**生信分析报告**

**项目标题： 脓毒症发病标志物 ;**

**单 号： BSJF241107 ;**

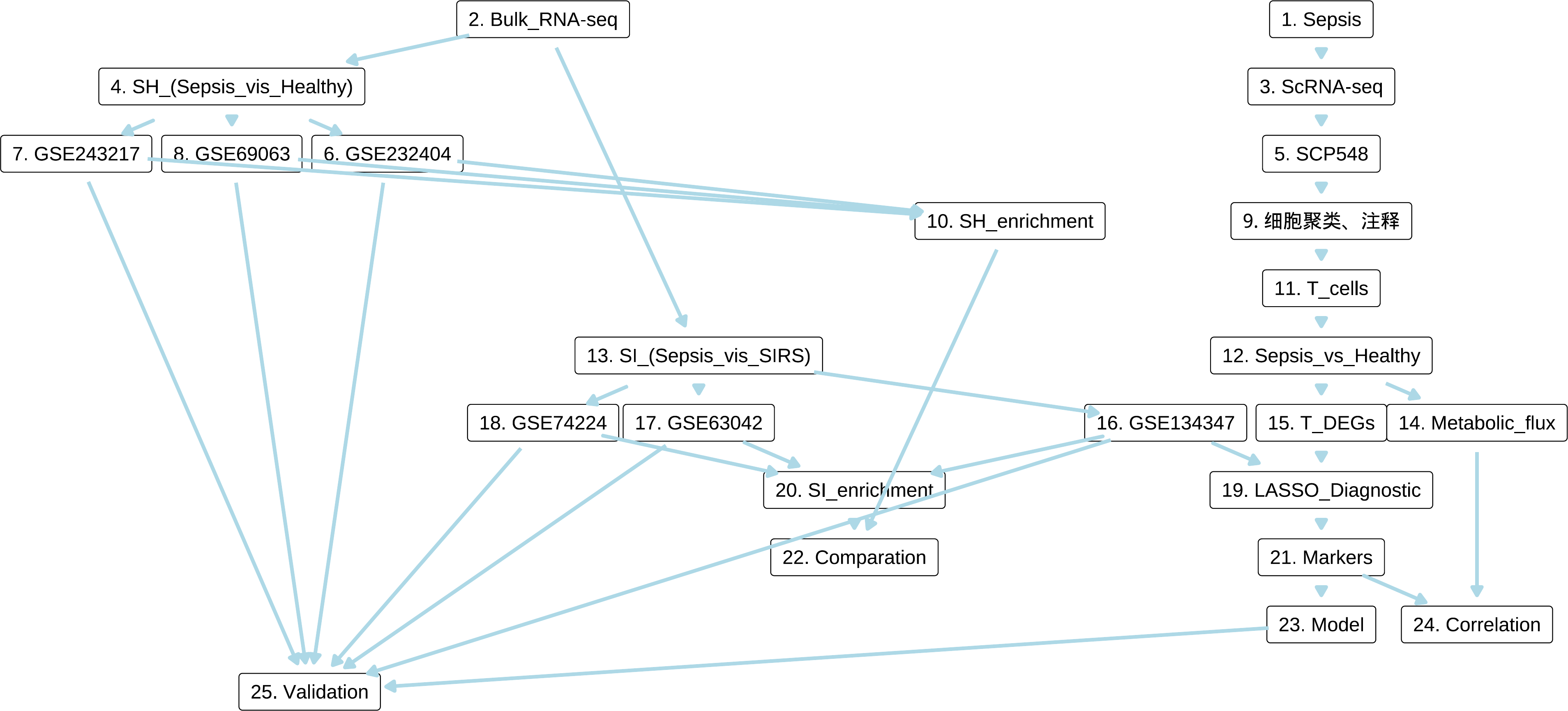
**分析人员： 黄礼闯 ;**

**分析类型： 生信分析 ;**

**委 托 人： 许冠华 ;**

**受 托 人： 杭州铂赛生物科技有限公司 .**

# 1 分析流程



**Fig.** **1** Route

**(File path: Figure+Table/1.0.0\_分析流程\_{#abstract}/Route.pdf)**

# 2 材料和方法

## 2.1 数据分析平台

在 Linux pop-os x86\_64 (6.9.3-76060903-generic) 上，使用 R version 4.4.2 (2024-10-31) (<https://www.r-project.org/>) 对数据统计分析与整合分析。

## 2.2 Seurat 单细胞数据分析 (Dataset: SEPSIS)

使用 Seurat R 包 (5.1.0) 进行单细胞数据质量控制 (QC) 和下游分析。依据 <Tutorial: <https://satijalab.org/seurat/articles/pbmc3k_tutorial.html>> 为指导对单细胞数据预处理。 一个细胞至少应有 0 个基因，并且基因数量小于 4000。线粒体基因的比例小于 10%。根据上述条件，获得用于下游分析的高质量细胞。执行标准 Seurat 分析工作流 (NormalizeData, FindVariableFeatures, ScaleData, RunPCA)。以 ElbowPlot 判断后续分析的 PC 维度。 在 1-15 PC 维度下，以 Seurat::FindNeighbors 构建 Nearest-neighbor Graph。随后在 1.2 分辨率下，以 Seurat::FindClusters 函数识别细胞群并以 Seurat::RunUMAP 进行 UMAP 聚类。

## 2.3 ClusterProfiler 富集分析 (Dataset: BULK)

以 ClusterProfiler R 包 (4.15.0.2) (2021, **IF:33.2**, Q1, The Innovation)1进行 KEGG 和 GO 富集分析。以 p.adjust 表示显著水平。

## 2.4 GEO 数据获取 (Dataset: PB\_GSE134347)

以 R 包 GEOquery (2.74.0) 获取 GSE134347 数据集。

## 2.5 Limma 差异分析 (Dataset: PB\_GSE134347)

以 limma (3.62.1) (2005)2 差异分析。创建设计矩阵，对比矩阵，差异分析：sepsis vs noninfectious。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 P.Value 小于 0.05，|Log2(FC)| 大于 0.5 的统计结果。

## 2.6 GEO 数据获取 (Dataset: PB\_GSE63042)

以 R 包 GEOquery (2.74.0) 获取 GSE63042 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 缺失样本: GSM1548259, GSM1548263, GSM1548283, GSM1548285, GSM1548310, GSM1548326, GSM1548332, GSM1548333, GSM1548334, GSM1548335, GSM1548340, GSM1548345 (‘NCBI-generated data’ 缺失样本计数数据的原因包括运行未通过 50% 的对齐率或由于技术原因处理失败) 以及基因注释。

## 2.7 Limma 差异分析 (Dataset: PB\_GSE63042)

以 R 包 edgeR (4.4.0) ()3 对数据预处理。以 edgeR::filterByExpr 过滤 count 数量小于 10 的基因。以 edgeR::calcNormFactors，limma::voom 转化 count 数据为 log2 counts-per-million (logCPM)。 以 limma (3.62.1) (2005)2 差异分析。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。创建设计矩阵，对比矩阵，差异分析：Sepsis vs SIRS。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 P.Value 小于 0.05，|Log2(FC)| 大于 0.5 的统计结果。

## 2.8 GEO 数据获取 (Dataset: PB\_GSE74224)

以 R 包 GEOquery (2.74.0) 获取 GSE74224 数据集。

## 2.9 Limma 差异分析 (Dataset: PB\_GSE74224)

以 limma (3.62.1) (2005)2 差异分析。创建设计矩阵，对比矩阵，差异分析：Sepsis vs SIRS。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 P.Value 小于 0.05，|Log2(FC)| 大于 0.5 的统计结果。

## 2.10 GEO 数据获取 (Dataset: BL\_GSE243217)

以 R 包 GEOquery (2.74.0) 获取 GSE243217 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 缺失样本: GSM7780773, GSM7780777, GSM7780778, GSM7780779, GSM7780780, GSM7780781, GSM7780782, GSM7780784, GSM7780785, GSM7780789, GSM7780790, GSM7780791, GSM7780792, GSM7780793, GSM7780795, GSM7780797, GSM7780798, GSM7780800, GSM7780801, GSM7780802, GSM7780803, GSM7780804, GSM7780805, GSM7780806, GSM7780807, GSM7780808, GSM7780809, GSM7780813, GSM7780815, GSM7780817, GSM7780818, GSM7780819, GSM7780820, GSM7780822, GSM7780823, GSM7780826, GSM7780827, GSM7780828, GSM7780829, GSM7780830, GSM7780831, GSM7780832, GSM7780833, GSM7780834, GSM7780835, GSM7780837, GSM7780838, GSM7780840, GSM7780841, GSM7780842 (‘NCBI-generated data’ 缺失样本计数数据的原因包括运行未通过 50% 的对齐率或由于技术原因处理失败) 以及基因注释。

## 2.11 Limma 差异分析 (Dataset: BL\_GSE243217)

以 R 包 edgeR (4.4.0) ()3 对数据预处理。以 edgeR::filterByExpr 过滤 count 数量小于 10 的基因。以 edgeR::calcNormFactors，limma::voom 转化 count 数据为 log2 counts-per-million (logCPM)。 以 limma (3.62.1) (2005)2 差异分析。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。创建设计矩阵，对比矩阵，差异分析：Sepsis vs healthy\_donor。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 adj.P.Val 小于 0.05，|Log2(FC)| 大于 1 的统计结果。

## 2.12 GEO 数据获取 (Dataset: PB\_GSE232404)

以 R 包 GEOquery (2.74.0) 获取 GSE232404 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 以及基因注释。

## 2.13 Limma 差异分析 (Dataset: PB\_GSE232404)

以 R 包 edgeR (4.4.0) ()3 对数据预处理。以 edgeR::filterByExpr 过滤 count 数量小于 10 的基因。以 edgeR::calcNormFactors，limma::voom 转化 count 数据为 log2 counts-per-million (logCPM)。 以 limma (3.62.1) (2005)2 差异分析。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。创建设计矩阵，对比矩阵，差异分析：SAKI vs HC。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 adj.P.Val 小于 0.05，|Log2(FC)| 大于 1 的统计结果。

## 2.14 GEO 数据获取 (Dataset: PB\_GSE69063)

以 R 包 GEOquery (2.74.0) 获取 GSE69063 数据集。

## 2.15 Limma 差异分析 (Dataset: PB\_GSE69063)

以 limma (3.62.1) (2005)2 差异分析。创建设计矩阵，对比矩阵，差异分析：Sepsis vs Healthy\_control。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 adj.P.Val 小于 0.05，|Log2(FC)| 大于 1 的统计结果。

## 2.16 ClusterProfiler 富集分析 (Dataset: SH)

以 ClusterProfiler R 包 (4.15.0.2) (2021, **IF:33.2**, Q1, The Innovation)1进行 KEGG 和 GO 富集分析。以 p.adjust 表示显著水平。

## 2.17 Lasso 诊断模型建立 (Dataset: BULK)

以 R 包 glmnet (4.1.8) 作 lasso 处罚的 binomial 回归，以 cv.glmnet 函数作 10 交叉验证获得模型。

## 2.18 scFEA 单细胞数据的代谢通量预测 (Dataset: SEPSIS)

将 Seurat 的 RNA Assay (‘counts’) 作为输入数据，以 scFEA 预测细胞的代谢通量 (2021, **IF:6.2**, Q1, Genome research)4。参考 <https://github.com/changwn/scFEA/blob/master/scFEA_tutorial1.ipynb> 和 <https://github.com/changwn/scFEA/blob/master/scFEA_tutorial2.ipynb>。

## 2.19 Limma 代谢通量差异分析 (Dataset: SEPSIS\_FLUX)

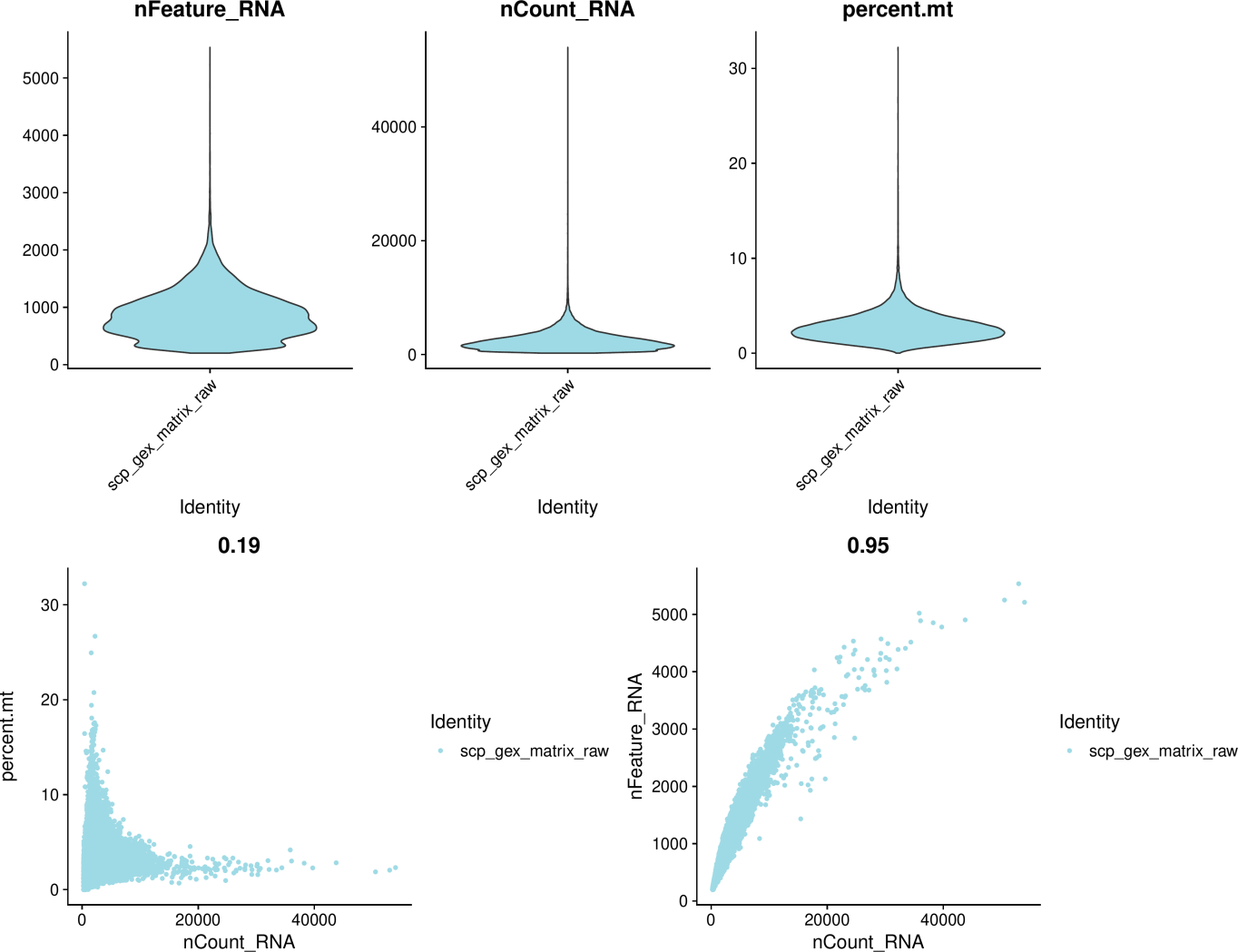
以 limma (3.62.1) (2005)2 差异分析。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。创建设计矩阵，对比矩阵，差异分析：Sepsis vs Control。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 adj.P.Val 小于 0.05，|Log2(FC)| 大于 0.2 的统计结果。

# 3 分析结果

## 3.1 Seurat 单细胞数据分析 (SEPSIS)

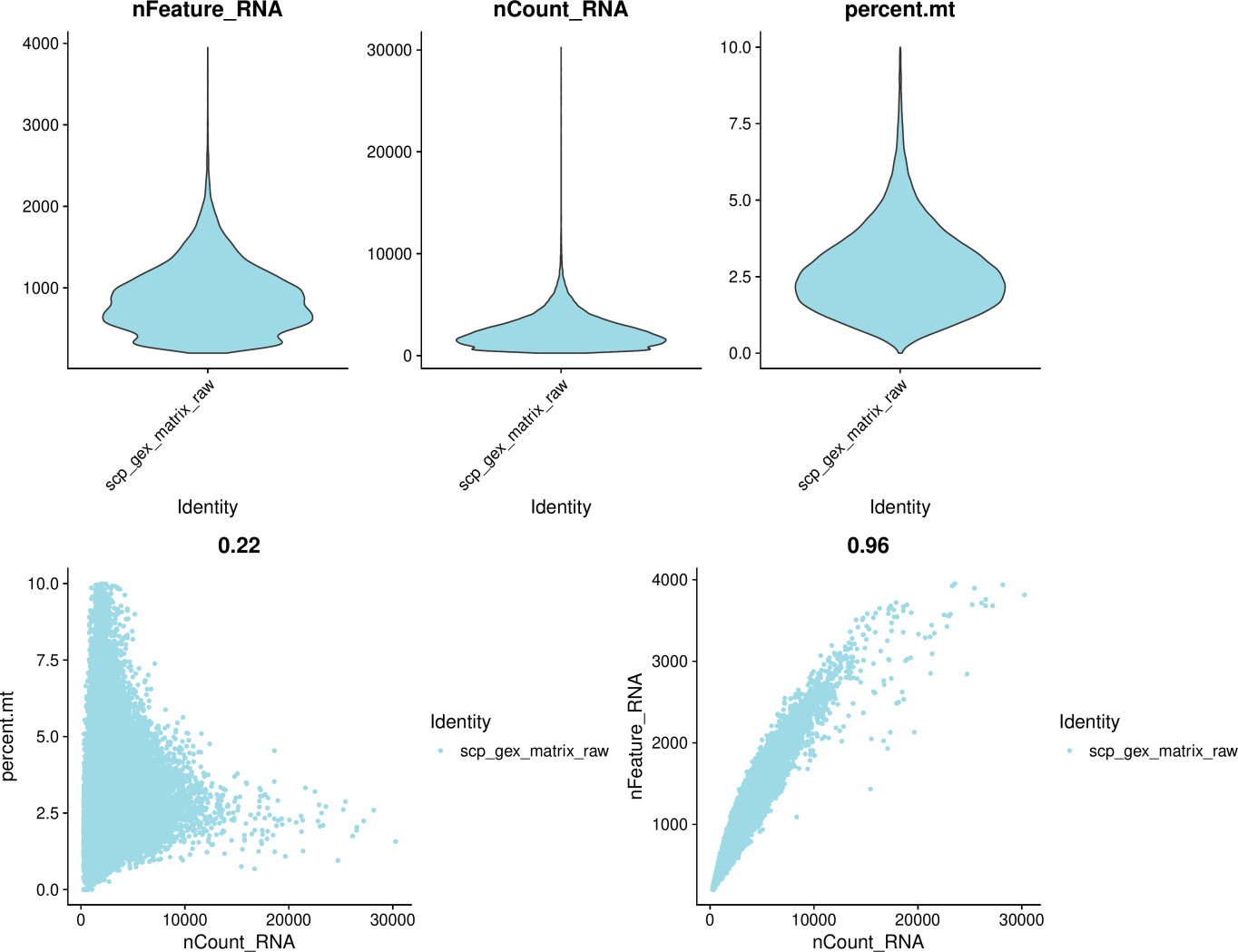
该数据集来源于 <https://singlecell.broadinstitute.org/single_cell/study/SCP548/an-immune-cell-signature-of-bacterial-sepsis-patient-pbmcs#study-summary> ((2020, **IF:58.7**, Q1, Nature medicine)5) 前期质量控制，一个细胞至少应有 0 个基因，并且基因数量小于 4000。过滤后，数据集共包含 21173 个基因，92918 个细胞。线粒体基因的比例小于 10%。数据归一化，PCA 聚类 (Seurat 标准工作流，见方法章节) 后，绘制 PC standard deviations 图。在 1-15 PC 维度，1.2 分辨率下，对细胞群 UMAP 聚类。 随后，以原作者的细胞注释 ((2020, **IF:58.7**, Q1, Nature medicine)5)， 映射到 UMAP 聚类中 (Fig. 为 Cell\_Sample 的 umap 聚类图。)。 对细胞群差异分析 (依据 Cell\_Sample, 分析 “T\_Sepsis vs T\_Control”)，筛选差异表达基因。

