**生信分析报告**

**项目标题： 黑色素瘤 ;**

**单 号： BSHQ240813 ;**

**分析人员： 黄礼闯 ;**

**分析类型： 生信分析 ;**

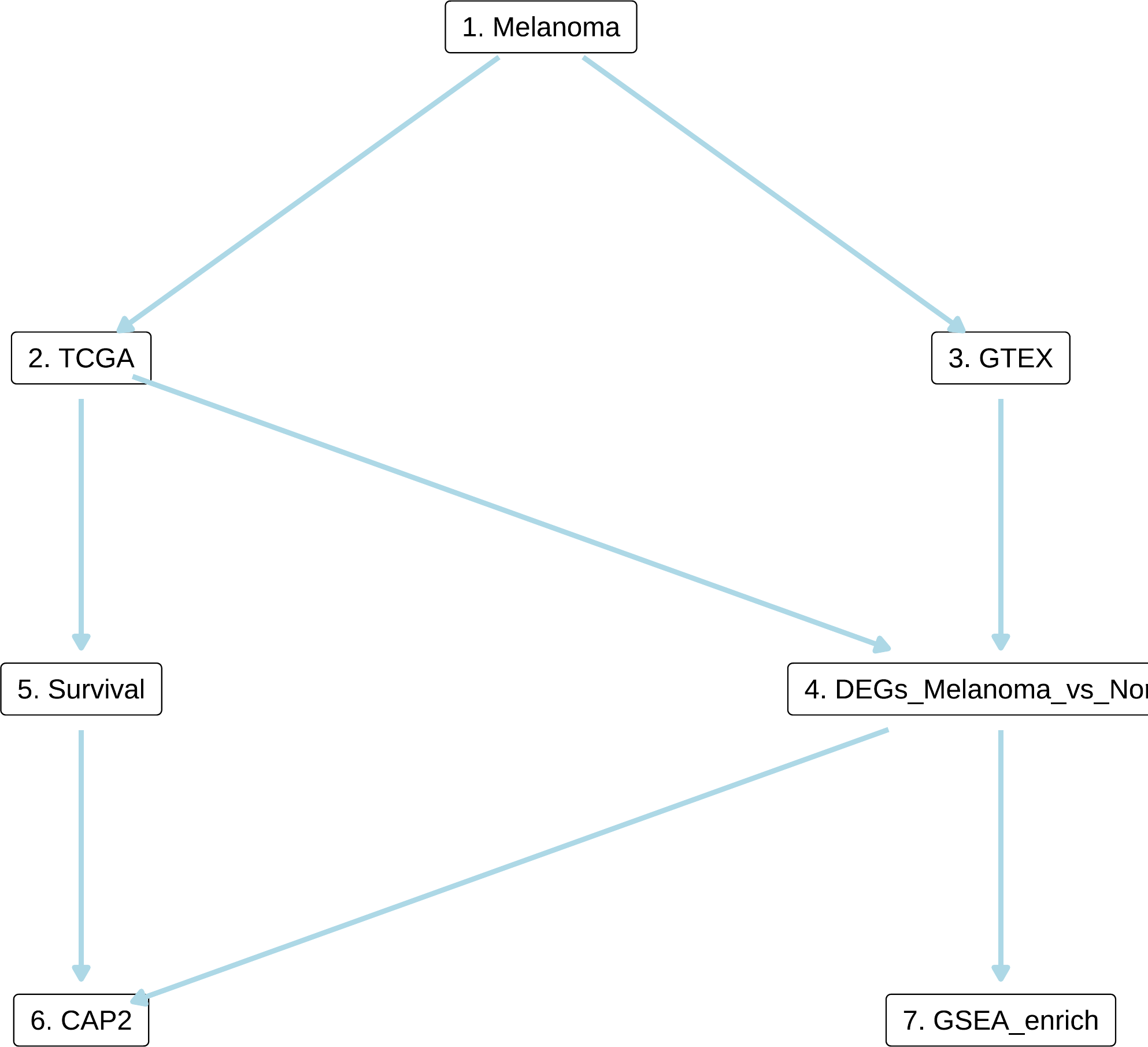
**委 托 人： 王凯 ;**

**受 托 人： 杭州铂赛生物科技有限公司 .**

# 1 分析流程

## 1.1 需求

1. 通过生物信息分析分析 CAP2 在黑色素瘤异常表达且与预后相关。HPA数据库分析。
2. 为了研究CAP2影响黑色素瘤进展的机制，进行GSEA等生物信息分析，寻找与CAP2R显著相关的下游信号通路



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

**(File path: Figure+Table/1.1\_需求/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 TCGA 数据获取 (Dataset: TCGA\_SKCM)

以 R 包 TCGAbiolinks (2.34.0) (2015, **IF:16.6**, Q1, Nucleic Acids Research)1 获取 TCGA-SKCM 数据集。

## 2.3 Survival 生存分析 (Dataset: TCGA\_SKCM)

以 R 包 edgeR (4.4.0) ()2 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)。使用标准化过的基因表达数据。 以 R 包 survival (3.7.0) 生存分析，以 R 包 survminer (0.5.0) 绘制生存曲线。以 R 包 timeROC (0.4) 绘制 1, 3, 5 年生存曲线。

## 2.4 UCSCXenaTools 癌症相关数据获取 (Dataset: SKCM)

以 R 包 UCSCXenaTools (1.6.0) (2019, Journal of Open Source Software)3 获取 TcgaTargetGtex 类型数据。

## 2.5 Limma 差异分析 (Dataset: XENA\_SKCM)

以 R 包 edgeR (4.4.0) ()2 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)4 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：SKCM vs Normal。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。

## 2.6 ClusterProfiler GSEA 富集分析 (Dataset: XENA\_SKCM)

以 ClusterProfiler R 包 (4.15.0.2) (2021, **IF:33.2**, Q1, The Innovation)5 按 GSVA 算法 (clusterProfiler::gseGO, ClusterProfiler::gseKEGG)，进行 KEGG 和 GO 富集分析 (P-value Cutoff = 0.05) 。 以 R 包 msigdbr (7.5.1) 获取 MSigDB 数据库基因集，用于 clusterProfiler GSEA 富集分析。

## 2.7 Pathview 通路可视化 (Dataset: XENA\_SKCM)

以 R 包 pathview (2013, Bioinformatics (Oxford, England))6 将 clusterProfiler 以 GSEA 算法富集的 KEGG 通路可视化。

