performed through the function featureCounts of the R package Rsubread and the counts were normalized according to log2CPM.

**data_processing.4 :**

Differential expression analysis was performed by the R package edgeR (FDR $< 0.1$ was considered significant), comparing each male patient with sepsis with all male uninfected controls and the female patients with sepsis with all female uninfected controls.

**data_processing.5 :**

Assembly: hg38

**data_processing.6 :**

Supplementary files format and content: tab-delimited text file contains results of differential expression analysis in edgeR

**data_processing.7 :**

Supplementary files format and content: columns indicate gene ID, logFc, p-value and FDR

Table 1为表格 metadata of GSE237861 概览。

**(对应文件为 `Figure+Table/metadata-of-GSE237861.csv`)**

> 注：表格共有 82 行 2 列，以下预览的表格可能省略部分数据；表格含有 82 个唯一 'title'。

Table 1: Metadata of GSE237861

| title | tissu… |
|---|---|
| sepsi… | prefr… |
| sepsi… | hippo… |
| sepsi… | heart |
| sepsi… | lung |
| sepsi… | kidney |
| sepsi… | colon |
| sepsi… | brain… |
| sepsi… | hippo… |
| sepsi… | heart |
| sepsi… | lung |
| sepsi… | kidney |
| sepsi… | brain… |
| sepsi… | hippo… |
| sepsi… | heart |
| sepsi… | lung |
| … | … |