Fig. 为 QC (质量控制) 图 (数据过滤前) 。 Fig. 为数据过滤后的 QC 图。 Fig. 为主成分 (PC) 的 Standard deviations。 Fig. Fig. 为 Cell\_Type 的 umap 聚类图。 (为原作者的注释结果，映射到本次分析的 UMAP 中) Fig. 为使用 CellMarker 数据库中的特异性 Marker 对单细胞注释结果的验证热图。 Fig. Tab. 细胞群差异表达基因附表 (其中 ‘contrast’ 列为比较的两类细胞) (|log2(FC)| > 1, P-Adjust < 0.5)。



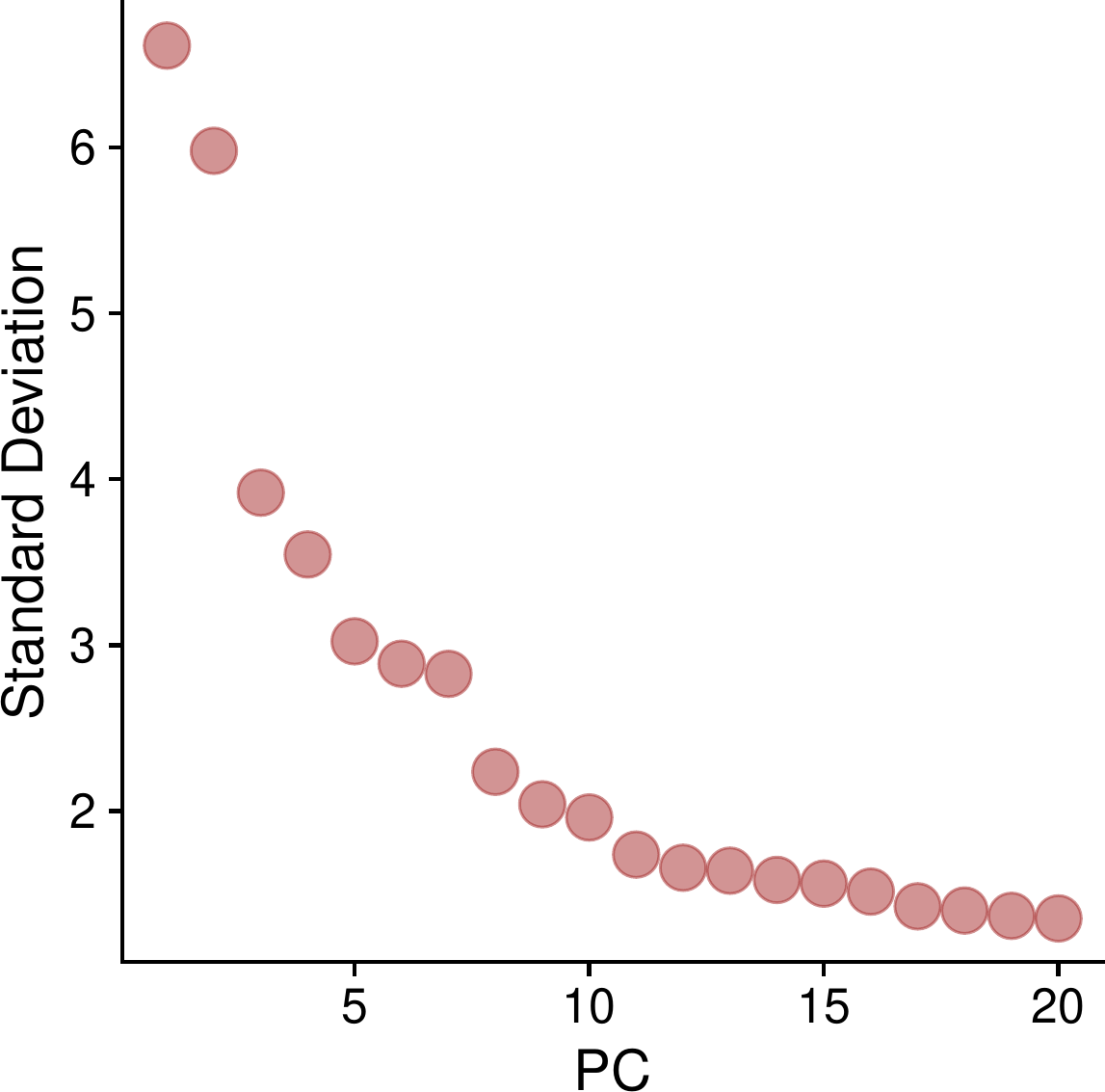
**Fig.** **2** SEPSIS Pre Quality control

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-Pre-Quality-control.pdf)**



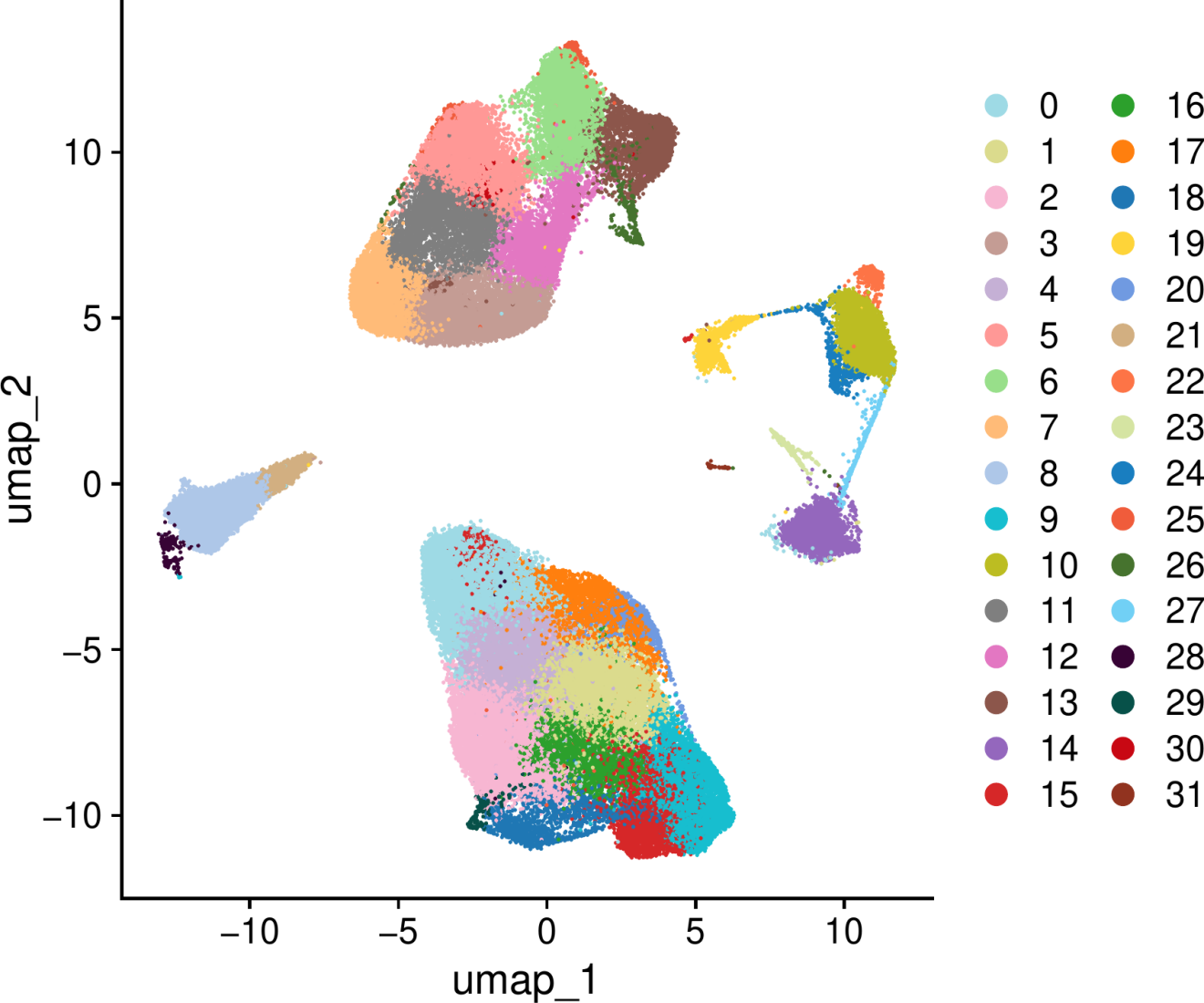
**Fig.** **3** SEPSIS After Quality control

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-After-Quality-control.pdf)**



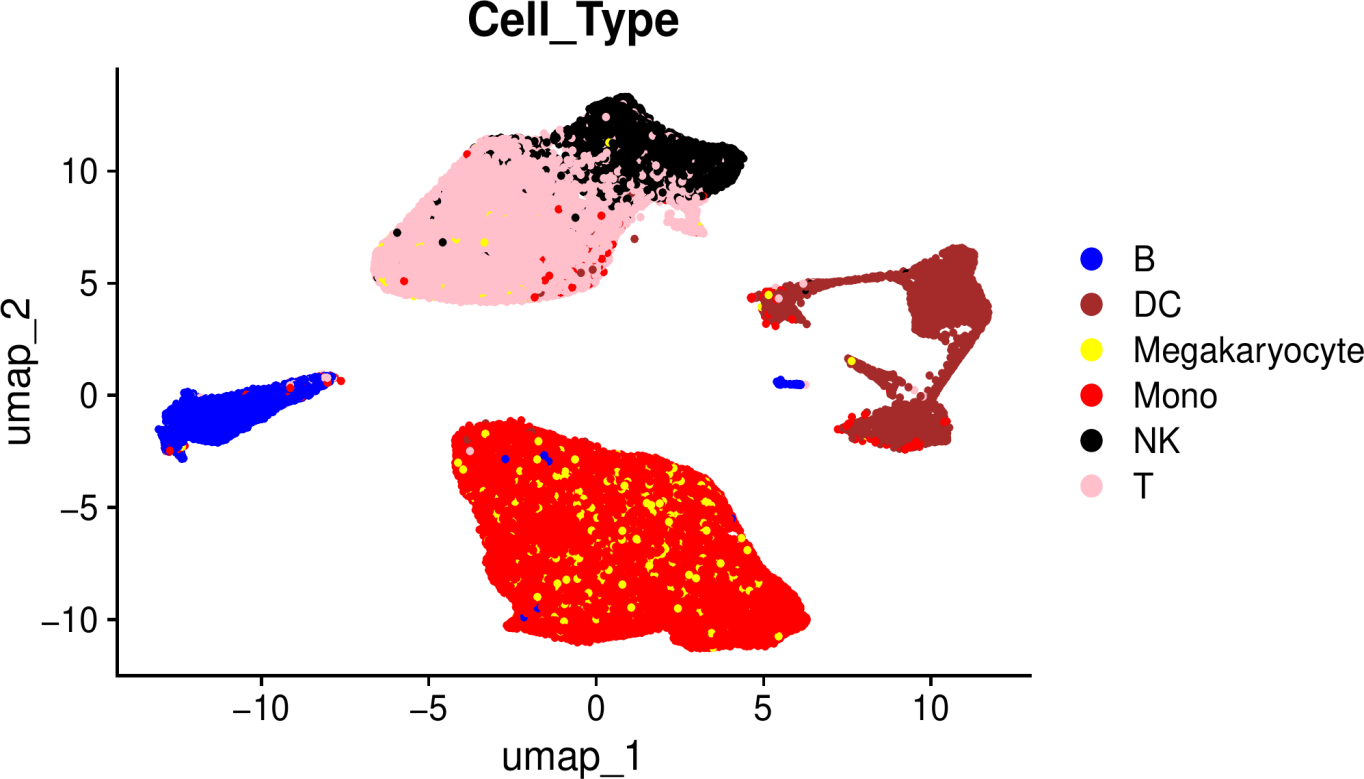
**Fig.** **4** SEPSIS Standard deviations of PCs

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-Standard-deviations-of-PCs.pdf)**



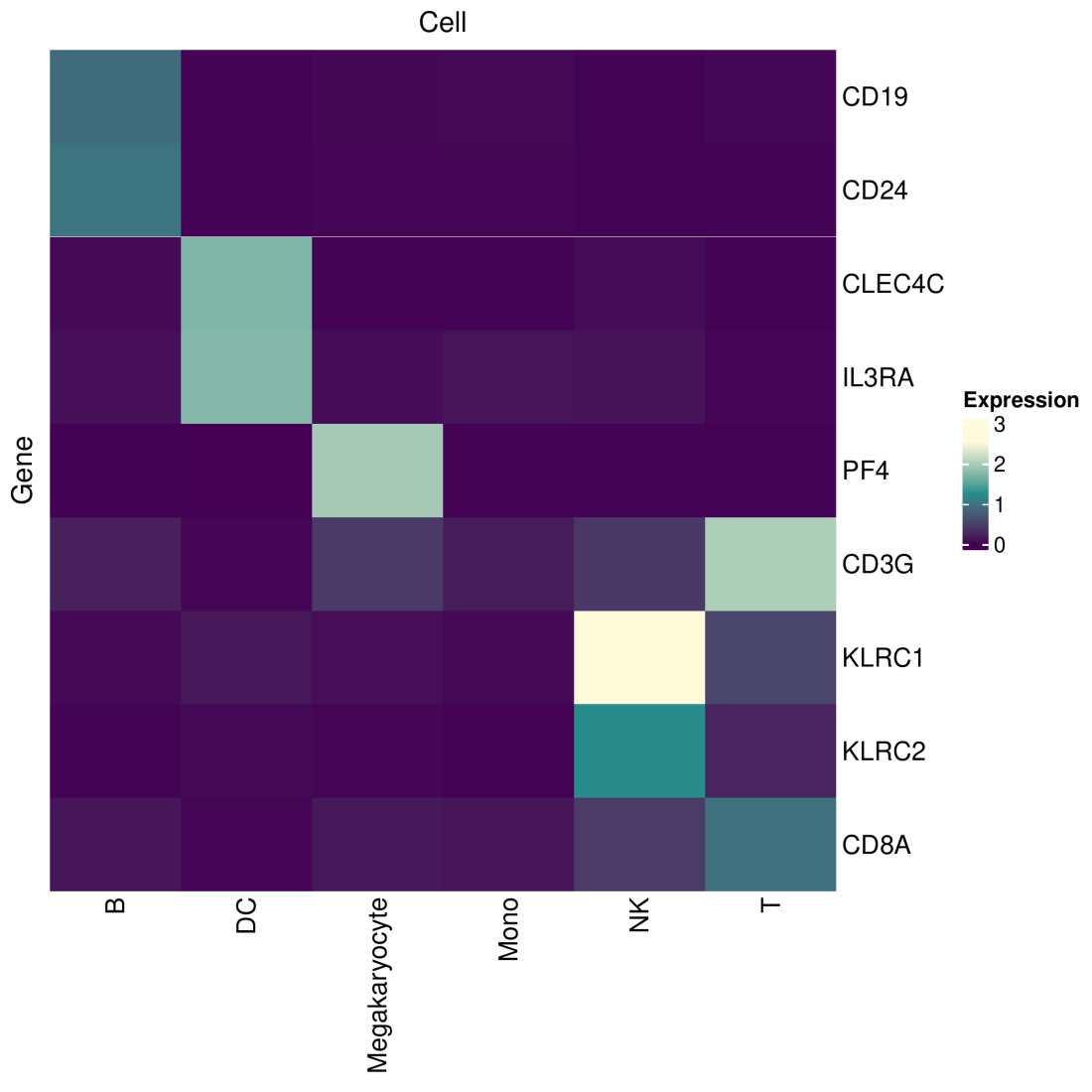
**Fig.** **5** SEPSIS UMAP Clustering

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-UMAP-Clustering.pdf)**



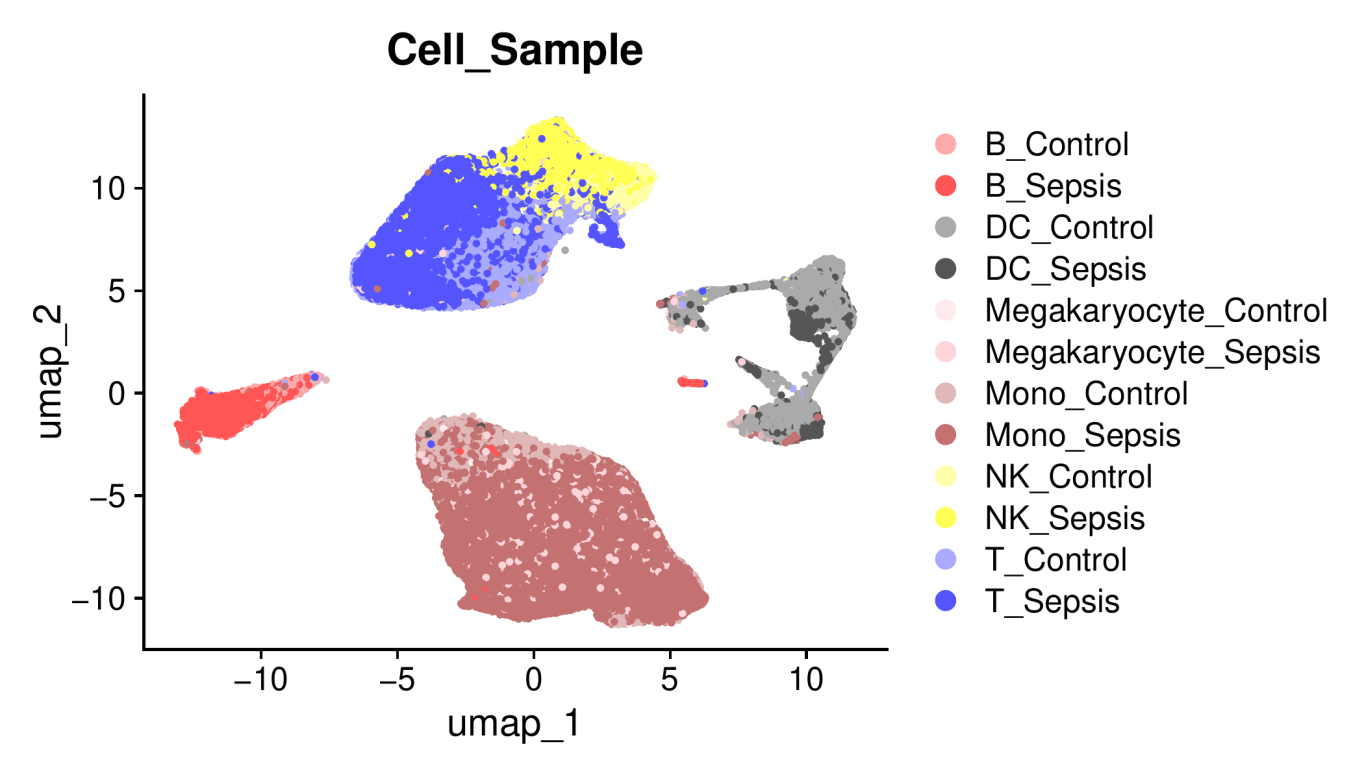
**Fig.** **6** SEPSIS The Cell Type

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-The-Cell-Type.pdf)**



**Fig.** **7** SEPSIS CellMarker Validation

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-CellMarker-Validation.pdf)**