# 3 分析结果

## 3.1 TCGA 数据获取 (TCGA\_SKCM)

获取 TCGA-SKCM 数据。

## 3.2 Survival 生存分析 (TCGA\_SKCM)

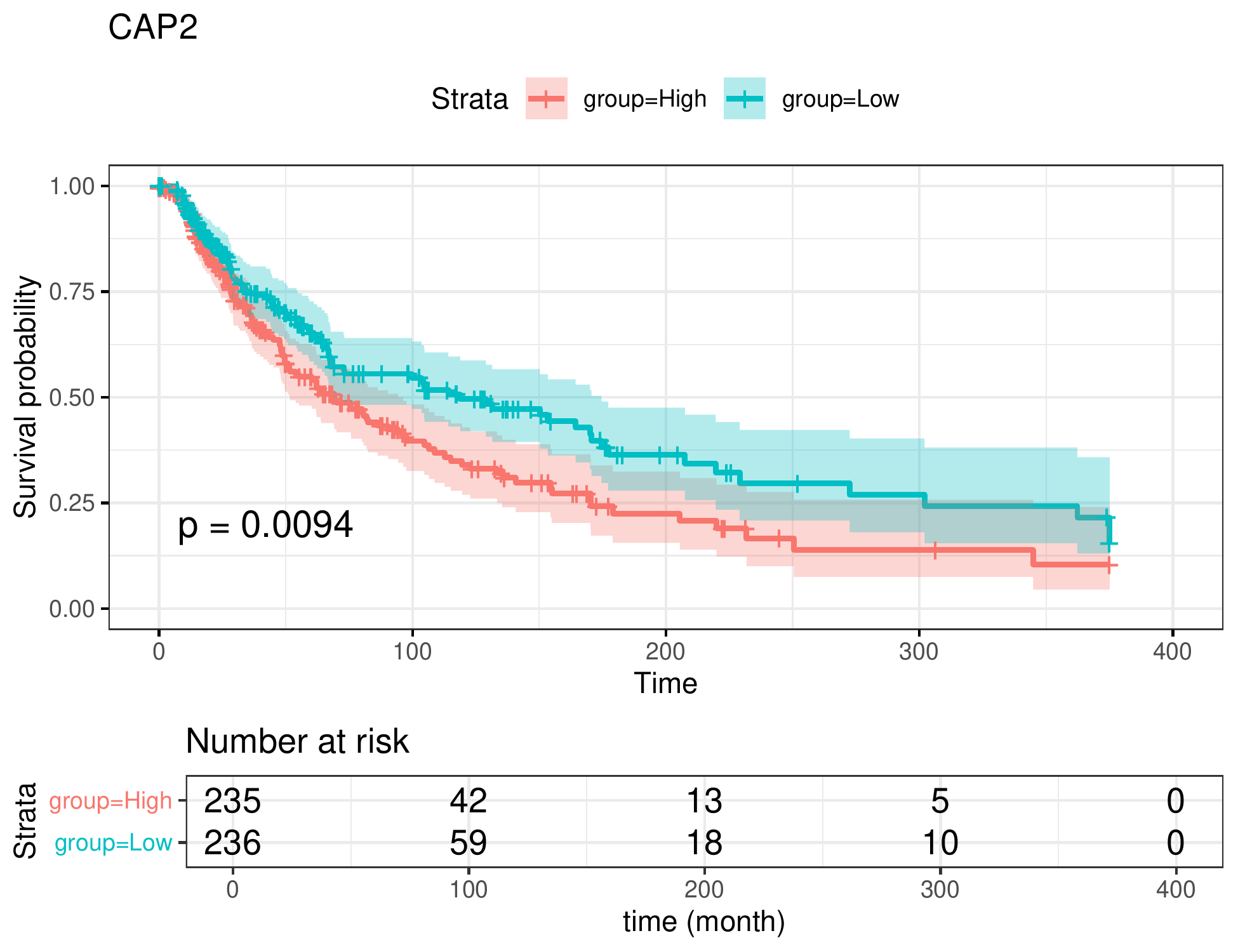
按中位风险评分，将样本分为 Low 和 High 风险组。生存数据为TCGA-SKCM，使用标准化过的基因表达数据。根据元数据信息 (即临床数据) ，去除了生存状态未知的样例。

Tab. **[1](#TCGA-SKCM-Significant-Survival-PValue)** 。 Fig. **[2](#TCGA-SKCM-survival-curve-of-CAP2)** 为 CAP2 生存曲线。

**Tab.** **1** TCGA SKCM Significant Survival PValue

| Name | Pvalue |
| --- | --- |
| CAP2 | 0.009385 |

**(File path: Figure+Table/3.2\_Survival\_生存分析\_(TCGA\_SKCM)/TCGA-SKCM-Significant-Survival-PValue.csv)**



**Fig.** **2** TCGA SKCM survival curve of CAP2

**(File path: Figure+Table/3.2\_Survival\_生存分析\_(TCGA\_SKCM)/TCGA-SKCM-survival-curve-of-CAP2.pdf)**

## 3.3 UCSCXenaTools 癌症相关数据获取 (SKCM)

获取 UCSC Xena 的 TcgaTargetGtex 数据。共 1023 个数据。样本组织为：Skin (n=1023) 。样本类型为：Metastatic (n=366) , Normal Tissue (n=555) , Primary Tumor (n=102) 。数据来源为：GTEX (n=555) , TCGA (n=468) 。