Figure 8为图 DEGs number in sepsis of mutiple tissue of GEO dataset 概览。

**(对应文件为 `Figure+Table/DEGs-number-in-sepsis-of-mutiple-tissue-of-GEO-dataset.pdf`)**

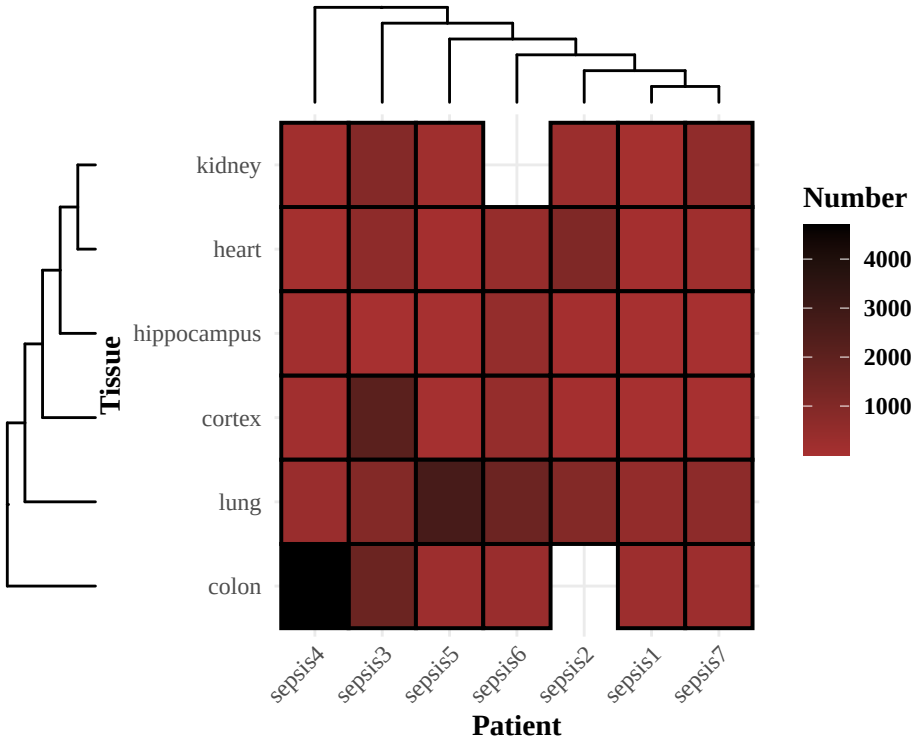



Figure 8: DEGs number in sepsis of mutiple tissue of GEO dataset

在六种不同的 sepsis 组织中，共有 51 个共同的交集基因（Fig. 9）。

Figure 9为图 intersection of DEGs of mutiple tissue of sepsis 概览。

**(对应文件为 `Figure+Table/intersection-of-DEGs-of-mutiple-tissue-of-sepsis.pdf`)**

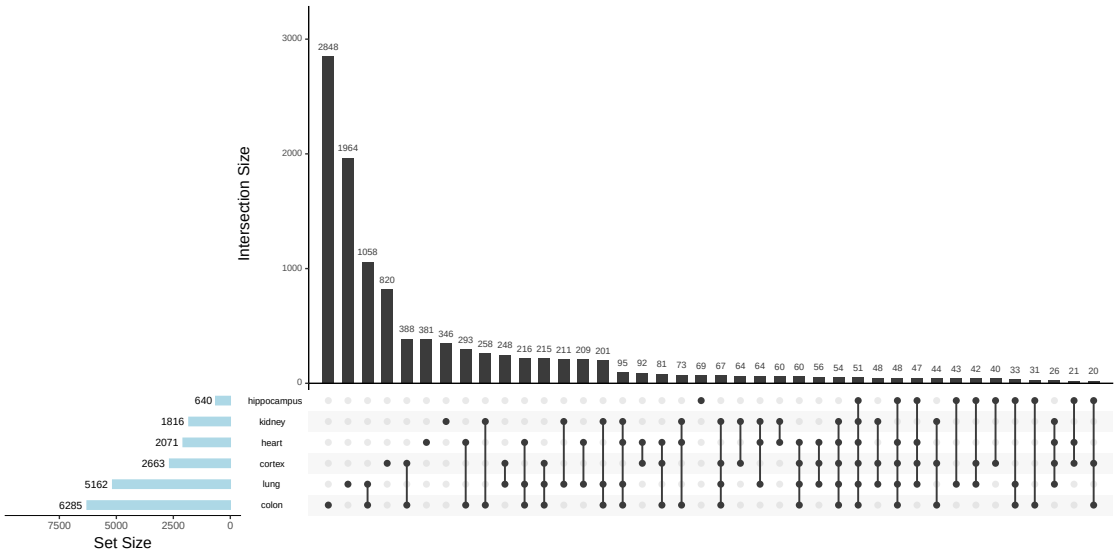



Figure 9: Intersection of DEGs of mutiple tissue of sepsis

## 6.3 整合：测序数据和 GEO 数据

### 6.3.1 关联基因

以 GSE237861 验证 TNA 作用的 Hippo 通路基因，属于 sepsis 哪些组织的差异表达基因，以确认 TNA 是否对其具有疗效。

- BIRC3、ID1 在 4 种组织中差异表达
- DLG4 在 3 种组织中差异表达
- ...