**Fig.** **8** SEPSIS The Cell Sample

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-The-Cell-Sample.pdf)**

**Tab.** **1** SEPSIS DEGs of the contrasts

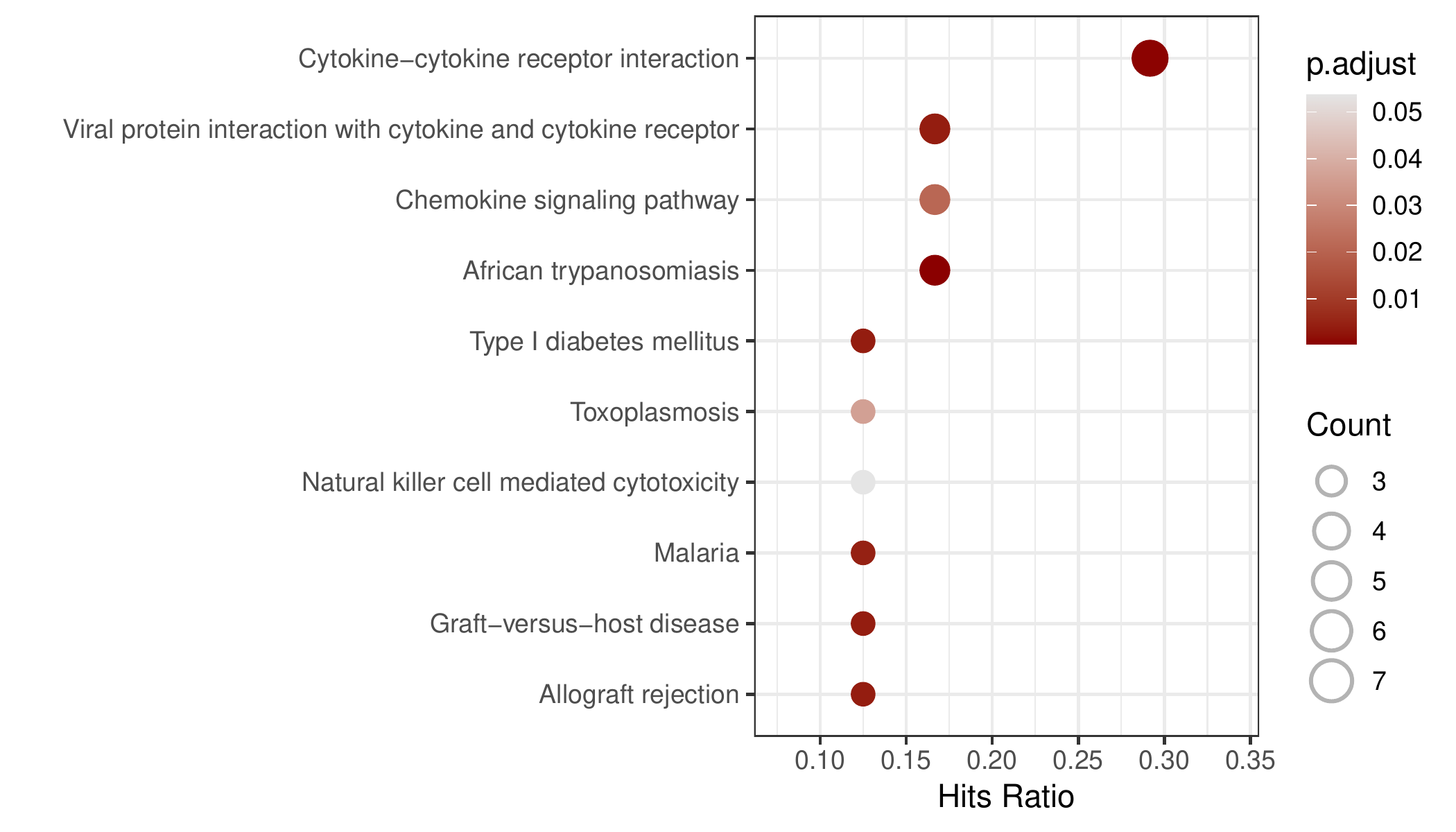
| Contrast | P val | Avg log2FC | Pct.1 | Pct.2 |
| --- | --- | --- | --- | --- |
| T Sepsis vs T Con... | 0 | 1.188 | 0.479 | 0.209 |
| T Sepsis vs T Con... | 1.452e-261 | 1.018 | 0.365 | 0.161 |
| T Sepsis vs T Con... | 1.821e-119 | 1.874 | 0.091 | 0.024 |
| T Sepsis vs T Con... | 2.257e-97 | 1.034 | 0.095 | 0.03 |
| T Sepsis vs T Con... | 8.143e-77 | 1.95 | 0.05 | 0.011 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.1.0\_Seurat\_单细胞数据分析\_(SEPSIS)/SEPSIS-DEGs-of-the-contrasts.csv)**

## 3.2 ClusterProfiler 富集分析 (BULK)

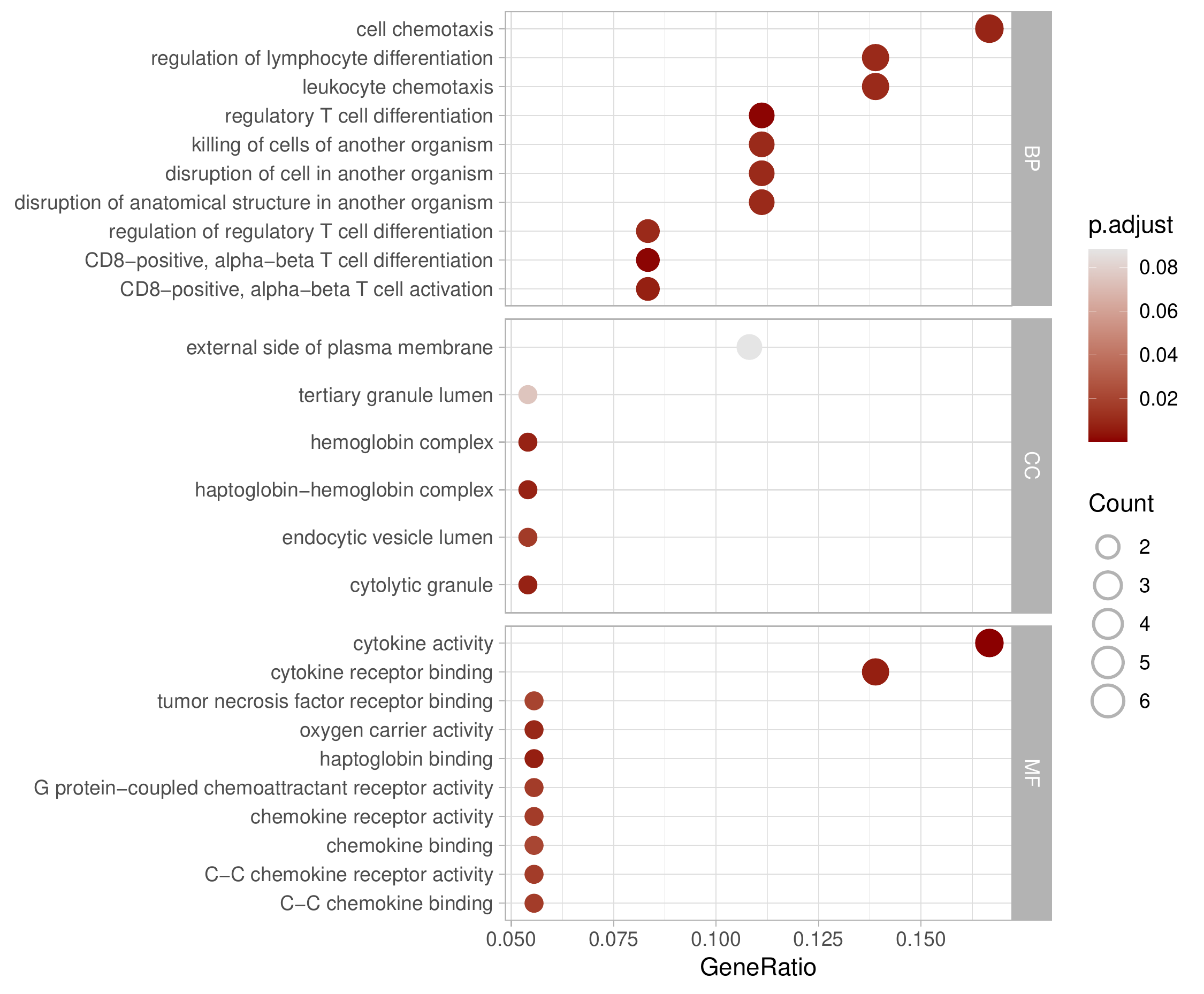
对基因集 (T\_Sepsis\_vs\_T\_Control, 来自于Seurat 单细胞数据分析[Section: SEPSIS]) 进行ClusterProfiler 富集分析。

Fig. 为 GO 富集分析气泡图。 Fig. 为 GO 富集分析气泡图。 Tab. 为 KEGG 富集分析统计表。 Tab. 为 GO 富集分析统计表。



**Fig.** **9** BULK KEGG enrichment

**(File path: Figure+Table/3.2.0\_ClusterProfiler\_富集分析\_(BULK)/BULK-KEGG-enrichment.pdf)**



**Fig.** **10** BULK GO enrichment

**(File path: Figure+Table/3.2.0\_ClusterProfiler\_富集分析\_(BULK)/BULK-GO-enrichment.pdf)**

**Tab.** **2** BULK KEGG enrichment data

| Category | Subcategory | ID | Description | GeneRatio |
| --- | --- | --- | --- | --- |
| Human Diseases | Infectious diseas... | Hsa05143 | African trypanoso... | 4/24 |
| Environmental Inf... | Signaling molecul... | Hsa04060 | Cytokine-cytokine... | 7/24 |
| Environmental Inf... | Signaling molecul... | Hsa04061 | Viral protein int... | 4/24 |
| Human Diseases | Immune disease | Hsa05330 | Allograft rejection | 3/24 |
| Human Diseases | Endocrine and met... | Hsa04940 | Type I diabetes m... | 3/24 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.2.0\_ClusterProfiler\_富集分析\_(BULK)/BULK-KEGG-enrichment-data.xlsx)**

**Tab.** **3** BULK GO enrichment data

| Ont | ID | Description | GeneRatio | BgRatio |
| --- | --- | --- | --- | --- |
| BP | GO:0045066 | Regulatory T cell... | 4/36 | 42/18986 |
| BP | GO:0043374 | CD8-positive, alp... | 3/36 | 15/18986 |
| BP | GO:0036037 | CD8-positive, alp... | 3/36 | 28/18986 |
| BP | GO:0060326 | Cell chemotaxis | 6/36 | 328/18986 |
| BP | GO:0045589 | Regulation of reg... | 3/36 | 38/18986 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.2.0\_ClusterProfiler\_富集分析\_(BULK)/BULK-GO-enrichment-data.xlsx)**

## 3.3 Bulk RNA-seq (Sepsis vs Non-infectious)

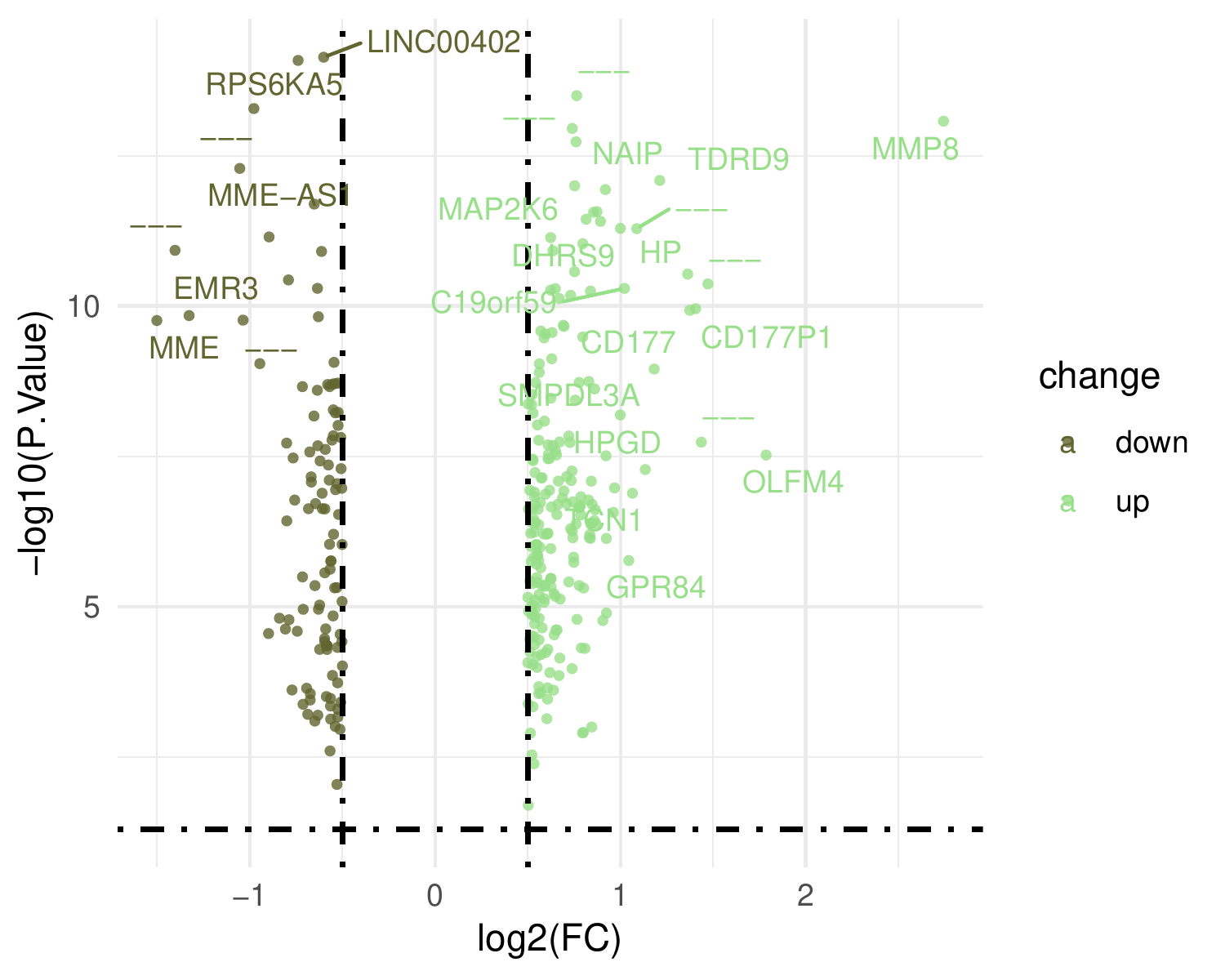
### 3.3.1 GEO 数据获取 (PB\_GSE134347)

以 GEOquery 获取 GSE134347 的数据信息。

### 3.3.2 Limma 差异分析 (PB\_GSE134347)

匹配 group 中包含“sepsis|noninfectious”的描述，最终得到 215 例数据。样本分组：noninfectious (n=59) , sepsis (n=156) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：sepsis vs noninfectious。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=126) , down (n=56)

Fig. 为 sepsis - noninfectious 差异分析火山图。 Tab. 为 sepsis - noninfectious 差异分析统计表格。



**Fig.** **11** PB GSE134347 sepsis vs noninfectious

**(File path: Figure+Table/3.3.2\_Limma\_差异分析\_(PB\_GSE134347)/PB-GSE134347-sepsis-vs-noninfectious.pdf)**

* P.Value cut-off: 0.05
* Log2(FC) cut-off: 0.5

**(See: Figure+Table/3.3.2\_Limma\_差异分析\_(PB\_GSE134347)/PB-GSE134347-sepsis-vs-noninfectious-content)**

**Tab.** **4** PB GSE134347 data sepsis vs noninfectious

| GENE SYMBOL | LogFC | P.Value | Rownames | ID |
| --- | --- | --- | --- | --- |
| LINC00402 | -0.6011 | 7.289e-15 | TC13000278.hg.1 | TC13000278.hg.1 |
| RPS6KA5 | -0.739 | 8.23e-15 | TC14001434.hg.1 | TC14001434.hg.1 |
| --- | 0.7637 | 3.176e-14 | TC05000307.hg.1 | TC05000307.hg.1 |
| --- | -0.9777 | 5.203e-14 | TC06002127.hg.1 | TC06002127.hg.1 |
| MMP8 | 2.743 | 8.448e-14 | TC11002231.hg.1 | TC11002231.hg.1 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.3.2\_Limma\_差异分析\_(PB\_GSE134347)/PB-GSE134347-data-sepsis-vs-noninfectious.tsv)**