## 3.4 Limma 差异分析 (XENA\_SKCM)

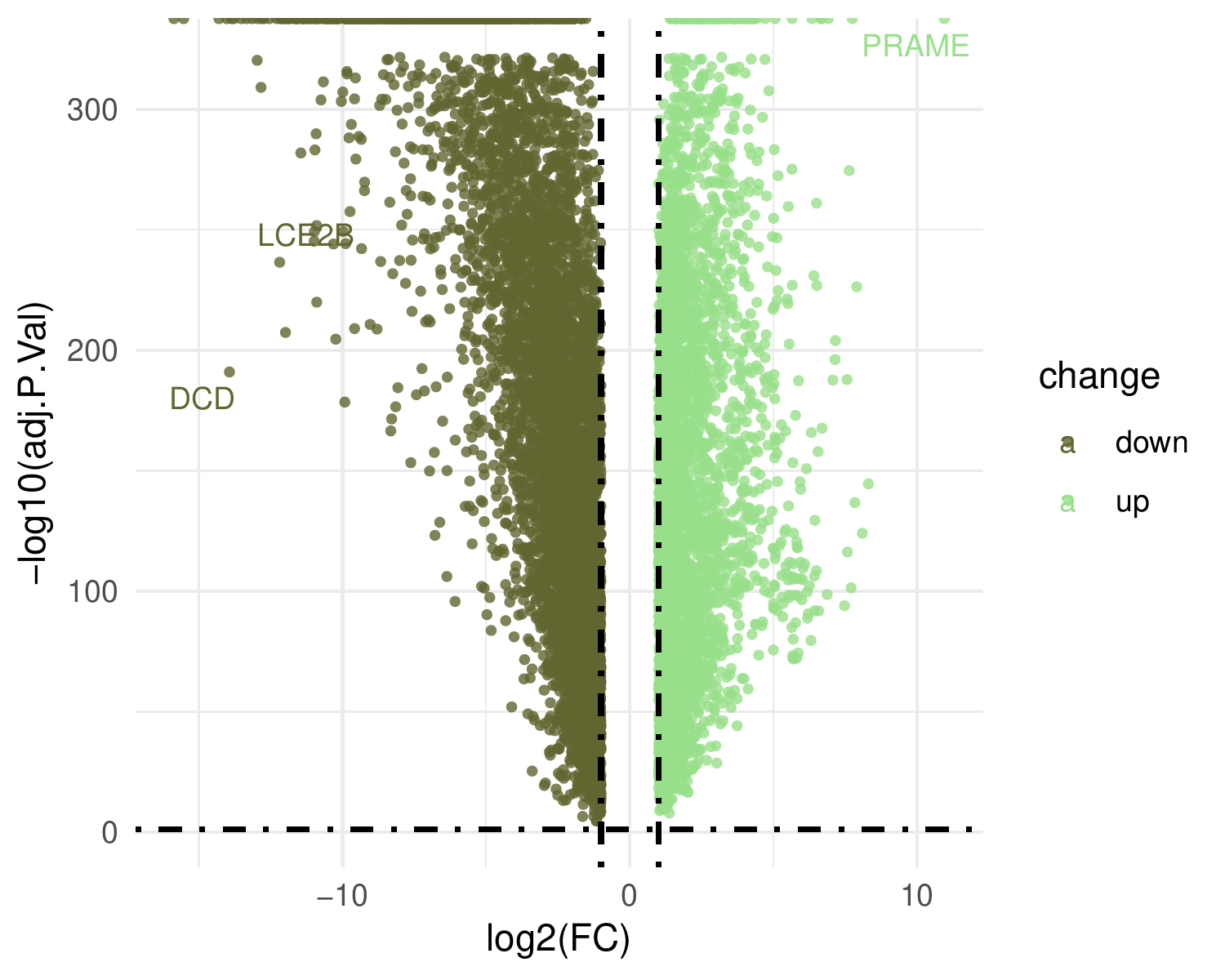
将 TcgaTargetGtex-SKCM 数据转化为 counts 型数据 (原数据为 log2(counts + 1)) 。样本分组：Normal (n=555) , SKCM (n=468) 。以 公式 ~ 0 + group + batch 创建设计矩阵 (design matrix) (Batch: GTEX (n=555) , TCGA (n=468) )。差异分析：SKCM vs Normal。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=3819) , down (n=5963) 对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。将基因 (CAP2 -> CDKN2A, MCM2, ORC1, …(n = 97)) 关联分析。将基因 (CAP2 -> CDKN2A, MCM2, ORC1, …(n = 97)) 关联分析。

Tab. **[2](#XENA-SKCM-data-SKCM-vs-Normal)** 为 SKCM - Normal 差异分析统计表格。 Fig. **[3](#XENA-SKCM-SKCM-vs-Normal)** 为 SKCM - Normal 差异分析火山图。 Fig. **[4](#XENA-SKCM-Box-Plot-Of-DEGs)** 基因 CAP2 表达水平，以及对应的 limma 差异分析显著水平。

**Tab.** **2** XENA SKCM data SKCM vs Normal

| Hgnc s... | LogFC | Adj.P.Val | Rownames | Id | Chrom | ChromS... | ChromEnd | Strand | AveExpr |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AC0090... | 6.644 | 0 | ENSG00... | ENSG00... | Chr16 | 2254000 | 2262000 | + | -2.928 |
| RP11-4... | 7.747 | 0 | ENSG00... | ENSG00... | Chr9 | 87530000 | 87530000 | + | -1.144 |
| VIPR1-AS1 | -7.064 | 0 | ENSG00... | ENSG00... | Chr3 | 42510000 | 42530000 | - | -1.973 |
| HNRNPCP2 | 3.7 | 0 | ENSG00... | ENSG00... | Chr2 | 189900000 | 189900000 | + | 1.128 |
| IL20RB... | -6.995 | 0 | ENSG00... | ENSG00... | Chr3 | 1.37e+08 | 1.37e+08 | - | -2.823 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.4\_Limma\_差异分析\_(XENA\_SKCM)/XENA-SKCM-data-SKCM-vs-Normal.csv)**