Figure 10为图 Target genes of TNA in mutiple tissue of sepsis of Hippo pathway 概览。

**(对应文件为 `Figure+Table/Target-genes-of-TNA-in-mutiple-tissue-of-sepsis-of-Hippo-pathway.pdf`)**

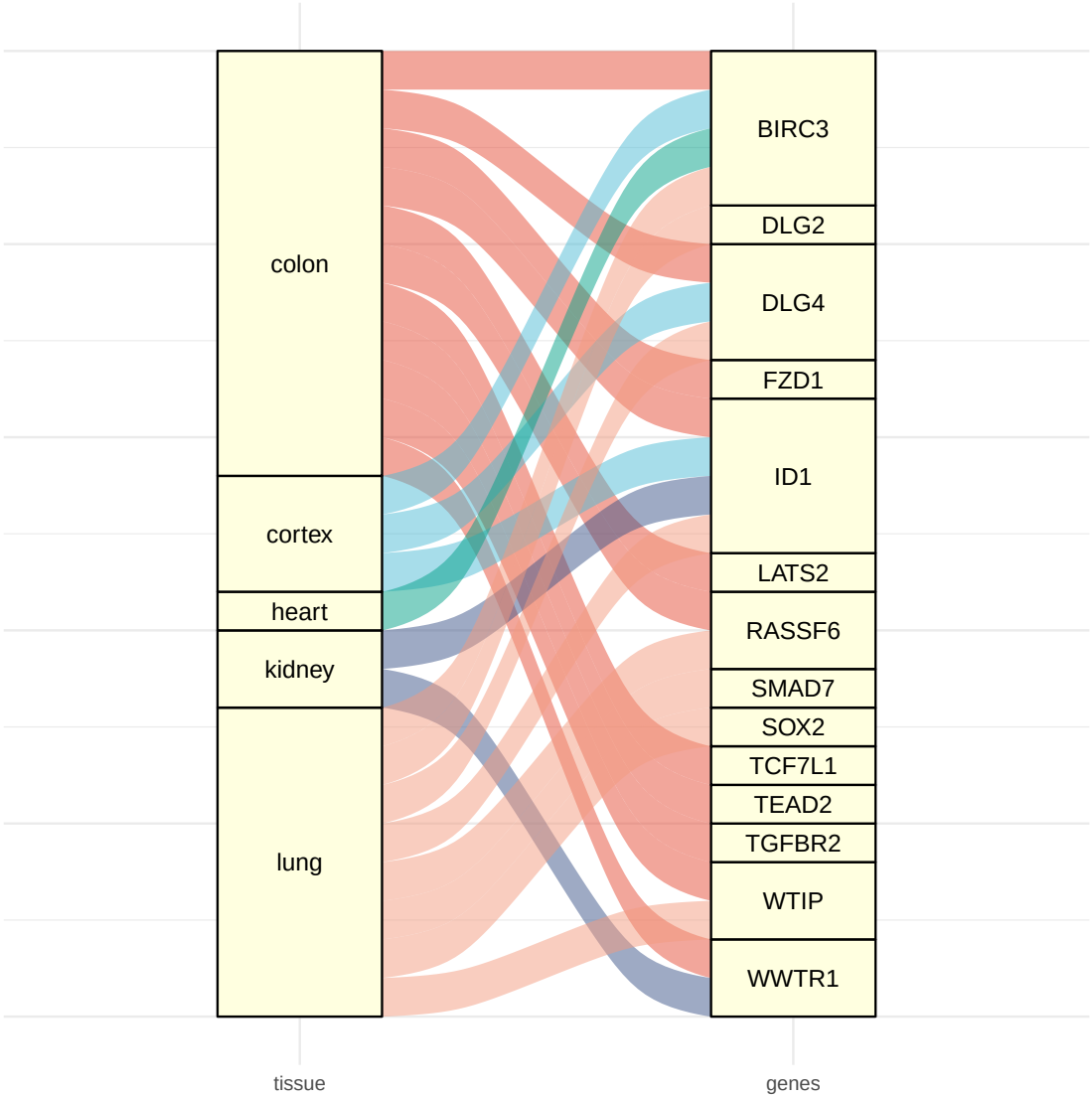



Figure 10: Target genes of TNA in mutiple tissue of sepsis of Hippo pathway

### 6.3.2 分子对接

丹参酮 I（Tanshinone I, CID:114917）

以 AutoDock Vina 对 Fig. 10 所示基因的蛋白以 Tanshinone I 分子对接。

结果显示，SMAD7, SOX2, TGFBR2, DLG4, DLG2 具有良好亲和度。

结合 Fig. 10 所示的多组织差异表达，DLG4 同时在 3 种组织 sepsis 差异表达，且为 TNA 作用靶点，表现良好对接亲和度，可能是 TNA 治疗的关键靶点之一。