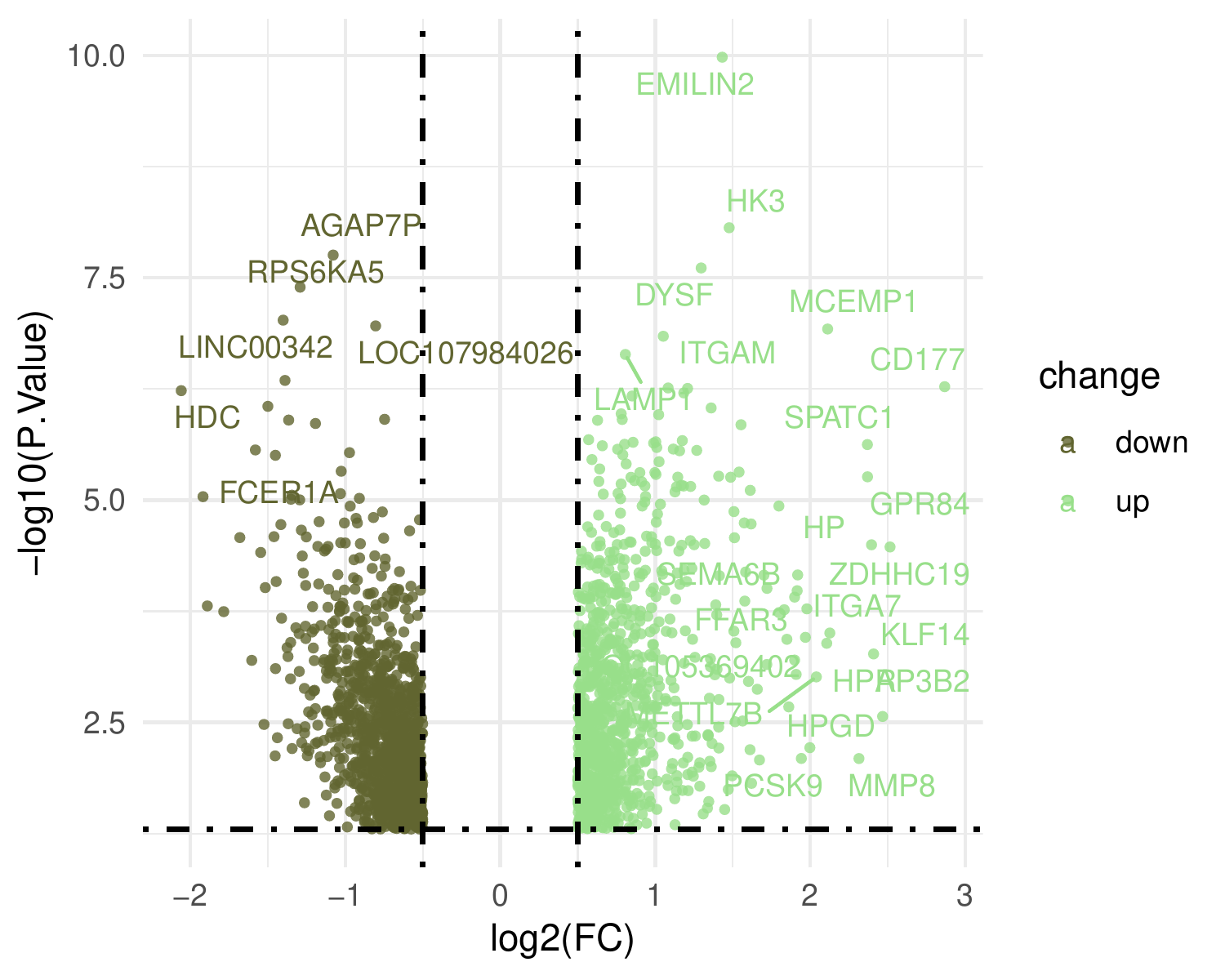
### 3.3.3 GEO 数据获取 (PB\_GSE63042)

以 GEOquery 获取 GSE63042 的数据信息。

### 3.3.4 Limma 差异分析 (PB\_GSE63042)

样本分组：Sepsis (n=96) , SIRS (n=21) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Sepsis vs SIRS。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=973) , down (n=1217) (SIRS, systemic inflammatory response syndrome，文中的 SIRS 指的是 Infection-Negative Systemic Inflammation)

Fig. 为 Sepsis - SIRS 差异分析火山图。 Tab. 为 Sepsis - SIRS 差异分析统计表格。



**Fig.** **12** PB GSE63042 Sepsis vs SIRS

**(File path: Figure+Table/3.3.4\_Limma\_差异分析\_(PB\_GSE63042)/PB-GSE63042-Sepsis-vs-SIRS.pdf)**

* P.Value cut-off: 0.05
* Log2(FC) cut-off: 0.5

**(See: Figure+Table/3.3.4\_Limma\_差异分析\_(PB\_GSE63042)/PB-GSE63042-Sepsis-vs-SIRS-content)**

**Tab.** **5** PB GSE63042 data Sepsis vs SIRS

| Hgnc symbol | LogFC | P.Value | Rownames | GeneID |
| --- | --- | --- | --- | --- |
| EMILIN2 | 1.433 | 1.046e-10 | 84034 | 84030 |
| HK3 | 1.477 | 8.651e-09 | 3101 | 3101 |
| AGAP7P | -1.078 | 1.755e-08 | 653268 | 653300 |
| DYSF | 1.297 | 2.451e-08 | 8291 | 8291 |
| RPS6KA5 | -1.292 | 4.028e-08 | 9252 | 9252 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.3.4\_Limma\_差异分析\_(PB\_GSE63042)/PB-GSE63042-data-Sepsis-vs-SIRS.tsv)**

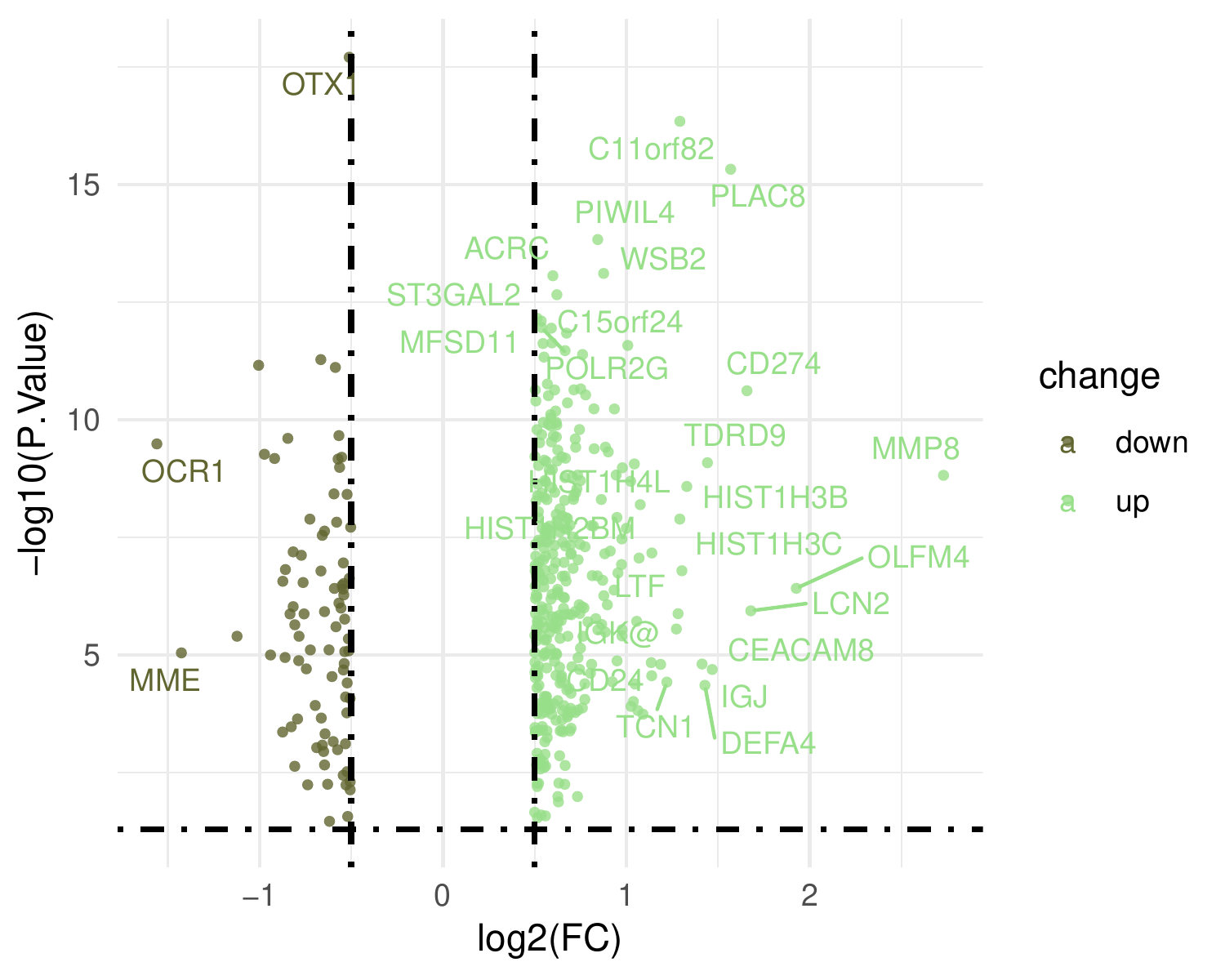
### 3.3.5 GEO 数据获取 (PB\_GSE74224)

以 GEOquery 获取 GSE74224 的数据信息。

### 3.3.6 Limma 差异分析 (PB\_GSE74224)

样本分组：Sepsis (n=74) , SIRS (n=31) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Sepsis vs SIRS。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=344) , down (n=79)

Fig. 为 Sepsis - SIRS 差异分析火山图。 Tab. 为 Sepsis - SIRS 差异分析统计表格。



**Fig.** **13** PB GSE74224 Sepsis vs SIRS

**(File path: Figure+Table/3.3.6\_Limma\_差异分析\_(PB\_GSE74224)/PB-GSE74224-Sepsis-vs-SIRS.pdf)**

* P.Value cut-off: 0.05
* Log2(FC) cut-off: 0.5

**(See: Figure+Table/3.3.6\_Limma\_差异分析\_(PB\_GSE74224)/PB-GSE74224-Sepsis-vs-SIRS-content)**

**Tab.** **6** PB GSE74224 data Sepsis vs SIRS

| GENE SYMBOL | LogFC | P.Value | Rownames | ID |
| --- | --- | --- | --- | --- |
| OTX1 | -0.5107 | 1.965e-18 | 2485112 | 2485000 |
| C11orf82 | 1.292 | 4.518e-17 | 3342426 | 3342000 |
| PLAC8 | 1.569 | 4.724e-16 | 2775909 | 2776000 |
| PIWIL4 | 0.8442 | 1.473e-14 | 3345157 | 3345000 |
| WSB2 | 0.8763 | 7.73e-14 | 3473727 | 3474000 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.3.6\_Limma\_差异分析\_(PB\_GSE74224)/PB-GSE74224-data-Sepsis-vs-SIRS.tsv)**

## 3.4 Bulk RNA-seq (Sepsis vs Healthy)

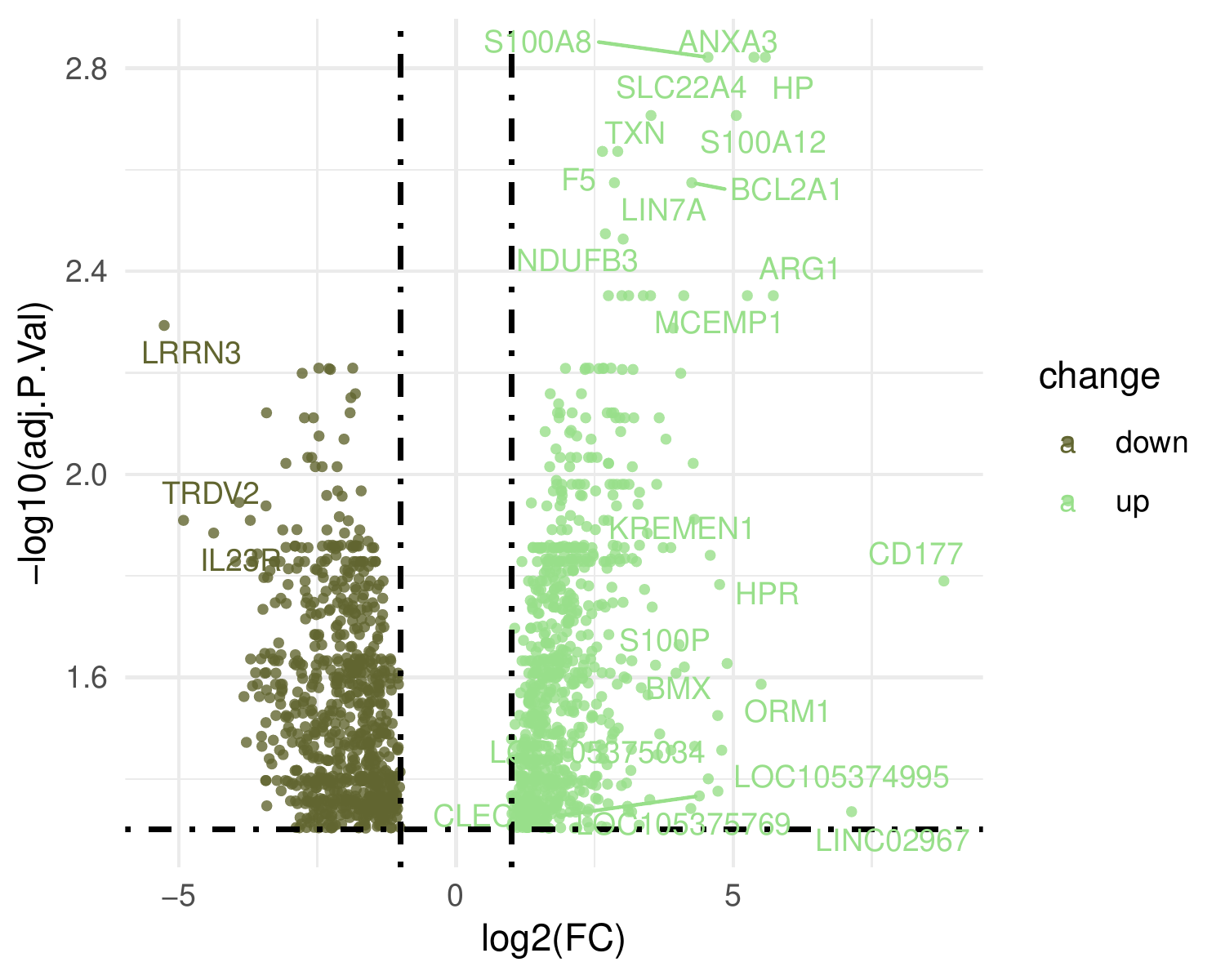
### 3.4.1 GEO 数据获取 (PB\_GSE243217)

以 GEOquery 获取 GSE243217 的数据信息。

### 3.4.2 Limma 差异分析 (PB\_GSE243217)

匹配 group 中包含“healthy\_donor|Sepsis”的描述，最终得到 11 例数据。样本分组：healthy\_donor (n=3) , Sepsis (n=8) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Sepsis vs healthy\_donor。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=775) , down (n=782)

Fig. 为 Sepsis - healthy\_donor 差异分析火山图。 Tab. 为 Sepsis - healthy\_donor 差异分析统计表格。



**Fig.** **14** BL GSE243217 Sepsis vs healthy donor

**(File path: Figure+Table/3.4.2\_Limma\_差异分析\_(PB\_GSE243217)/BL-GSE243217-Sepsis-vs-healthy-donor.pdf)**

* adj.P.Val cut-off: 0.05
* Log2(FC) cut-off: 1

**(See: Figure+Table/3.4.2\_Limma\_差异分析\_(PB\_GSE243217)/BL-GSE243217-Sepsis-vs-healthy-donor-content)**

**Tab.** **7** BL GSE243217 data Sepsis vs healthy donor

| Hgnc symbol | LogFC | Adj.P.Val | Rownames | GeneID |
| --- | --- | --- | --- | --- |
| HP | 5.58 | 0.001508 | 3240 | 3240 |
| ANXA3 | 5.376 | 0.001508 | 306 | 306 |
| S100A8 | 4.546 | 0.001508 | 6279 | 6279 |
| SLC22A4 | 3.518 | 0.001964 | 6583 | 6583 |
| S100A12 | 5.056 | 0.001964 | 6283 | 6283 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.4.2\_Limma\_差异分析\_(PB\_GSE243217)/BL-GSE243217-data-Sepsis-vs-healthy-donor.tsv)**

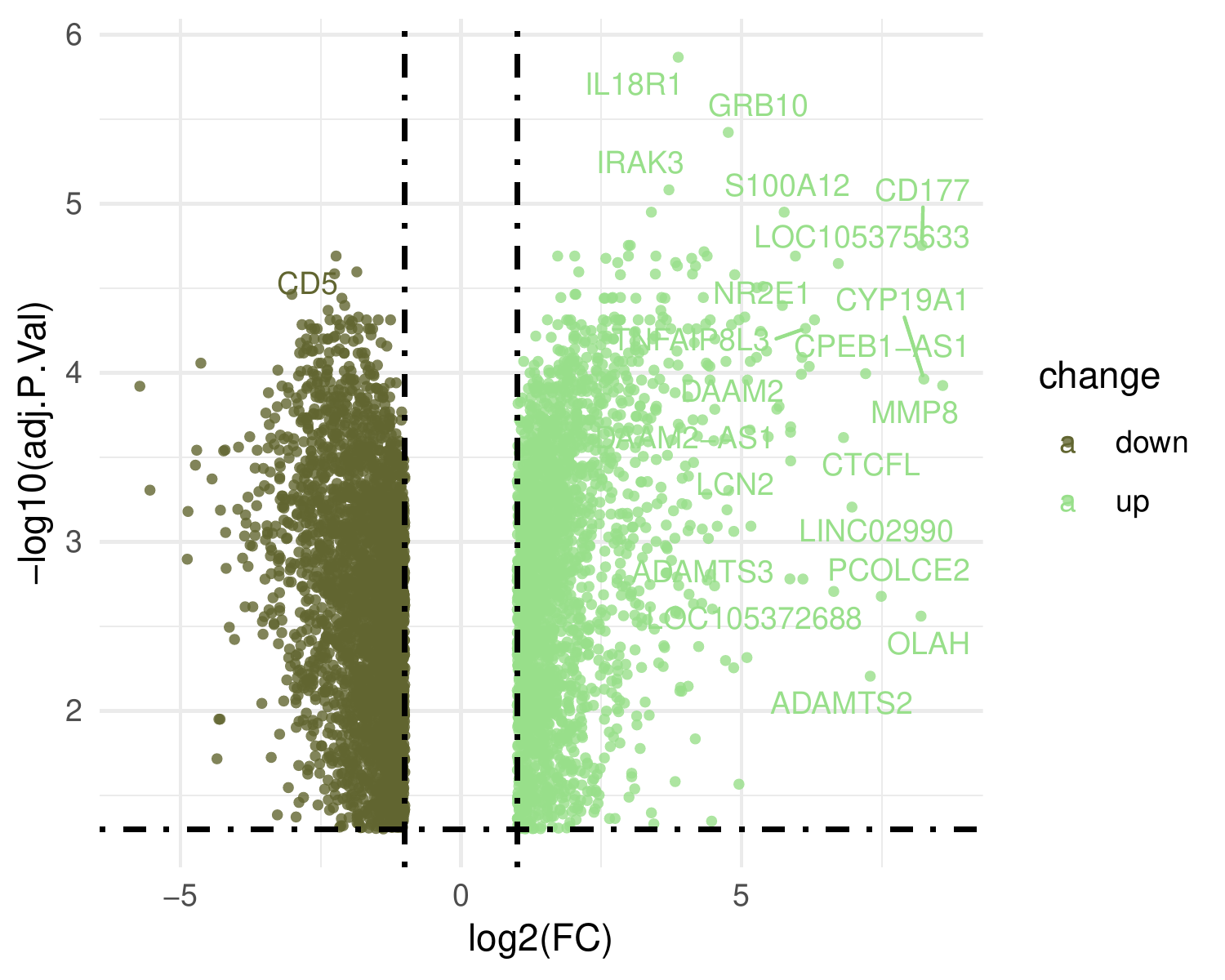
### 3.4.3 GEO 数据获取 (PB\_GSE232404)