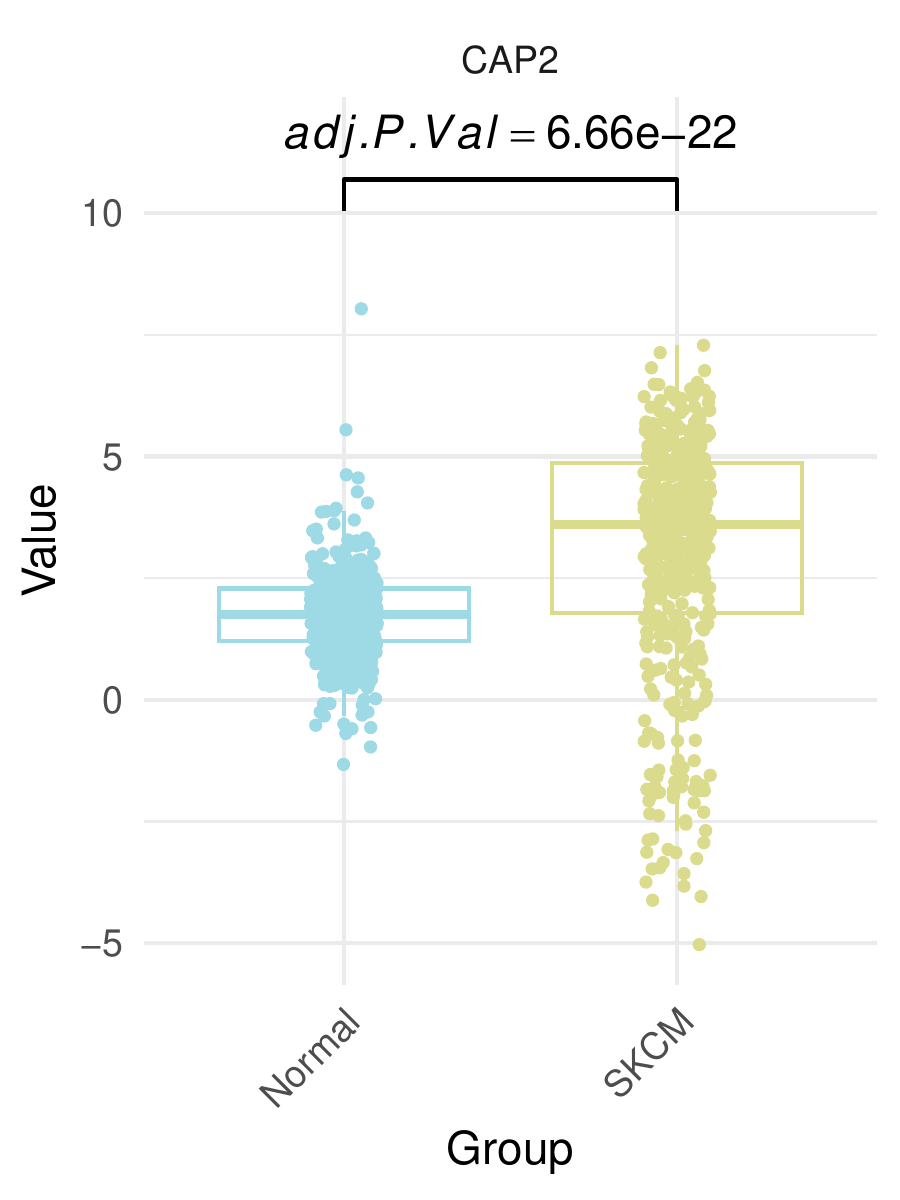


**Fig.** **3** XENA SKCM SKCM vs Normal

**(File path: Figure+Table/3.4\_Limma\_差异分析\_(XENA\_SKCM)/XENA-SKCM-SKCM-vs-Normal.pdf)**

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

**(See: Figure+Table/3.4\_Limma\_差异分析\_(XENA\_SKCM)/XENA-SKCM-SKCM-vs-Normal-content)**



**Fig.** **4** XENA SKCM Box Plot Of DEGs

**(File path: Figure+Table/3.4\_Limma\_差异分析\_(XENA\_SKCM)/XENA-SKCM-Box-Plot-Of-DEGs.pdf)**

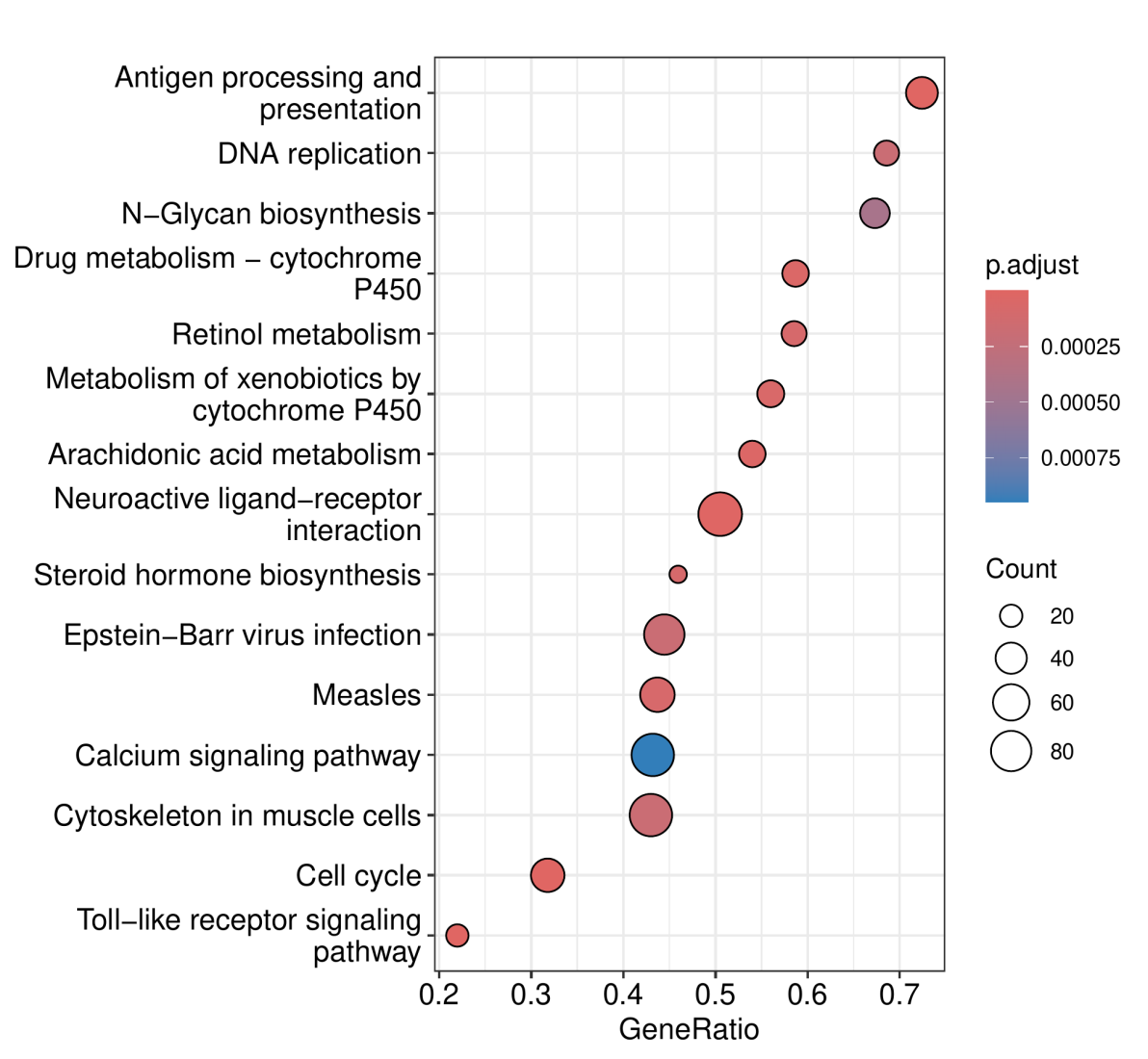
## 3.5 ClusterProfiler GSEA 富集分析 (XENA\_SKCM)