Figure 11为图 docking affinity 概览。

**(对应文件为 `Figure+Table/docking-affinity.pdf`)**

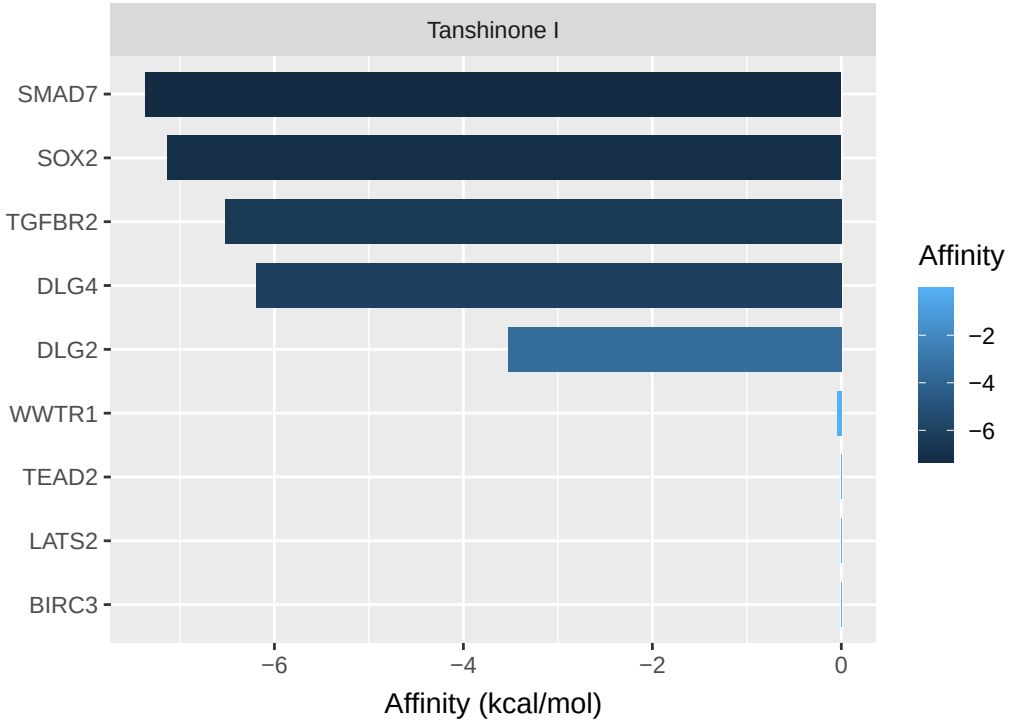



Figure 11: Docking affinity

Figure 12为图 Tanshinone I binding with protein DLG4 概览。

**(对应文件为 `Figure+Table/114917_into_1kef.png`)**

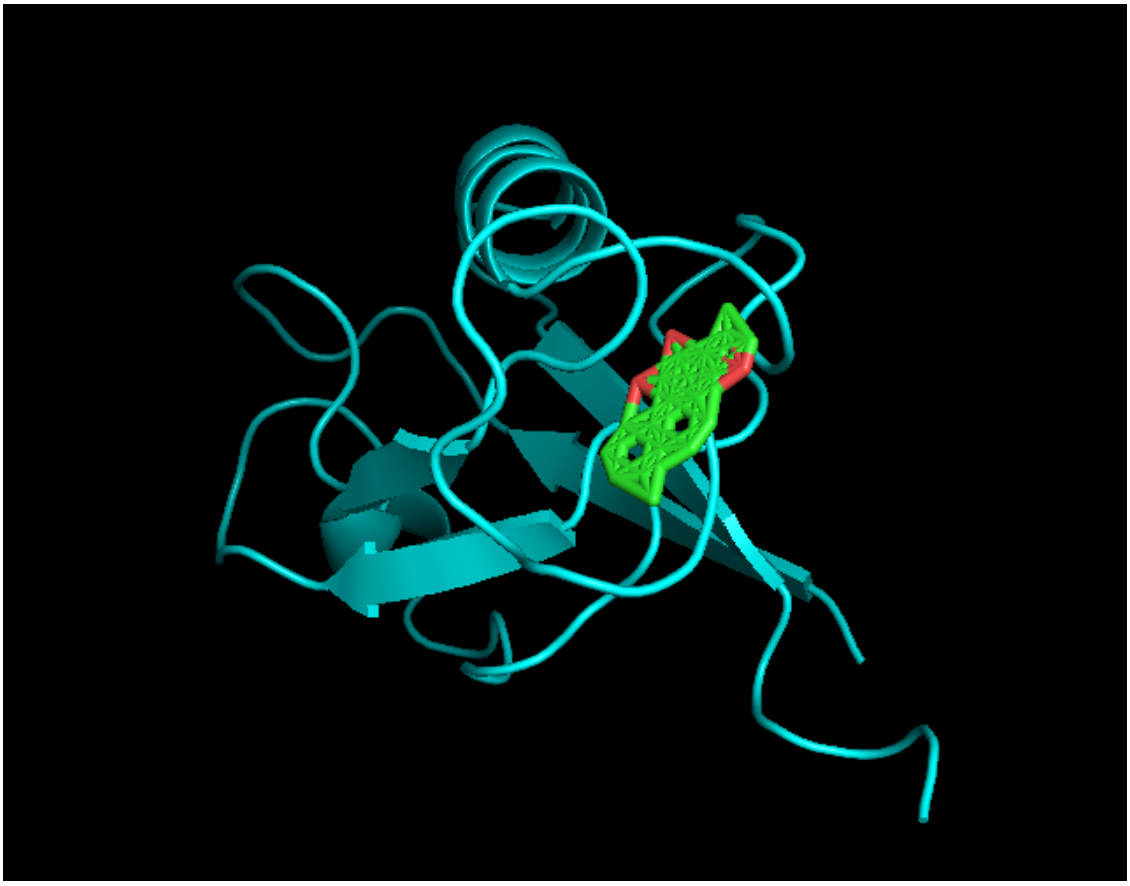


Figure 12: Tanshinone I binding with protein DLG4