以 GEOquery 获取 GSE232404 的数据信息。

### 3.4.4 Limma 差异分析 (PB\_GSE232404)

样本分组：HC (n=5) , SAKI (n=5) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：SAKI vs HC。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=2410) , down (n=3099)

Fig. 为 SAKI - HC 差异分析火山图。 Tab. 为 SAKI - HC 差异分析统计表格。



**Fig.** **15** PB GSE232404 SAKI vs HC

**(File path: Figure+Table/3.4.4\_Limma\_差异分析\_(PB\_GSE232404)/PB-GSE232404-SAKI-vs-HC.pdf)**

* adj.P.Val cut-off: 0.05
* Log2(FC) cut-off: 1

**(See: Figure+Table/3.4.4\_Limma\_差异分析\_(PB\_GSE232404)/PB-GSE232404-SAKI-vs-HC-content)**

**Tab.** **8** PB GSE232404 data SAKI vs HC

| Hgnc symbol | LogFC | Adj.P.Val | Rownames | GeneID |
| --- | --- | --- | --- | --- |
| IL18R1 | 3.87 | 1.361e-06 | 8809 | 8809 |
| GRB10 | 4.761 | 3.789e-06 | 2887 | 2887 |
| IRAK3 | 3.704 | 8.285e-06 | 11213 | 11210 |
| S100A12 | 5.757 | 1.122e-05 | 6283 | 6283 |
| MIR6502 | 3.391 | 1.122e-05 | 102465249 | 102500000 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.4.4\_Limma\_差异分析\_(PB\_GSE232404)/PB-GSE232404-data-SAKI-vs-HC.tsv)**

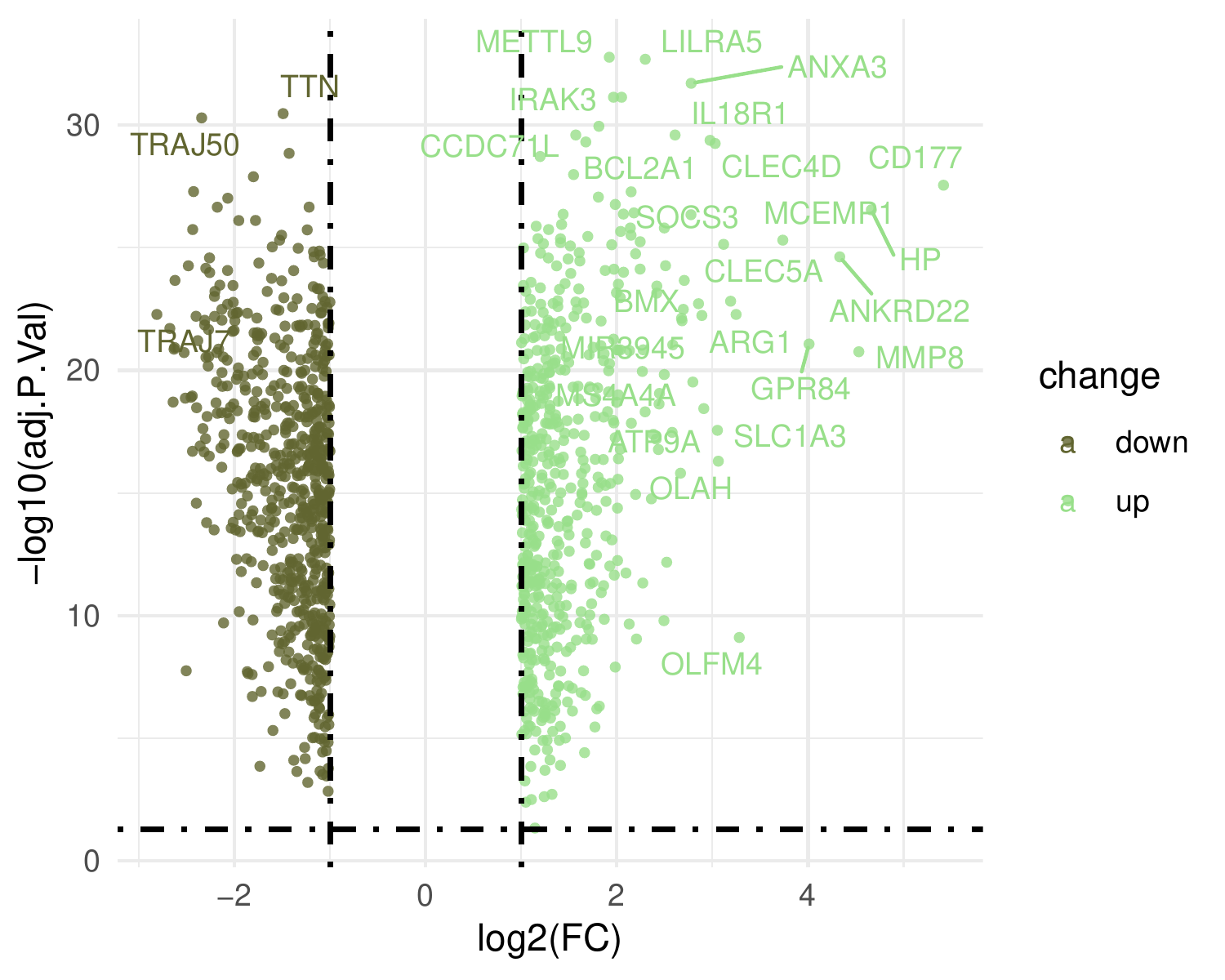
### 3.4.5 GEO 数据获取 (PB\_GSE69063)

以 GEOquery 获取 GSE69063 的数据信息。

### 3.4.6 Limma 差异分析 (PB\_GSE69063)

匹配 group 中包含“Sepsis|Healthy\_control”的描述，最终得到 90 例数据。样本分组：Healthy\_control (n=33) , Sepsis (n=57) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Sepsis vs Healthy\_control。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=538) , down (n=655)

Fig. 为 Sepsis - Healthy\_control 差异分析火山图。 Tab. 为 Sepsis - Healthy\_control 差异分析统计表格。



**Fig.** **16** PB GSE69063 Sepsis vs Healthy control

**(File path: Figure+Table/3.4.6\_Limma\_差异分析\_(PB\_GSE69063)/PB-GSE69063-Sepsis-vs-Healthy-control.pdf)**

* adj.P.Val cut-off: 0.05
* Log2(FC) cut-off: 1

**(See: Figure+Table/3.4.6\_Limma\_差异分析\_(PB\_GSE69063)/PB-GSE69063-Sepsis-vs-Healthy-control-content)**

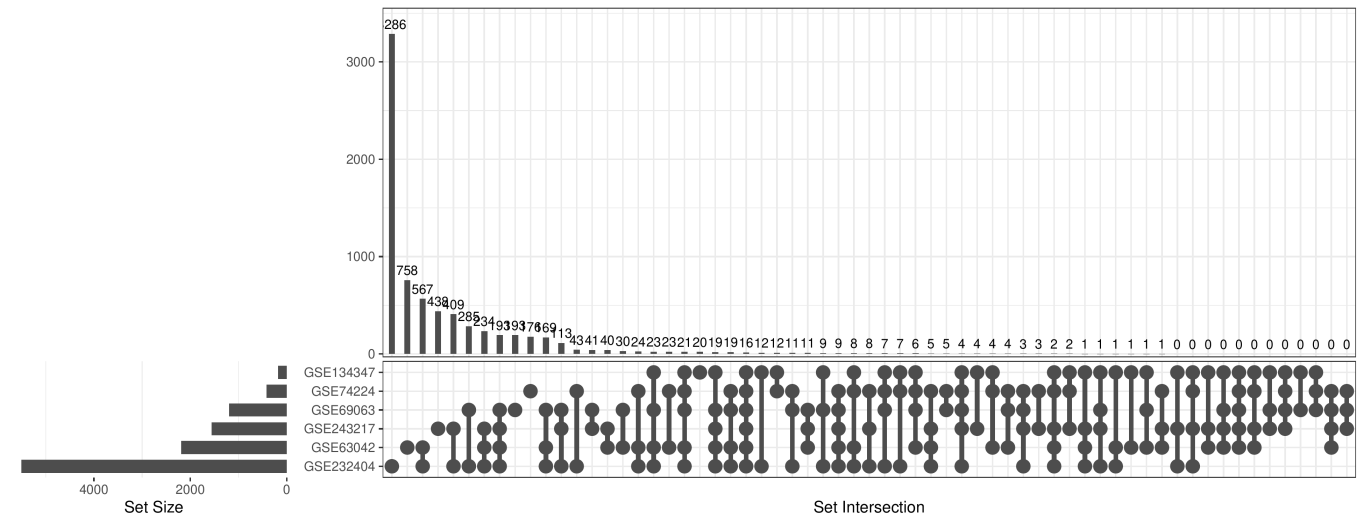
**Tab.** **9** PB GSE69063 data Sepsis vs Healthy control

| SYMBOL | LogFC | Adj.P.Val | Rownames | ID |
| --- | --- | --- | --- | --- |
| METTL9 | 1.922 | 1.712e-33 | 51108 at | 51108 at |
| LILRA5 | 2.298 | 2.068e-33 | 353514 at | 353514 at |
| ANXA3 | 2.779 | 1.979e-32 | 306 at | 306 at |
| FCAR | 2.05 | 7.326e-32 | 2204 at | 2204 at |
| S100A12 | 1.966 | 7.326e-32 | 6283 at | 6283 at |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.4.6\_Limma\_差异分析\_(PB\_GSE69063)/PB-GSE69063-data-Sepsis-vs-Healthy-control.csv)**

## 3.5 汇总: GSE134347 + GSE232404 + GSE243217 + GSE63042 + GSE69063 + GSE74224 (BULK)

不同数据集的差异表达基因的交集。 Fig. 为 GSE134347, GSE232404, GSE243217, GSE63042, GSE69063, GSE74224 各自交集。 Tab. 为多个数据集的元数据信息 (分组信息)



**Fig.** **17** BULK Intersection of GSE134347 with GSE232404 with GSE243217 with GSE63042 with GSE69063 with GSE74224

**(File path: Figure+Table/3.5.0\_汇总:\_GSE134347\_+\_GSE232404\_+\_GSE243217\_+\_GSE63042\_+\_GSE69063\_+\_GSE74224\_(BULK)/BULK-Intersection-of-GSE134347-with-GSE232404-with-GSE243217-with-GSE63042-with-GSE69063-with-GSE74224.pdf)**

* All\_intersection: TDRD9, MAP2K6, STOM, NLRC4, HP, CD177, HK3, ZNF438, LILRA5, ANXA3, VNN1, TLR5, BPI, GYG1, ALPL, FOLR3

**(See: Figure+Table/3.5.0\_汇总:\_GSE134347\_+\_GSE232404\_+\_GSE243217\_+\_GSE63042\_+\_GSE69063\_+\_GSE74224\_(BULK)/BULK-Intersection-of-GSE134347-with-GSE232404-with-GSE243217-with-GSE63042-with-GSE69063-with-GSE74224-content)**

**Tab.** **10** Metadata of mutiple datasets

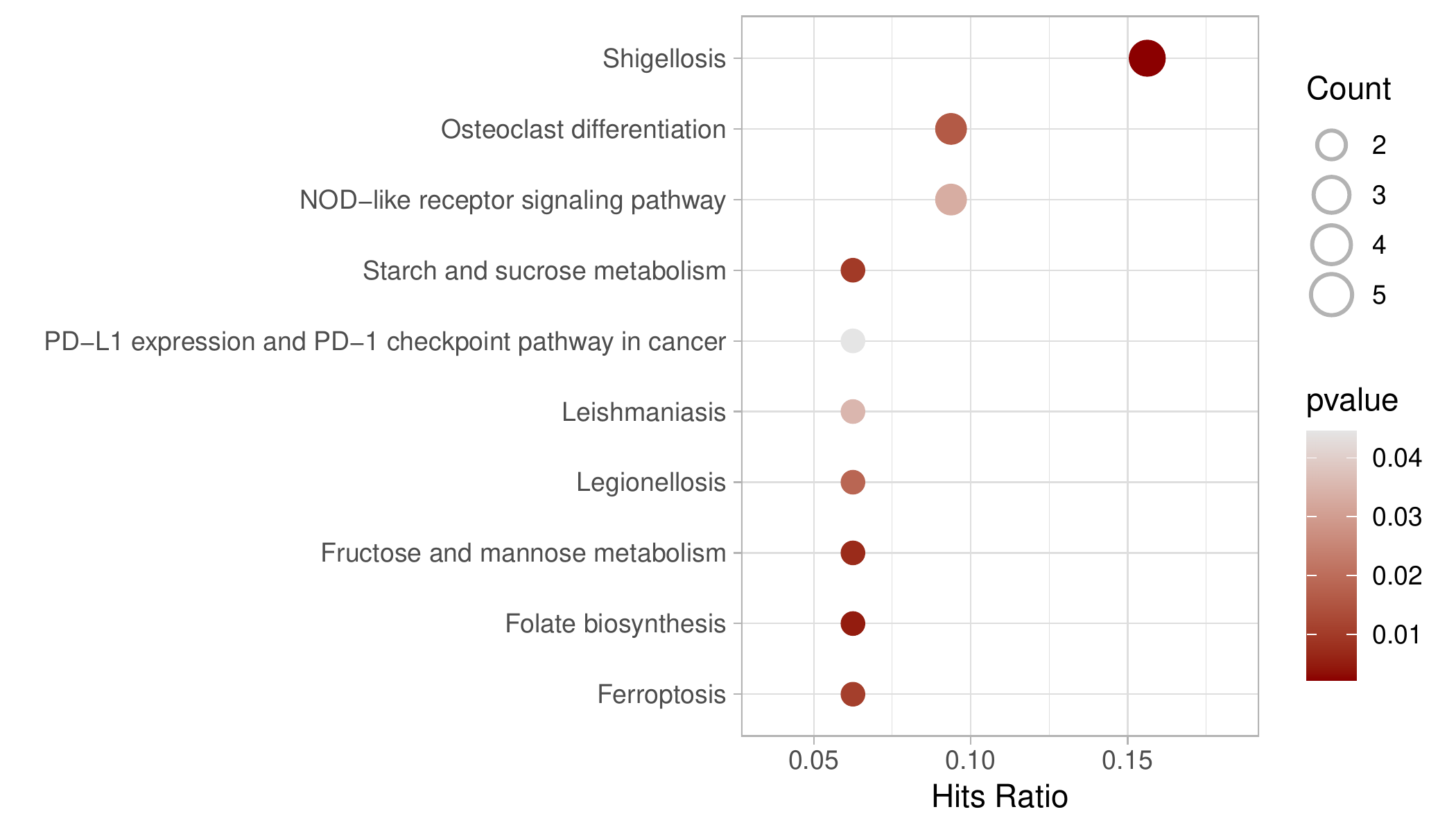
| Project | Versus | Group | Tissue |
| --- | --- | --- | --- |
| GSE134347 | Sepsis - noninfec... | SI | Blood |
| GSE232404 | SAKI - HC | SH | Blood |
| GSE243217 | Sepsis - healthy ... | SH | Blood |
| GSE63042 | Sepsis - SIRS | SI | Blood |
| GSE69063 | Sepsis - Healthy ... | SH | Blood |
| ... | ... | ... | ... |

**(File path: Figure+Table/3.5.0\_汇总:\_GSE134347\_+\_GSE232404\_+\_GSE243217\_+\_GSE63042\_+\_GSE69063\_+\_GSE74224\_(BULK)/Metadata-of-mutiple-datasets.csv)**