对基因集 (SKCM - Normal, 来自于Limma 差异分析(Section: XENA\_SKCM)) 进行ClusterProfiler GSEA 富集分析。按矫正 P 值排序 (BH, p.adjust) ，最显著的三条通路为: Cell cycle, Antigen processing and presentation, Toll-like receptor signaling pathway。均显著上调。 Fig. **[6](#XENA-SKCM-GSEA-plot-of-the-pathways)** 为 GSEA KEGG 富集条码图 (以 clusterProfiler 仿 GSEA 软件绘图)。

所有富集的通路都不包含 CAP2 基因 (该基因不被 KEGG、GO 等数据库记录到通路中，研究较少)。

Fig. **[5](#XENA-SKCM-GSEA-KEGG-enrichment)** 为 GSEA KEGG 富集分析气泡图。 Tab. **[3](#XENA-SKCM-GSEA-KEGG-enrichment-data)** 为 GSEA KEGG 富集分析统计附表。

从富集结果中筛选包含CAP2的通路。筛选到“Genes down-regulated in MCF-7 cells (breast cancer) by selective estrogen receptor modulators (SERM) 4-hydroxytamoxifen, raloxifene, or ICI 182780 but not by estradiol [PubChem=44959;5035;3478439;5757].”, “Genes down-regulated by estradiol [PubChem=5757] and down-regulated by ESRRA [GeneID=2101] in MCF-7 cells (breast cancer).”, “Genes down-regulated in lobular carcinoma vs normal lobular breast cells.”, …(n = 4)。 Tab. **[4](#XENA-SKCM-filter-by-match-genes-g)** 以CAP2筛选到的通路。 为 C2 基因集富集结果表格 (包含 CAP2 的通路) (这些似乎与 CAP2 在黑色素瘤中无关)



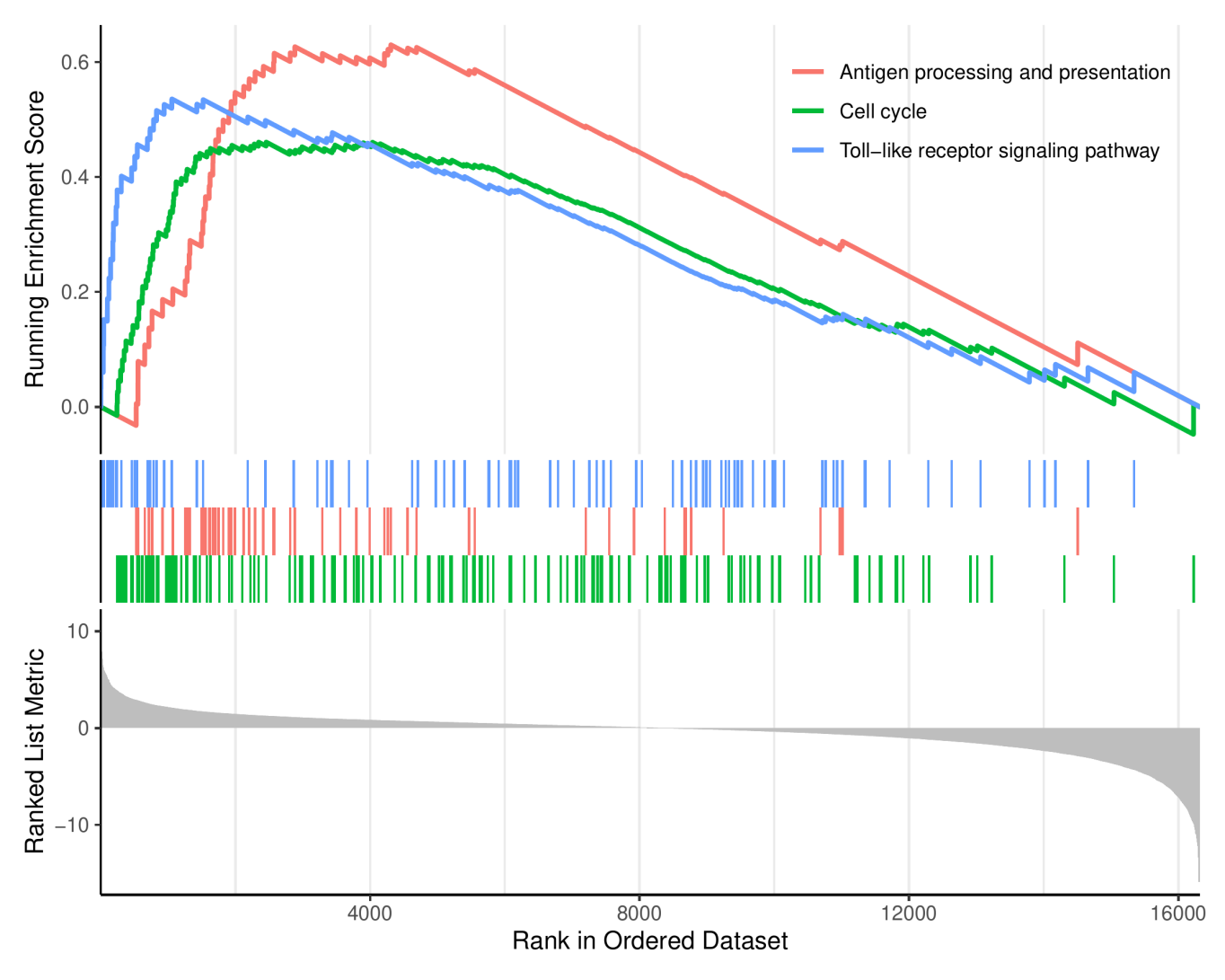
**Fig.** **5** XENA SKCM GSEA KEGG enrichment

**(File path: Figure+Table/3.5\_ClusterProfiler\_GSEA\_富集分析\_(XENA\_SKCM)/XENA-SKCM-GSEA-KEGG-enrichment.pdf)**

**Tab.** **3** XENA SKCM GSEA KEGG enrichment data

| ID | Descri... | SetSize | Enrich... | NES | Pvalue | P.adjust | Qvalue | Rank | Leadin... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hsa04110 | Cell c... | 151 | 0.4606 | 2.205 | 2.009e-09 | 6.591e-07 | 4.315e-07 | 2457 | Tags=3... |
| Hsa04612 | Antige... | 58 | 0.6299 | 2.563 | 5.678e-09 | 9.312e-07 | 6.097e-07 | 4307 | Tags=7... |
| Hsa04620 | Toll-l... | 91 | 0.5357 | 2.348 | 9.066e-09 | 9.912e-07 | 6.489e-07 | 1057 | Tags=2... |
| Hsa04080 | Neuroa... | 196 | -0.5477 | -1.691 | 1.054e-07 | 8.642e-06 | 5.658e-06 | 3314 | Tags=5... |
| Hsa00590 | Arachi... | 50 | -0.7093 | -1.989 | 4.937e-07 | 3.239e-05 | 2.12e-05 | 2602 | Tags=5... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.5\_ClusterProfiler\_GSEA\_富集分析\_(XENA\_SKCM)/XENA-SKCM-GSEA-KEGG-enrichment-data.xlsx)**



**Fig.** **6** XENA SKCM GSEA plot of the pathways

**(File path: Figure+Table/3.5\_ClusterProfiler\_GSEA\_富集分析\_(XENA\_SKCM)/XENA-SKCM-GSEA-plot-of-the-pathways.pdf)**

**Tab.** **4** XENA SKCM filter by match genes g

| ID | Descri... | SetSize | Enrich... | NES | Pvalue | P.adjust | Qvalue | Rank | Leadin... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M11250 | Genes ... | 48 | 0.597 | 2.246 | 1.5e-06 | 3.154e-05 | 2.371e-05 | 3614 | Tags=6... |
| M19002 | Genes ... | 40 | 0.5963 | 2.165 | 4.86e-05 | 0.0006841 | 0.0005143 | 3013 | Tags=5... |
| M13547 | Genes ... | 72 | 0.4115 | 1.729 | 0.0006924 | 0.006467 | 0.004862 | 3187 | Tags=4... |
| M1472 | Genes ... | 56 | 0.4276 | 1.657 | 0.002809 | 0.01923 | 0.01446 | 2377 | Tags=3... |

**(File path: Figure+Table/3.5\_ClusterProfiler\_GSEA\_富集分析\_(XENA\_SKCM)/XENA-SKCM-filter-by-match-genes-g.xlsx)**

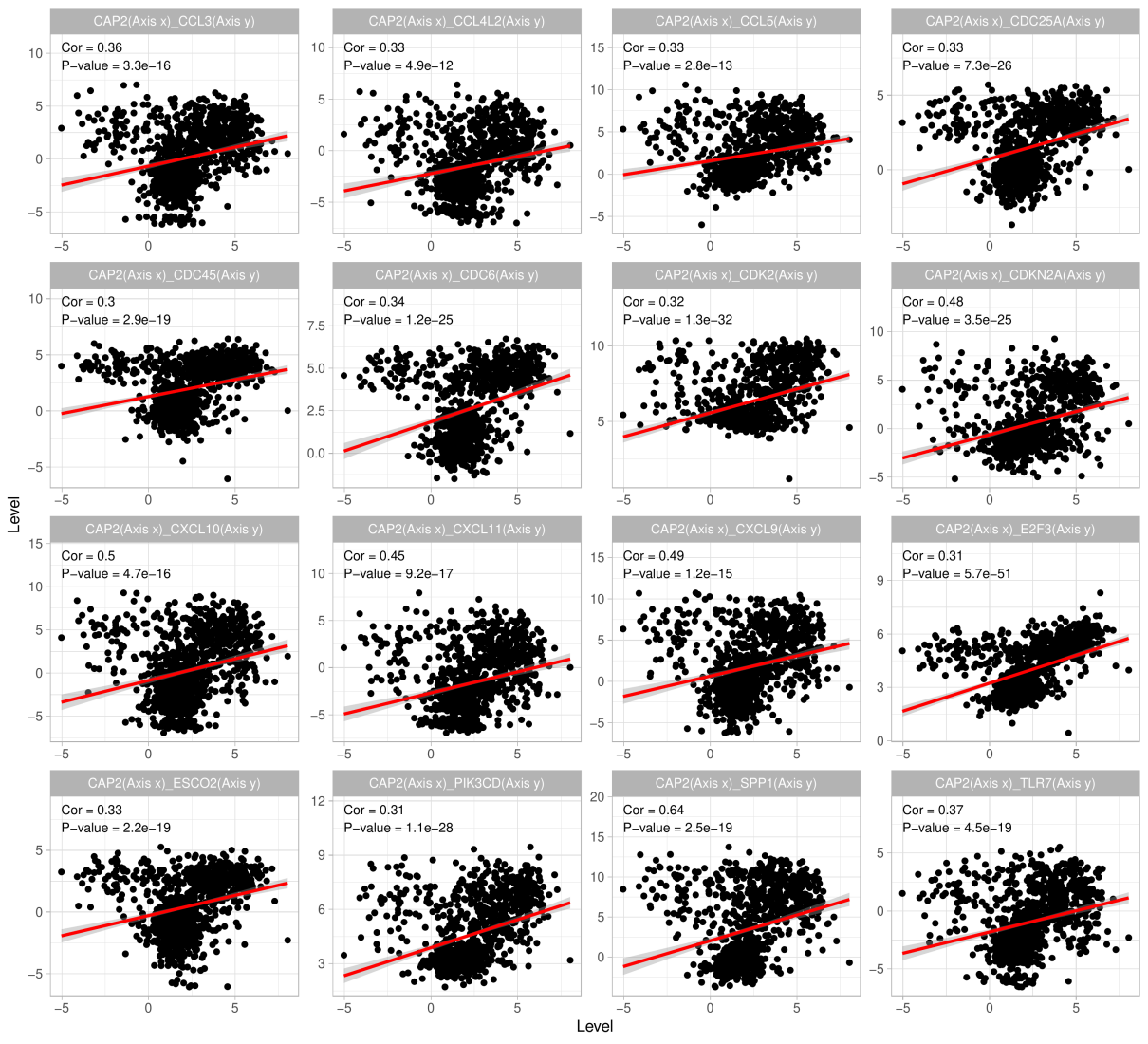
## 3.6 关联分析 (XENA\_SKCM)

将基因 (CAP2 -> 基因集 (来自于ClusterProfiler GSEA 富集分析(Section: XENA\_SKCM)) ) 关联分析。共得到 16 个显著的基因对 (P < 0.05, |Cor| > 0.3)。 Fig. **[7](#XENA-SKCM-significant-correlation-plots)** 为显著关联的基因的线型回归图。 Tab. **[5](#XENA-SKCM-significant-correlation-analysis-data)** 为关联分析统计附表 (P-value cutoff: 0.05, Cor (关联系数) cutoff: 0.3)。 从富集结果中筛选包含基因集 (CXCL9, CDC6, …(n = 14)., 来自于关联分析(Section: XENA\_SKCM)) 的通路。筛选到“Cell cycle”, “Toll-like receptor signaling pathway”, “Measles”, …(n = 17)。