## 3.6 Sepsis vs Non-infectious 与 Sepsis vs Healthy

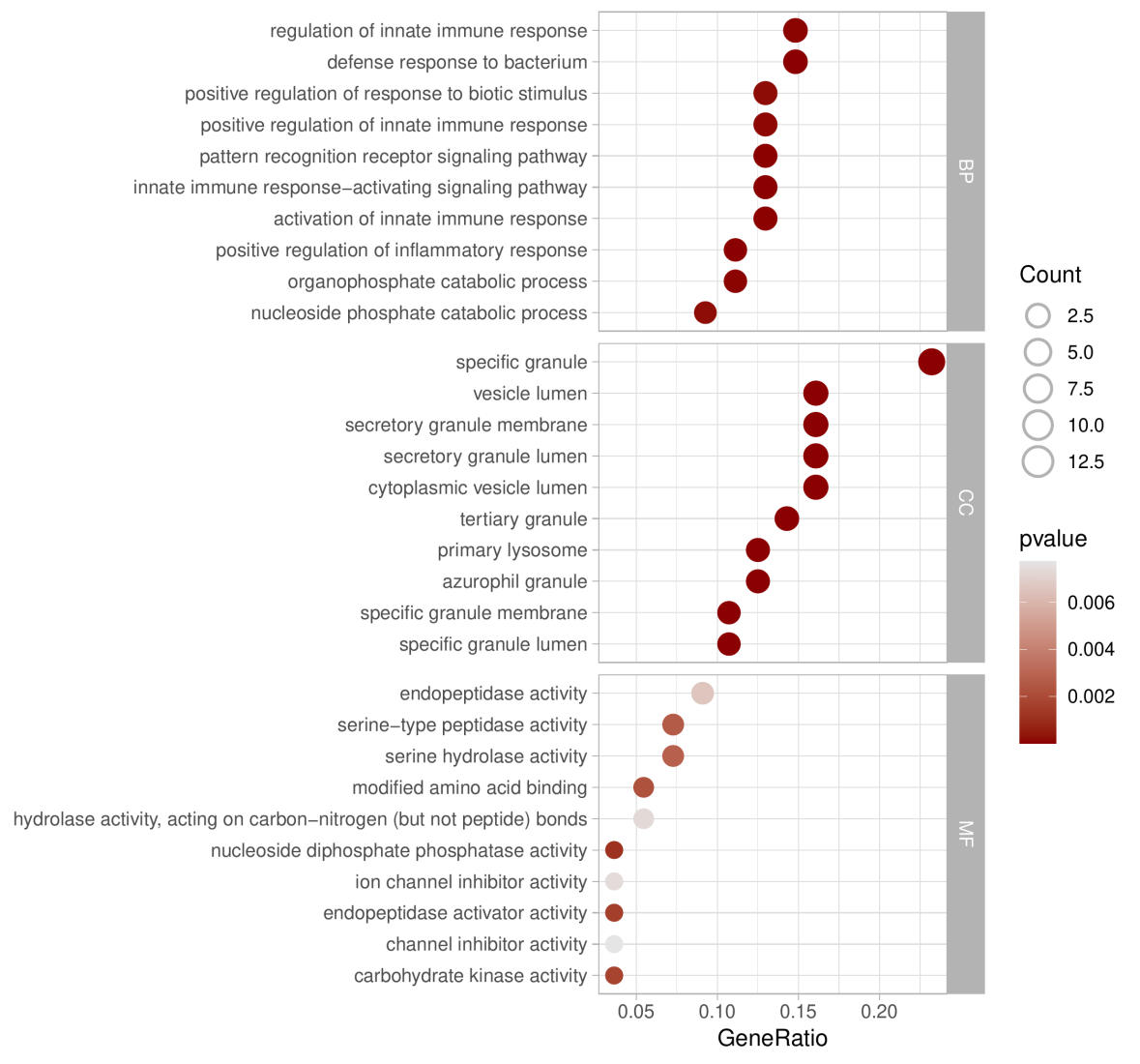
### 3.6.1 ClusterProfiler 富集分析 (SI)

Fig. 为 GO 富集分析气泡图。 Fig. 为 GO 富集分析气泡图。 Tab. 为 KEGG 富集分析统计表。 Tab. 为 GO 富集分析统计表。



**Fig.** **18** SI KEGG enrichment

**(File path: Figure+Table/3.6.1\_ClusterProfiler\_富集分析\_(SI)/SI-KEGG-enrichment.pdf)**



**Fig.** **19** SI GO enrichment

**(File path: Figure+Table/3.6.1\_ClusterProfiler\_富集分析\_(SI)/SI-GO-enrichment.pdf)**

**Tab.** **11** SI KEGG enrichment data

| Category | Subcategory | ID | Description | GeneRatio |
| --- | --- | --- | --- | --- |
| Human Diseases | Infectious diseas... | Hsa05131 | Shigellosis | 5/32 |
| Metabolism | Metabolism of cof... | Hsa00790 | Folate biosynthesis | 2/32 |
| Metabolism | Carbohydrate meta... | Hsa00051 | Fructose and mann... | 2/32 |
| Metabolism | Carbohydrate meta... | Hsa00500 | Starch and sucros... | 2/32 |
| Cellular Processes | Cell growth and d... | Hsa04216 | Ferroptosis | 2/32 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6.1\_ClusterProfiler\_富集分析\_(SI)/SI-KEGG-enrichment-data.xlsx)**

**Tab.** **12** SI GO enrichment data

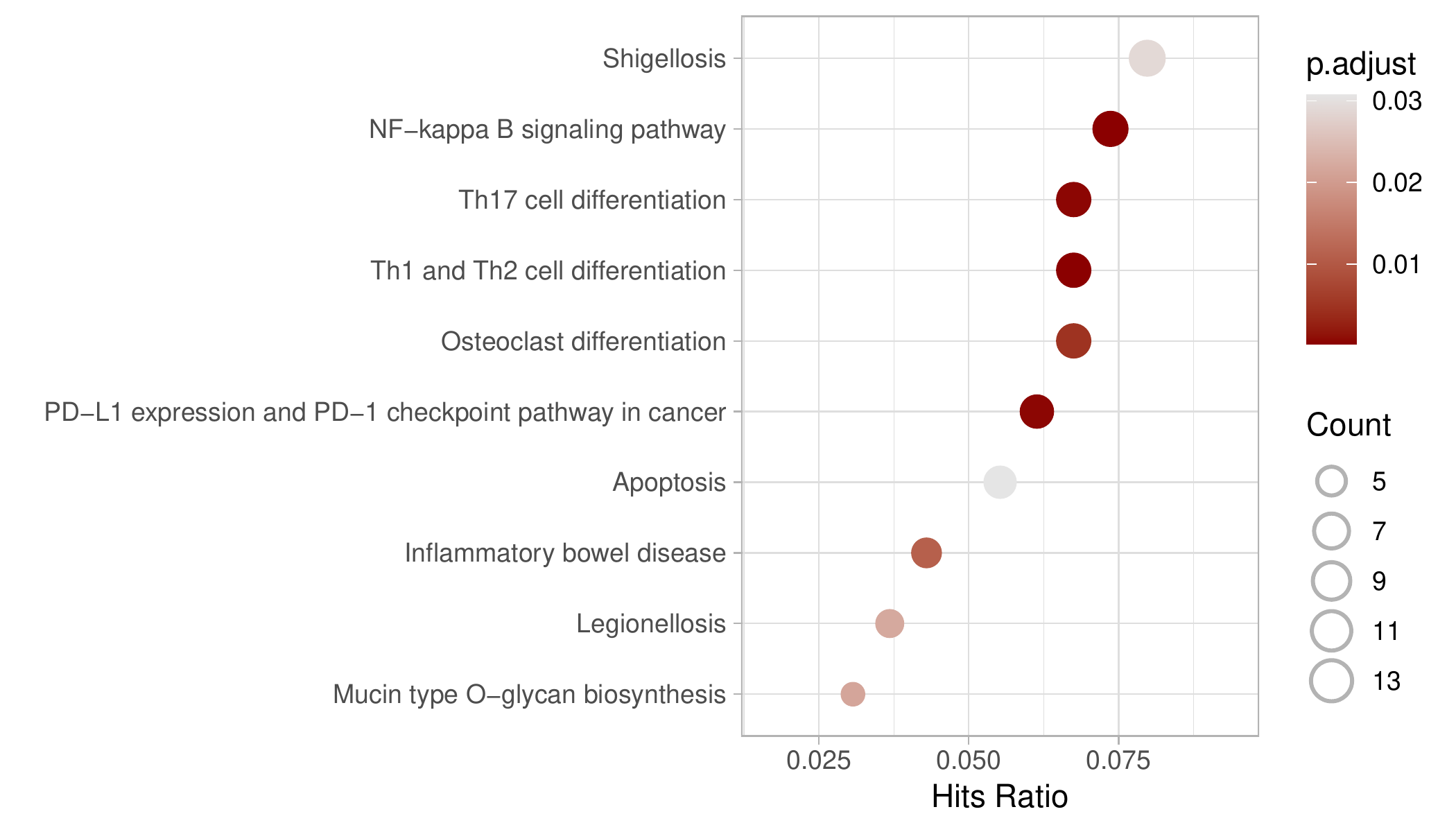
| Ont | ID | Description | GeneRatio | BgRatio |
| --- | --- | --- | --- | --- |
| BP | GO:0042742 | Defense response ... | 8/54 | 342/18986 |
| BP | GO:0050729 | Positive regulati... | 6/54 | 162/18986 |
| BP | GO:0002221 | Pattern recogniti... | 7/54 | 278/18986 |
| BP | GO:0002758 | Innate immune res... | 7/54 | 297/18986 |
| BP | GO:0002218 | Activation of inn... | 7/54 | 317/18986 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6.1\_ClusterProfiler\_富集分析\_(SI)/SI-GO-enrichment-data.xlsx)**

### 3.6.2 ClusterProfiler 富集分析 (SH)

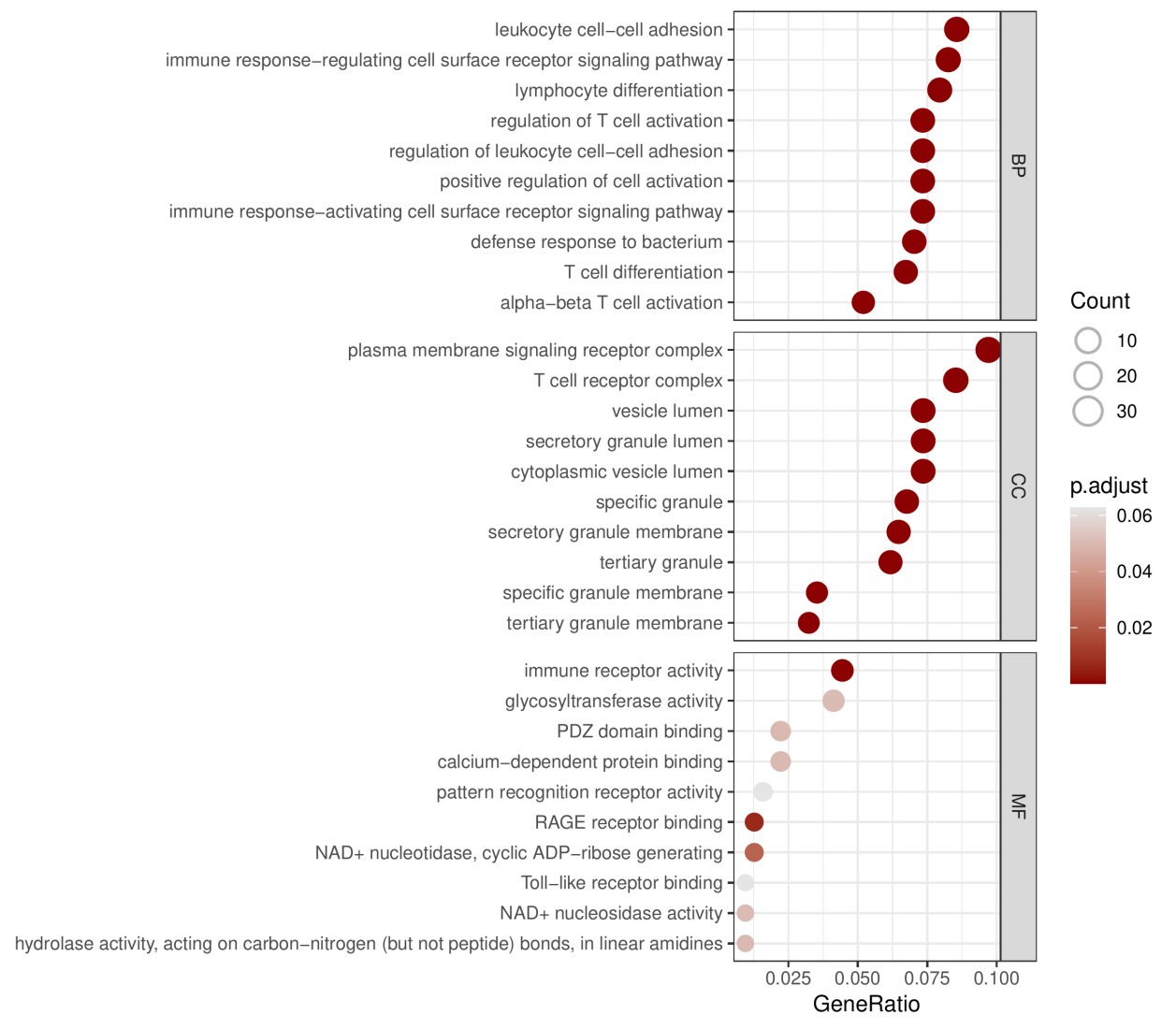
对基因集 (IRAK3, S100A12, CYP1B1, …[n = 358], 来自于 Venn 交集 (SH: GSE232404, GSE243217, GSE69063)[Section: BULK]) 进行ClusterProfiler 富集分析。

Fig. 为 GO 富集分析气泡图。 Fig. 为 GO 富集分析气泡图。 Tab. 为 KEGG 富集分析统计表。 Tab. 为 GO 富集分析统计表。



**Fig.** **20** SH KEGG enrichment

**(File path: Figure+Table/3.6.2\_ClusterProfiler\_富集分析\_(SH)/SH-KEGG-enrichment.pdf)**



**Fig.** **21** SH GO enrichment

**(File path: Figure+Table/3.6.2\_ClusterProfiler\_富集分析\_(SH)/SH-GO-enrichment.pdf)**

**Tab.** **13** SH KEGG enrichment data

| Category | Subcategory | ID | Description | GeneRatio |
| --- | --- | --- | --- | --- |
| Environmental Inf... | Signal transduction | Hsa04064 | NF-kappa B signal... | 12/163 |
| Organismal Systems | Immune system | Hsa04658 | Th1 and Th2 cell ... | 11/163 |
| Organismal Systems | Immune system | Hsa04659 | Th17 cell differe... | 11/163 |
| Human Diseases | Cancer: overview | Hsa05235 | PD-L1 expression ... | 10/163 |
| Organismal Systems | Development and r... | Hsa04380 | Osteoclast differ... | 11/163 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6.2\_ClusterProfiler\_富集分析\_(SH)/SH-KEGG-enrichment-data.xlsx)**

**Tab.** **14** SH GO enrichment data

| Ont | ID | Description | GeneRatio | BgRatio |
| --- | --- | --- | --- | --- |
| BP | GO:0002768 | Immune response-r... | 27/327 | 369/18986 |
| BP | GO:0007159 | Leukocyte cell-ce... | 28/327 | 430/18986 |
| BP | GO:0002429 | Immune response-a... | 24/327 | 336/18986 |
| BP | GO:0030098 | Lymphocyte differ... | 26/327 | 426/18986 |
| BP | GO:0042742 | Defense response ... | 23/327 | 342/18986 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6.2\_ClusterProfiler\_富集分析\_(SH)/SH-GO-enrichment-data.xlsx)**

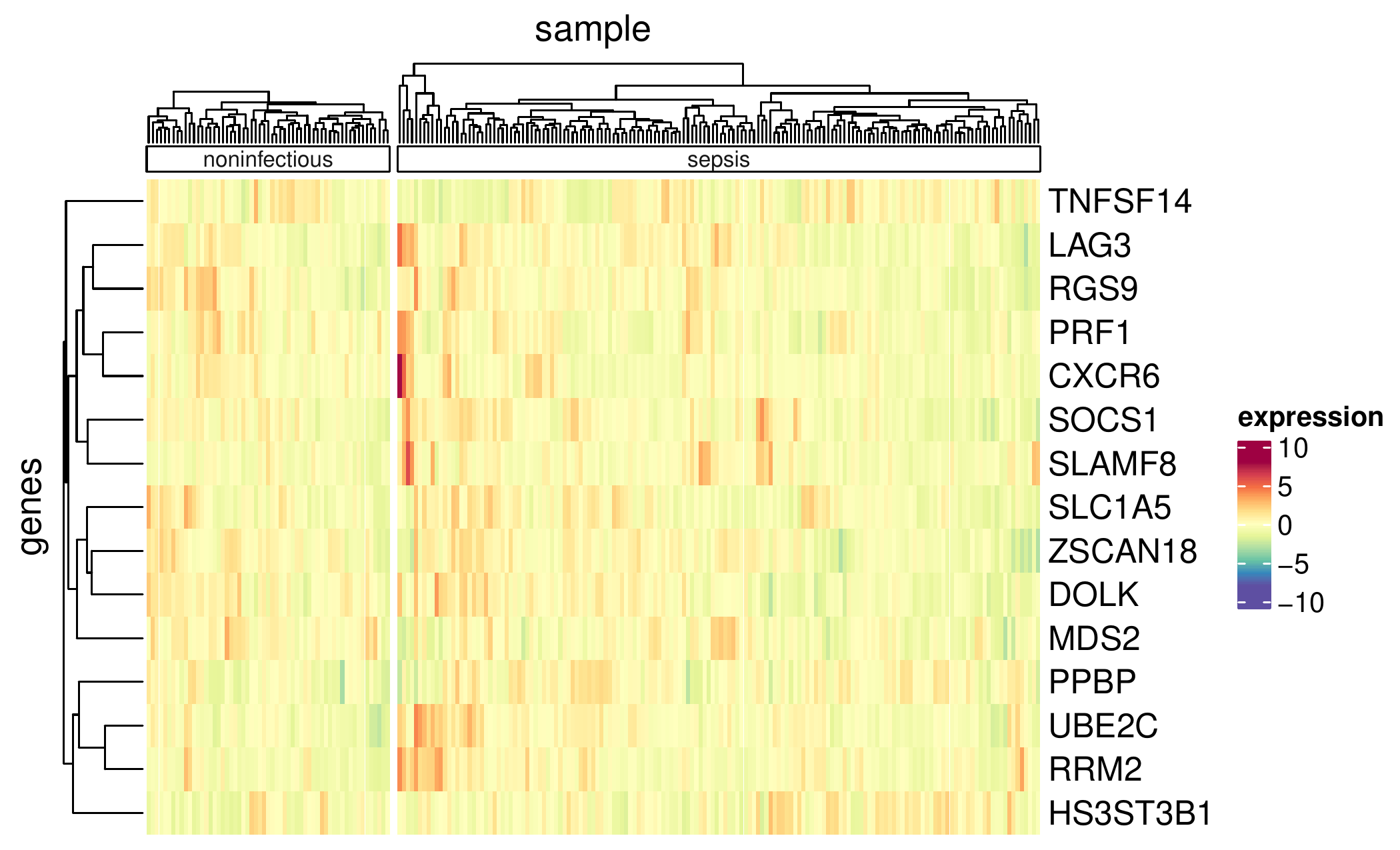
### 3.6.3 富集分析比较

上述 SI 交集与 SH 交集的富集分析，均表现出免疫行为的区别。 SH 的富集分析，富集的通路 (GO) 表现出 T 细胞行为的差异 (Fig. **[21](#SH-GO-enrichment)** )。 而 SI 的富集分析，主要体现对病菌的感染性反应 (Fig. **[19](#SI-GO-enrichment)** )。 可以推测，Sepsis 与 SIRS (Sepsis Negative) 的差异，可以体现在 T 细胞对病菌感染的差异行为。 因此，后续分析集中于 T 细胞。

## 3.7 Lasso 诊断模型建立 (BULK)

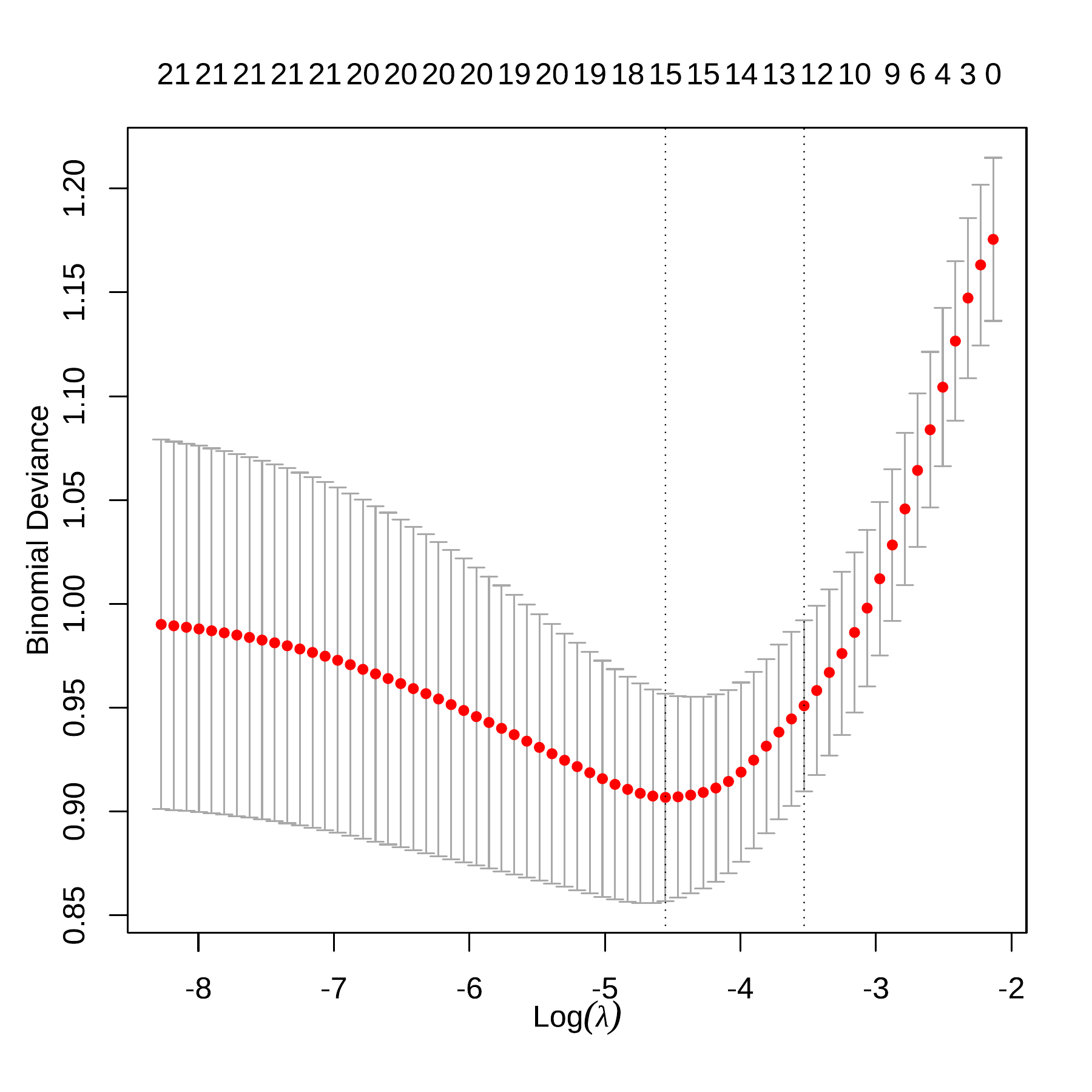
选定基因集 (T\_Sepsis\_vs\_T\_Control, 来自于Seurat 单细胞数据分析[Section: SEPSIS]) 。以 GSE134347 (sepsis - noninfectious) 创建诊断模型。使用 glmnet::cv.glmnet 作 10 倍交叉验证 (评估方式为 Binomial Deviance)，筛选 lambda 值。lambda.min, lambda.1se 值分别为 0.011, 0.029 (R 随机种子为 987456)。对应的特征数 (基因数) 分别为 15, 12。使用数据集 GSE134347 (SI), GSE232404 (SH), GSE243217 (SH), GSE63042 (SI), GSE69063 (SH), GSE74224 (SI) 验证。

Fig. 为特征基因在训练数据集 (GSE134347) 的表达热图。 Fig. 为模型的交叉验证曲线 (cross-validation curve)。 Fig. 为 lambda.min 下，模型预测 (以内部数据) 的 ROC 曲线。 Fig. 为 lambda.1se 下，模型预测 (以内部数据) 的 ROC 曲线。 Fig. 为 lambda.min 下各变量的系数。 Fig. 为 lambda.1se 下各变量的系数。 Fig. 为 lambda.min ROC 曲线。 Fig. 为 lambda.1se ROC 曲线。



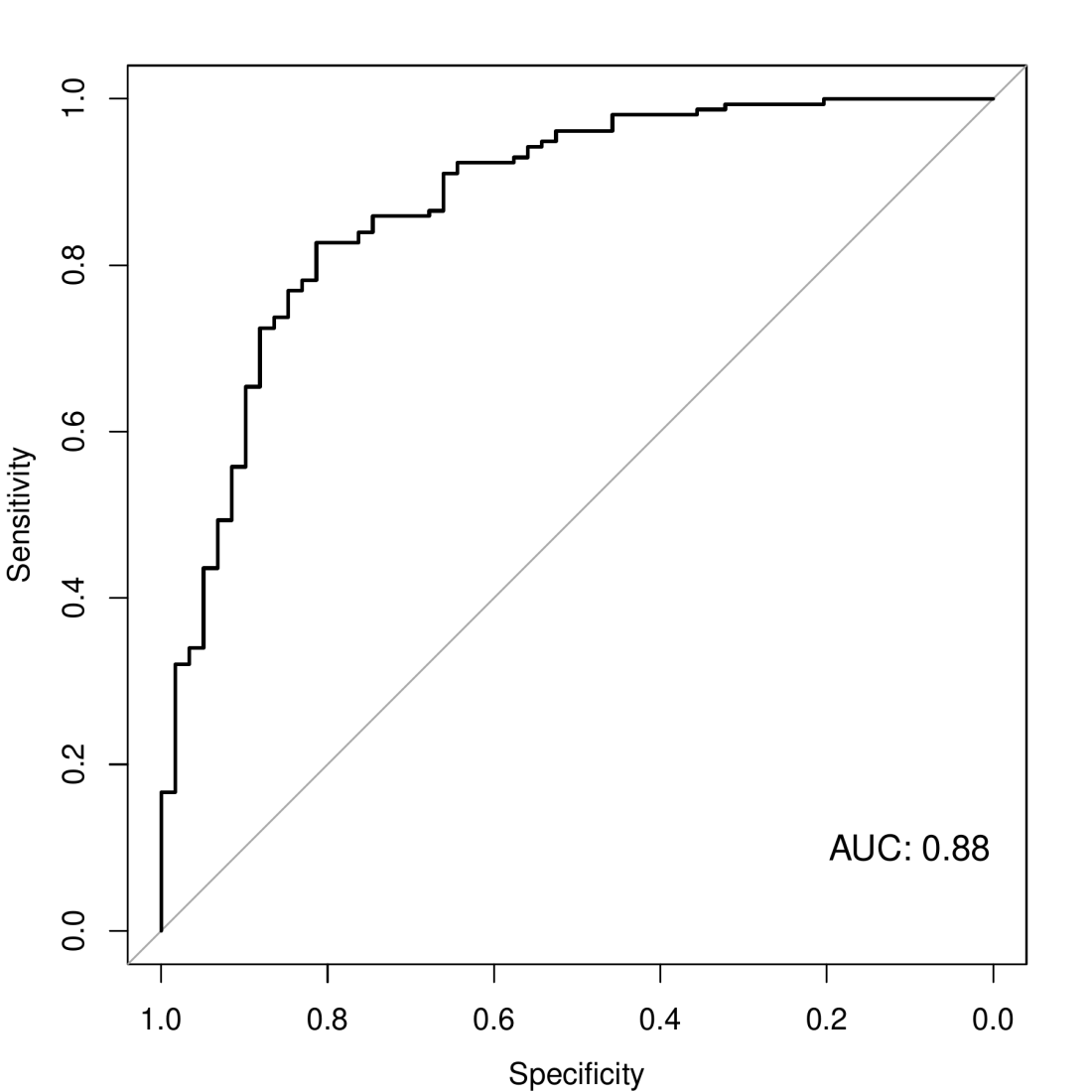
**Fig.** **22** Feature heatmap in validation dataset

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/Feature-heatmap-in-validation-dataset.pdf)**



**Fig.** **23** Lasso model

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/Lasso-model.pdf)**



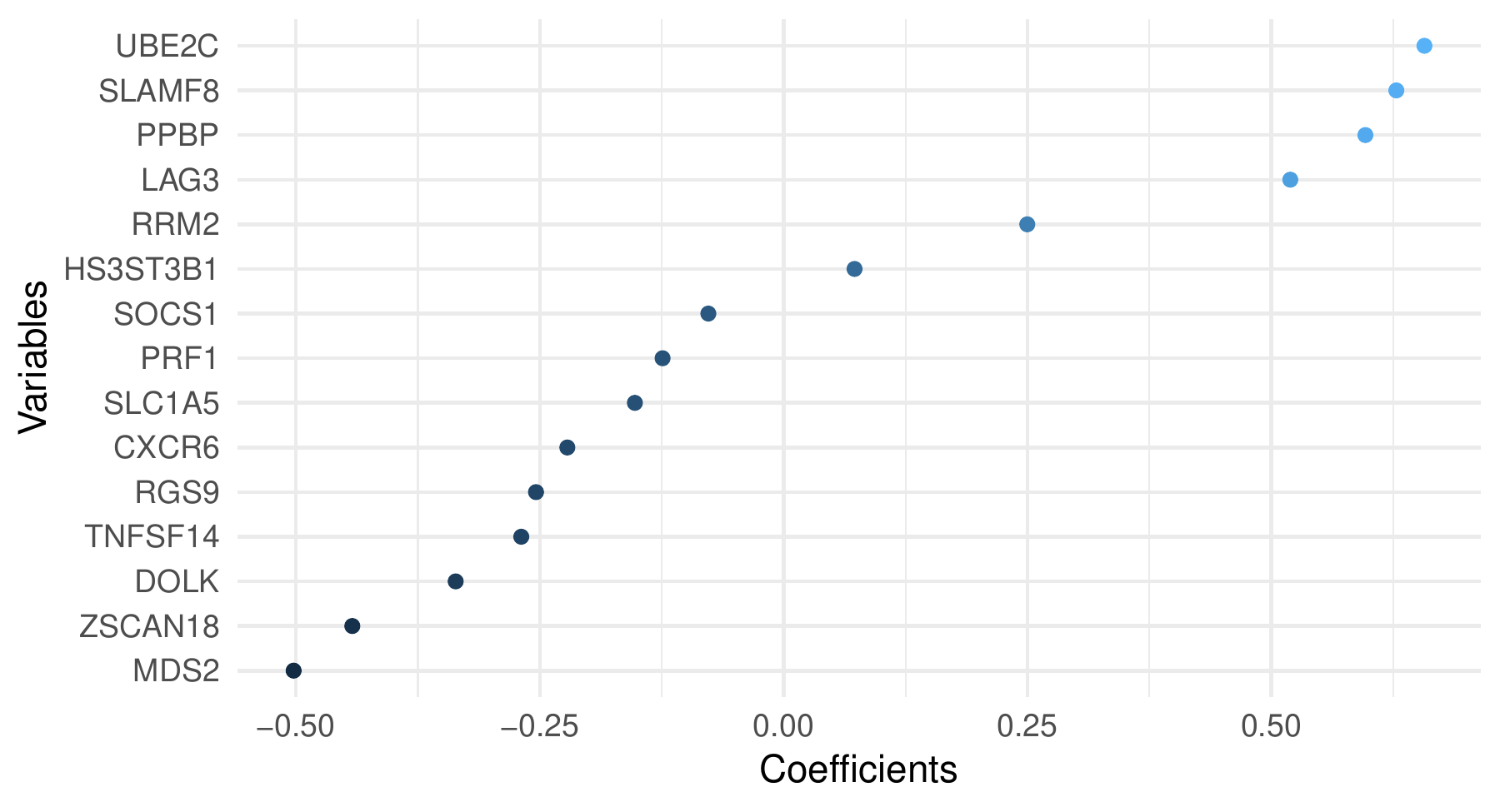
**Fig.** **24** Lasso ROC lambda min

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/Lasso-ROC-lambda-min.pdf)**



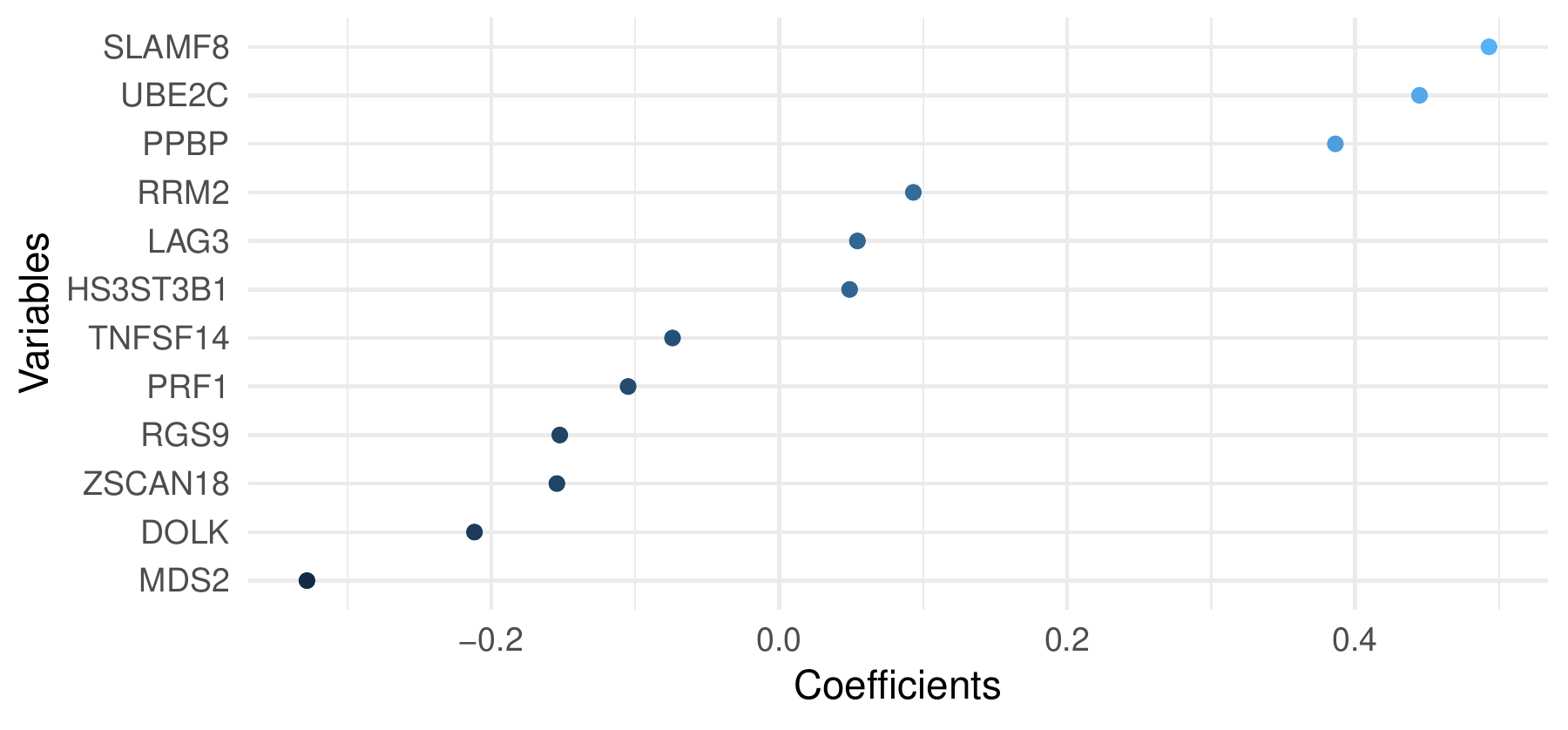
**Fig.** **25** Lasso ROC lambda 1se

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/Lasso-ROC-lambda-1se.pdf)**



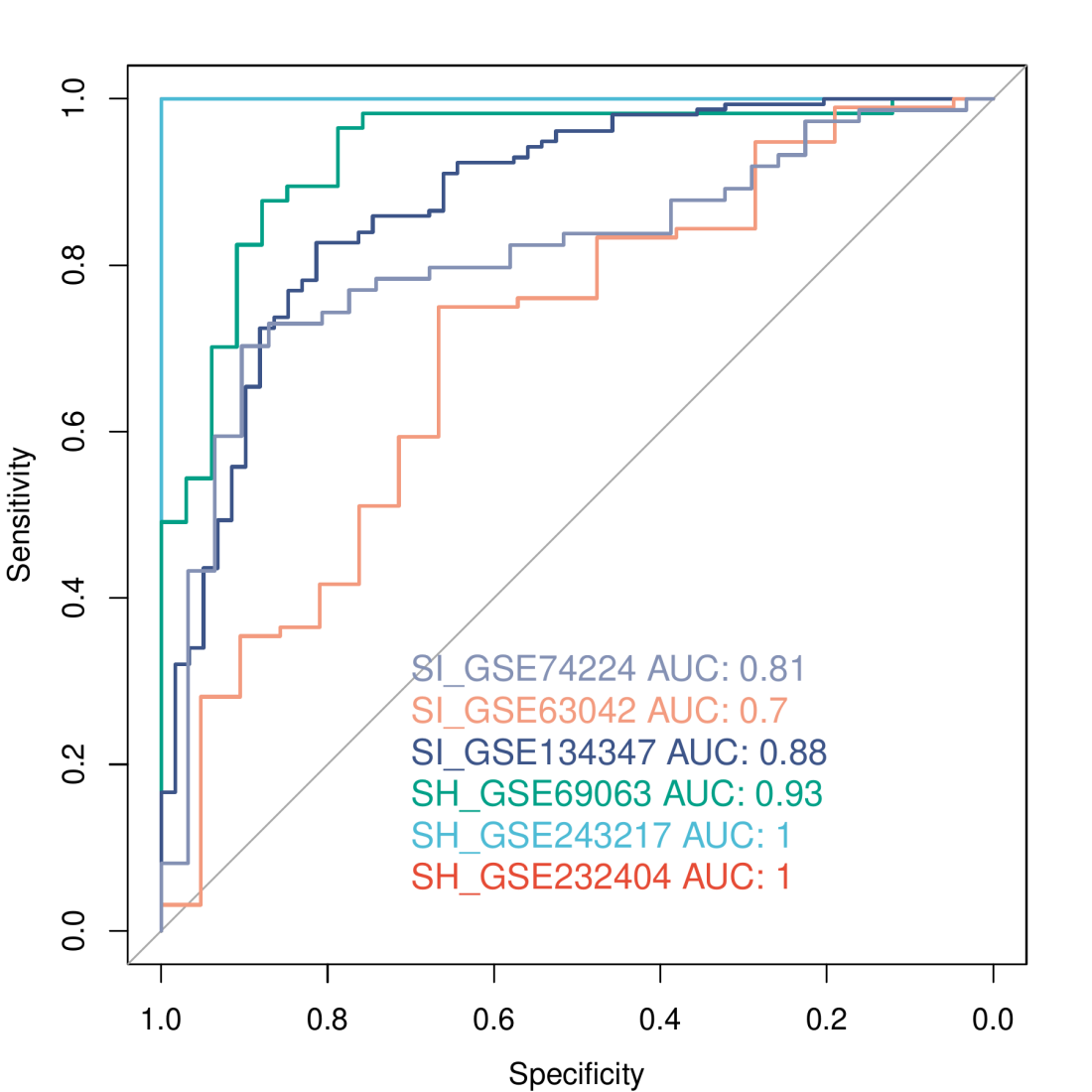
**Fig.** **26** Lasso coeffients lambda min