**Tab.** **5** XENA SKCM significant correlation analysis data

| From | To | Cor | Pvalue | Model | -log2(... | Signif... | Sign |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CAP2 | CXCL9 | 0.4906 | 1.202e-15 | C(TCGA... | 49.56 | < 0.001 | \*\* |
| CAP2 | CDC6 | 0.3415 | 1.16e-25 | C(TCGA... | 82.83 | < 0.001 | \*\* |
| CAP2 | CCL5 | 0.3275 | 2.844e-13 | C(TCGA... | 41.68 | < 0.001 | \*\* |
| CAP2 | SPP1 | 0.6418 | 2.517e-19 | C(TCGA... | 61.79 | < 0.001 | \*\* |
| CAP2 | CDK2 | 0.3156 | 1.269e-32 | C(TCGA... | 106 | < 0.001 | \*\* |
| ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6\_关联分析\_(XENA\_SKCM)/XENA-SKCM-significant-correlation-analysis-data.xlsx)**



**Fig.** **7** XENA SKCM significant correlation plots

**(File path: Figure+Table/3.6\_关联分析\_(XENA\_SKCM)/XENA-SKCM-significant-correlation-plots.pdf)**

**Tab.** **6** XENA SKCM filter by match genes f

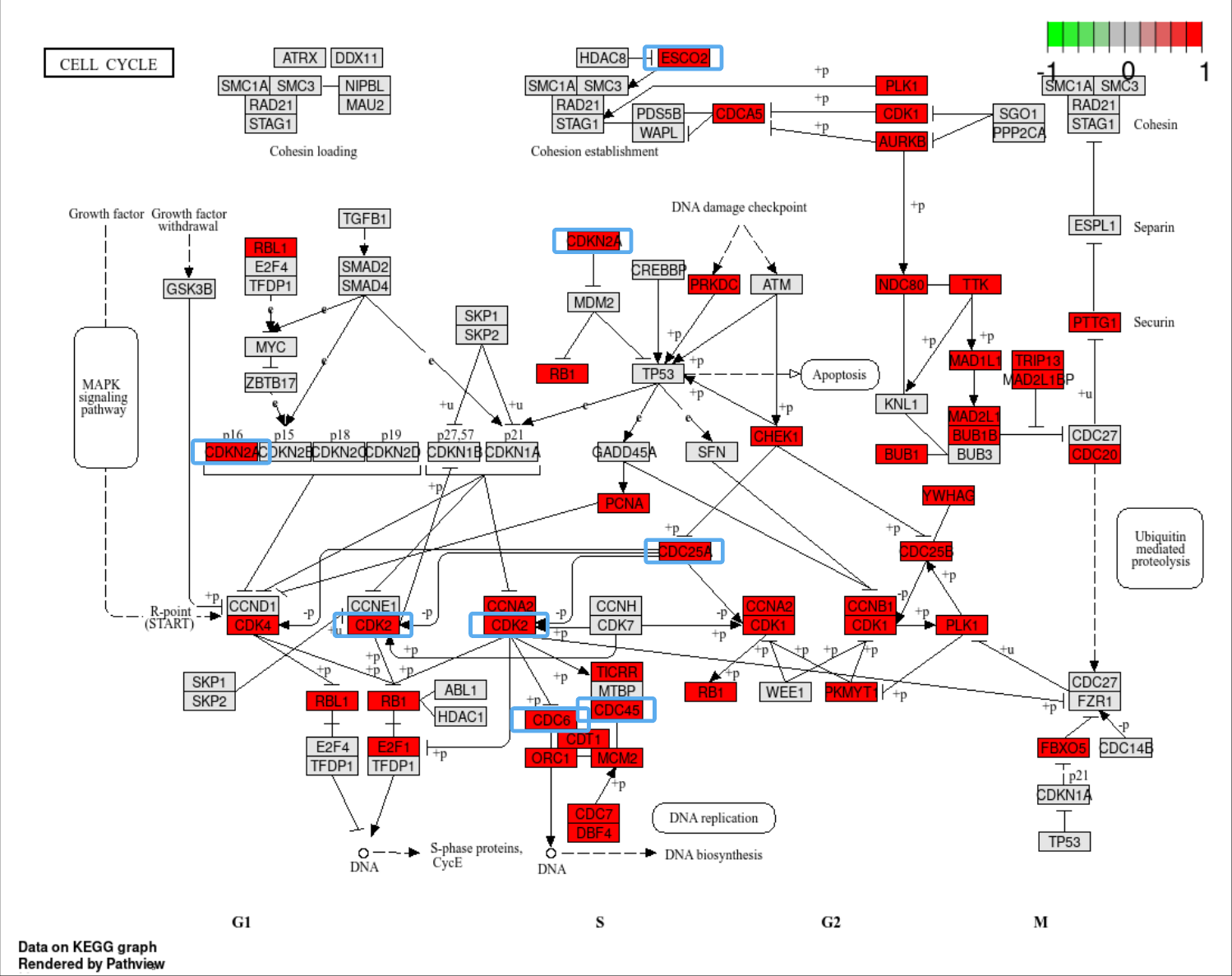
| ID | Descri... | SetSize | Enrich... | NES | Pvalue | P.adjust | Qvalue | Rank | Leadin... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hsa04110 | Cell c... | 151 | 0.4606 | 2.205 | 2.009e-09 | 6.591e-07 | 4.315e-07 | 2457 | Tags=3... |
| Hsa04620 | Toll-l... | 91 | 0.5357 | 2.348 | 9.066e-09 | 9.912e-07 | 6.489e-07 | 1057 | Tags=2... |
| Hsa05162 | Measles | 119 | 0.4323 | 1.98 | 1.999e-06 | 8.195e-05 | 5.365e-05 | 4156 | Tags=4... |
| Hsa05169 | Epstei... | 180 | 0.3578 | 1.698 | 6.742e-06 | 0.0001846 | 0.0001209 | 4881 | Tags=4... |
| Hsa04380 | Osteoc... | 128 | 0.3803 | 1.757 | 5.995e-05 | 0.001093 | 0.0007152 | 1242 | Tags=2... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6\_关联分析\_(XENA\_SKCM)/XENA-SKCM-filter-by-match-genes-f.xlsx)**

## 3.7 Pathview 通路可视化 (XENA\_SKCM)

以 pathview 探究基因集在通路 hsa04110, hsa04620 中的上下游关系。从KEGG hsa04110, hsa04620中，查找包含基因集 (CXCL9, CDC6, CCL5, …[n = 16], 来自于关联分析[Section: XENA\_SKCM]) 的通路。