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/Lasso-coeffients-lambda-min.pdf)**



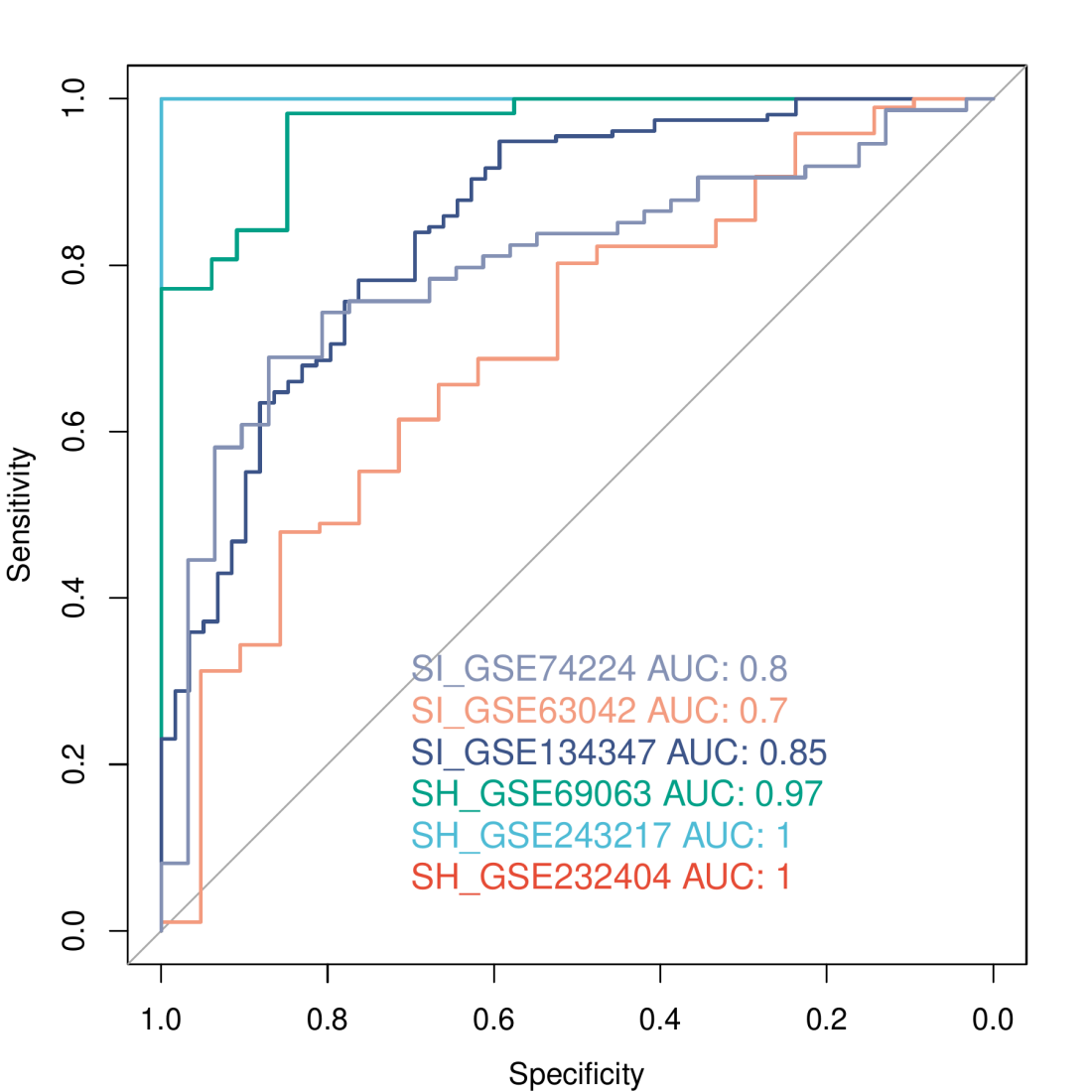
**Fig.** **27** Lasso coeffients lambda 1se

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/Lasso-coeffients-lambda-1se.pdf)**



**Fig.** **28** BULK lambda min ROC

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/BULK-lambda-min-ROC.pdf)**



**Fig.** **29** BULK lambda 1se ROC

**(File path: Figure+Table/3.7.0\_Lasso\_诊断模型建立\_(BULK)/BULK-lambda-1se-ROC.pdf)**

## 3.8 scFEA 单细胞数据的代谢通量预测 (SEPSIS)

将 Seurat (T 细胞) 以 scFEA 预测代谢通量。

Tab. 为细胞代谢通量矩阵 (各 M\_ 为代谢模块)。 Tab. 各代谢模块的注释。

**Tab.** **15** SEPSIS metabolic flux matrix

| V1 | M 1 | M 2 | M 3 | M 4 |
| --- | --- | --- | --- | --- |
| AAACCTGAGACGCTTT-1 | 9.235e-06 | 0.0397 | 0.01107 | 0.01938 |
| AAACCTGAGTTTAGGA-1 | 9.235e-06 | 0.02684 | 0.1845 | 0.01938 |
| AAACGGGAGACCTAGG-1 | 8.35e-06 | 0.08267 | 0.01107 | 0.09394 |
| AAAGATGTCTGCTGTC-1 | 8.35e-06 | 0.02641 | 0.01947 | 0.01938 |
| AAAGCAAAGGAGCGTT-1 | 9.235e-06 | 0.02645 | 0.01947 | 0.01938 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.8.0\_scFEA\_单细胞数据的代谢通量预测\_(SEPSIS)/SEPSIS-metabolic-flux-matrix.csv)**

**Tab.** **16** SEPSIS annotation of metabolic flux

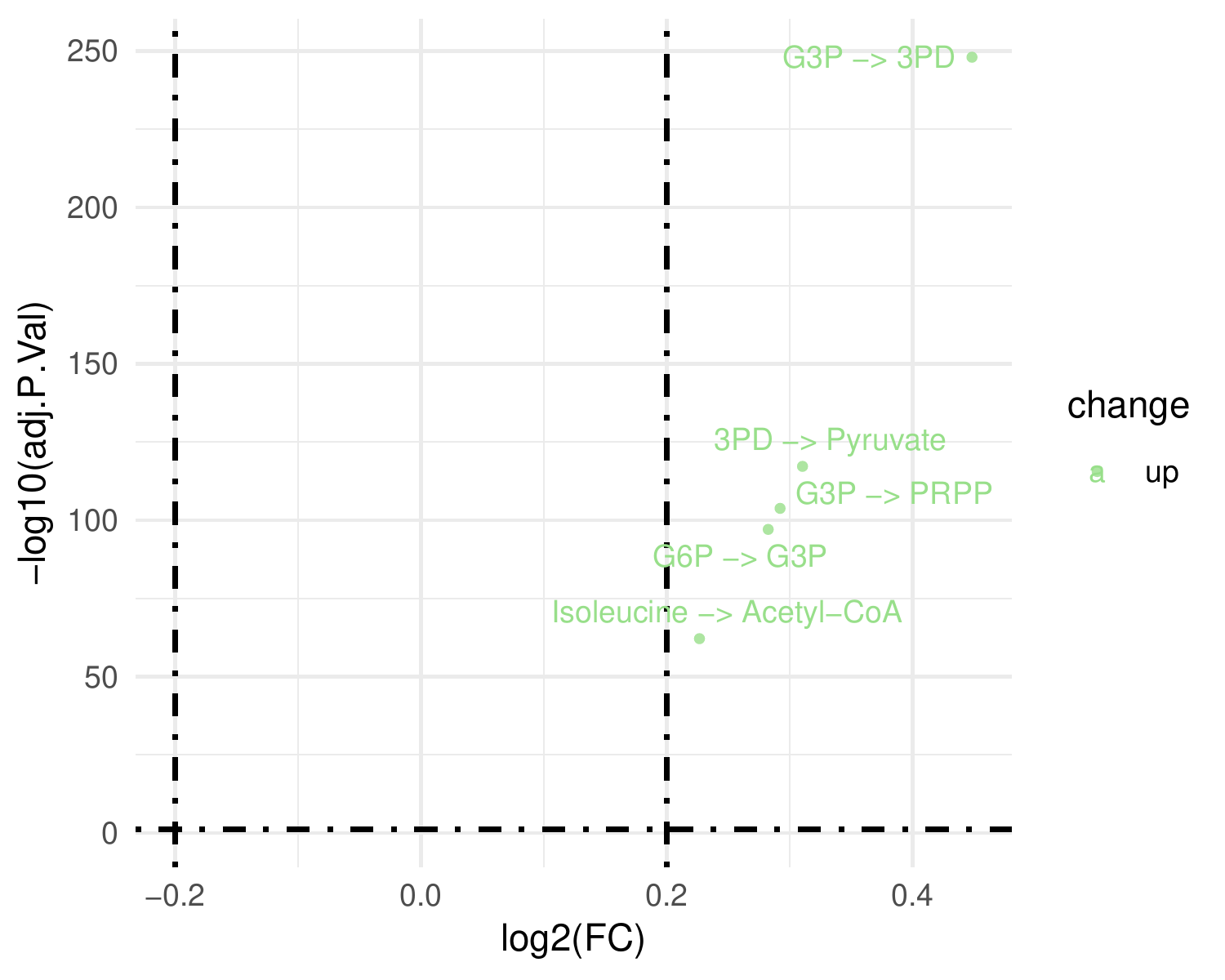
| V1 | Module id | Compound IN name | Compound IN ID | Compound OUT name |
| --- | --- | --- | --- | --- |
| M 1 | 1 | Glucose | C00267 | G6P |
| M 2 | 2 | G6P | C00668 | G3P |
| M 3 | 3 | G3P | C00118 | 3PD |
| M 4 | 4 | 3PD | C00197 | Pyruvate |
| M 5 | 5 | Pyruvate | C00022 | Acetyl-Coa |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.8.0\_scFEA\_单细胞数据的代谢通量预测\_(SEPSIS)/SEPSIS-annotation-of-metabolic-flux.xlsx)**

## 3.9 Limma 代谢通量差异分析 (SEPSIS\_FLUX)

以 limma 的线形分析策略，对细胞的代谢通量差异分析。样本分组：Control (n=14946) , Sepsis (n=8864) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Sepsis vs Control。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DMFs 统计：up (n=5) , down (n=0)

Fig. 为 Sepsis - Control 差异分析火山图。 Tab. 为 Sepsis - Control 差异分析统计表格。 Fig. 为细胞的代谢通量山脊图。



**Fig.** **30** SEPSIS FLUX Sepsis vs Control

**(File path: Figure+Table/3.9.0\_Limma\_代谢通量差异分析\_(SEPSIS\_FLUX)/SEPSIS-FLUX-Sepsis-vs-Control.pdf)**

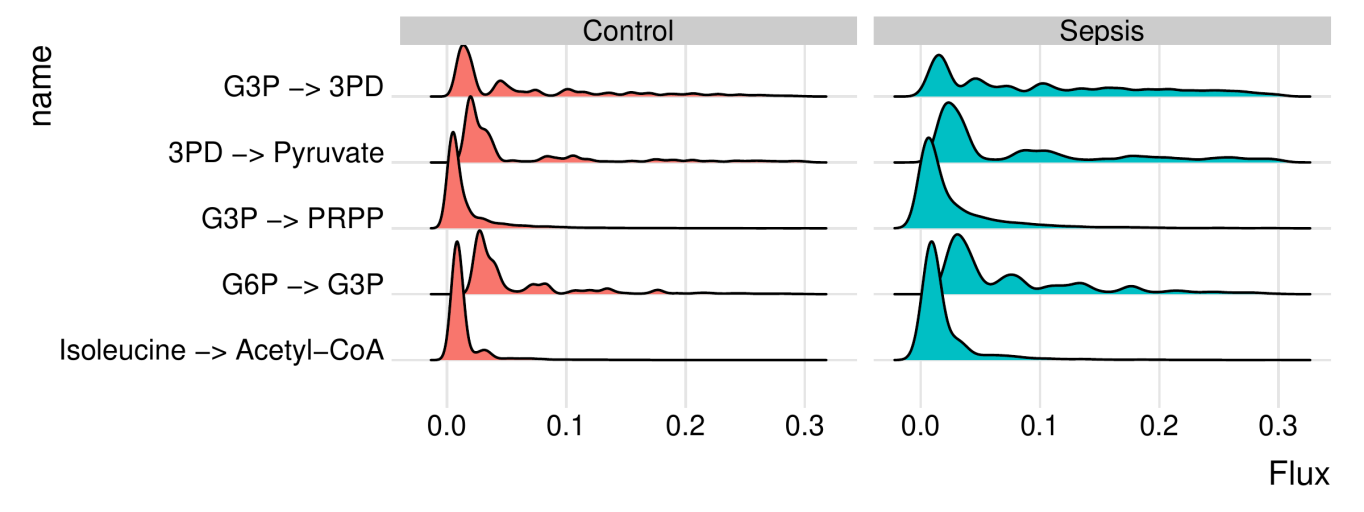
* adj.P.Val cut-off: 0.05
* Log2(FC) cut-off: 0.2

**(See: Figure+Table/3.9.0\_Limma\_代谢通量差异分析\_(SEPSIS\_FLUX)/SEPSIS-FLUX-Sepsis-vs-Control-content)**

**Tab.** **17** SEPSIS FLUX data Sepsis vs Control

| Name | LogFC | Adj.P.Val | Rownames | Module id |
| --- | --- | --- | --- | --- |
| G3P -> 3PD | 0.4482 | 1.049e-248 | M 3 | 3 |
| 3PD -> Pyruvate | 0.3103 | 6.047e-118 | M 4 | 4 |
| G3P -> PRPP | 0.2921 | 1.614e-104 | M 33 | 33 |
| G6P -> G3P | 0.2824 | 8.43e-98 | M 2 | 2 |
| Isoleucine -> Ace... | 0.2265 | 7.099e-63 | M 56 | 56 |

**(File path: Figure+Table/3.9.0\_Limma\_代谢通量差异分析\_(SEPSIS\_FLUX)/SEPSIS-FLUX-data-Sepsis-vs-Control.csv)**



**Fig.** **31** Cell flux ridge plot

**(File path: Figure+Table/3.9.0\_Limma\_代谢通量差异分析\_(SEPSIS\_FLUX)/Cell-flux-ridge-plot.pdf)**

## 3.10 关联分析

将 (基因集 (MDS2, SLAMF8, RRM2, …[n = 15], 来自于Lasso 诊断模型建立[Section: BULK]) ) 与 (代谢通量 (Sepsis - Control, 来自于Limma 代谢通量差异分析[Section: SEPSIS\_FLUX]) ) 关联分析。共得到 73 个显著的关联对 (P < 0.05)。

Fig. 为关联分析 (Diagnostic markers, Metabolic Flux) 热图。 Tab. 为关联分析统计附表 (P-value cutoff: 0.05)。



**Fig.** **32** Correlation heatmap

**(File path: Figure+Table/3.10.0\_关联分析/Correlation-heatmap.pdf)**

**Tab.** **18** Significant correlation

| Diagnostic.markers | Metabolic.Flux | Cor | Pvalue | -log2(P.value) |
| --- | --- | --- | --- | --- |
| RRM2 | G6P -> G3P | 0.29 | 0 | 16.61 |
| RRM2 | G3P -> PRPP | 0.29 | 0 | 16.61 |
| UBE2C | G6P -> G3P | 0.24 | 0 | 16.61 |
| UBE2C | G3P -> PRPP | 0.24 | 0 | 16.61 |
| RRM2 | Isoleucine -> Ace... | 0.24 | 0 | 16.61 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.10.0\_关联分析/Significant-correlation.csv)**

# 4 总结

(SIRS, systemic inflammatory response syndrome，文中的 SIRS 指的是 Infection-Negative Systemic Inflammation) 本研究从 T 细胞出发，探究 Sepsis 和 Non-infectious (Sepsis-Negative) 诊断标志物。以单细胞数据筛选 T 细胞在脓毒症和正常细胞中的表达差异。富集分析表明，这些基因与 T 细胞在免疫期间的行为以及与病菌感染密切相关 (Fig. **[9](#BULK-KEGG-enrichment)** , Fig. **[10](#BULK-GO-enrichment)** )。 同时，文中通过比较 Sepsis vs Healthy 与 Sepsis vis SIRS 的富集分析，也佐证了这一点 (见 [0](#compare))。 随后，结合这些基因，以 Bulk RNA-seq 建立了诊断模型。该模型同时以脓毒症与 SIRS ，脓毒症与健康对照组 的 Bulk RNA-seq 进行了大量的验证 (Fig. **[29](#BULK-lambda-1se-ROC)** , Fig. **[28](#BULK-lambda-min-ROC)** )。 本研究还针对 T 细胞在 Sepsis 状态下，与正常相比，代谢通量的改变 (通过对单细胞数据集预测代谢通量(2021, **IF:6.2**, Q1, Genome research)4) 做了差异分析，发现了五种于 Sepsis 下的代谢状态升高 (Fig. **[30](#SEPSIS-FLUX-Sepsis-vs-Control)** ， Fig. **[31](#Cell-flux-ridge-plot)** )。 关联分析表明，T 细胞的诊断标志物与这些代谢通量改变总体成负相关性 (Fig. **[32](#Correlation-heatmap)** )

# Reference

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

2. Smyth, G. K. Limma: Linear models for microarray data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer-Verlag, 2005). doi:[10.1007/0-387-29362-0\_23](https://doi.org/10.1007/0-387-29362-0_23).

3. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data users guide. 119.

4. Alghamdi, N. *et al.* A graph neural network model to estimate cell-wise metabolic flux using single-cell rna-seq data. *Genome research* **31**, 1867–1884 (2021).

5. Reyes, M. *et al.* An immune-cell signature of bacterial sepsis. *Nature medicine* **26**, 333–340 (2020).