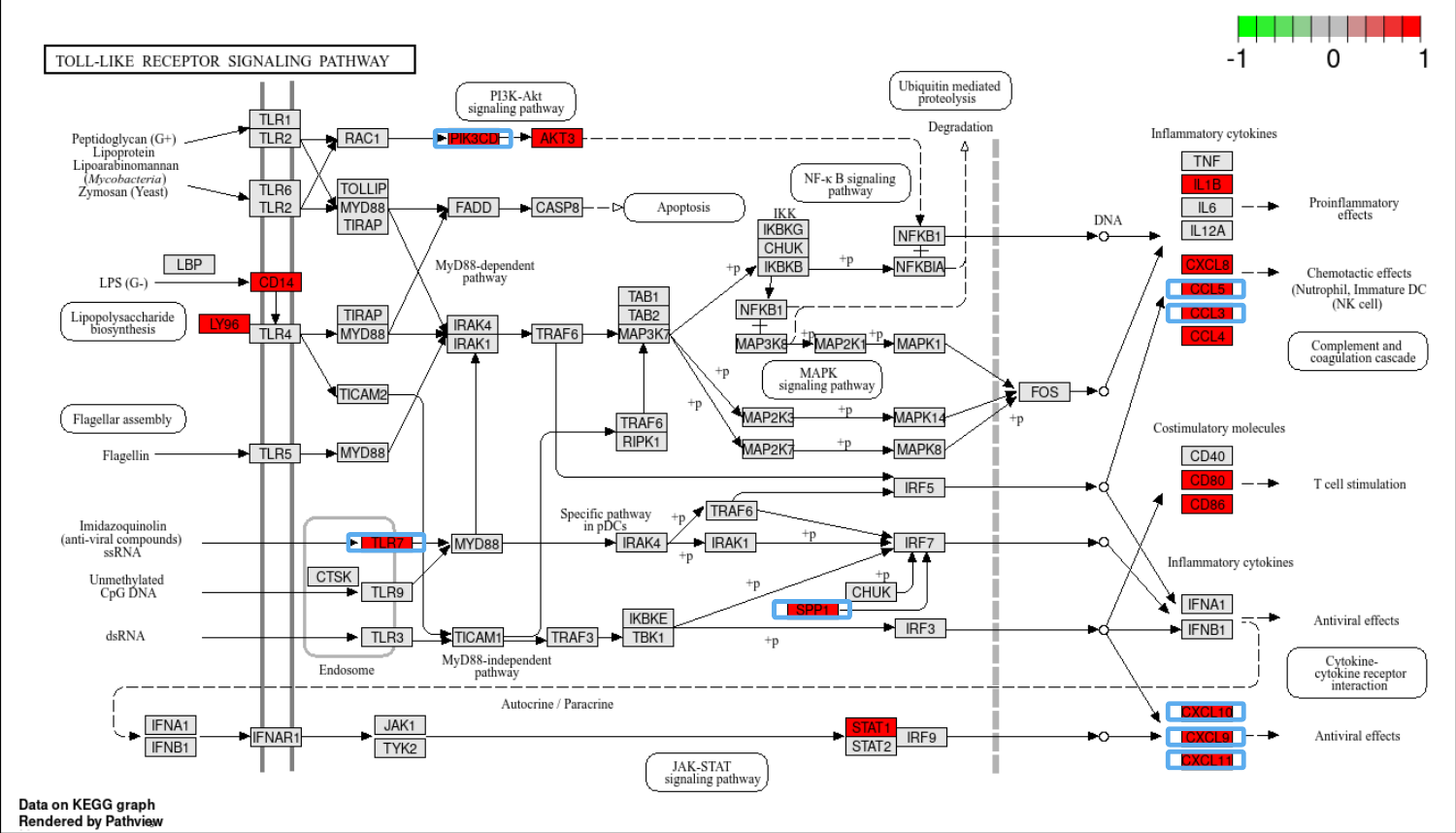
Fig. **[8](#XENA-SKCM-marked-pathway-hsa04110)** 为KEGG hsa04110通路图，包含基因上调或下调信息，并标记了基因集 (CXCL9, CDC6, CCL5, …[n = 16], 来自于关联分析[Section: XENA\_SKCM]) 。 Fig. **[9](#XENA-SKCM-marked-pathway-hsa04620)** 为KEGG hsa04620通路图，包含基因上调或下调信息，并标记了基因集 (CXCL9, CDC6, CCL5, …[n = 16], 来自于关联分析[Section: XENA\_SKCM]) 。



**Fig.** **8** XENA SKCM marked pathway hsa04110

**(File path: Figure+Table/3.7\_Pathview\_通路可视化\_(XENA\_SKCM)/XENA-SKCM-marked-pathway-hsa04110.pdf)**

* Interactive figure: <https://www.genome.jp/pathway/hsa04110>
* Enriched genes: CDKN2A, MCM2, ORC1, CDK1, BUB1B, PCNA, PLK1, DBF4, MAD2L1, BUB1, YWHAG, CHEK1, PRKDC, PKMYT1, PTTG1, RB1, CDC7, CDC20, CDC25B, CDC45, CDC6, CDC25A, CCNB1, CCNA2, CDK2, CDK4, MAD1L1, TTK, RBL1, E2F1, NDC80, AURKB, MAD2L1BP, TRIP13, CDT1, TICRR, ESCO2, CDCA5, FBXO5
* Marked genes: CDKN2A, CDC45, CDC6, CDC25A, CDK2, CDK2, CDKN2A, ESCO2



**Fig.** **9** XENA SKCM marked pathway hsa04620

**(File path: Figure+Table/3.7\_Pathview\_通路可视化\_(XENA\_SKCM)/XENA-SKCM-marked-pathway-hsa04620.pdf)**

* Interactive figure: <https://www.genome.jp/pathway/hsa04620>
* Enriched genes: SPP1, CXCL11, CXCL10, CD86, CXCL8, IL1B, CCL5, AKT3, STAT1, CD80, LY96, CD14, CCL3, CXCL9, PIK3CD, TLR7, CCL4
* Marked genes: SPP1, CXCL11, CXCL10, CCL5, CCL3, CXCL9, PIK3CD, TLR7

# 4 总结

CAP2 生存分析 (Fig. **[2](#TCGA-SKCM-survival-curve-of-CAP2)** ) 显著。 属于差异表达基因 (Fig. **[4](#XENA-SKCM-Box-Plot-Of-DEGs)** )。 Melanoma\_vs\_Normal GSEA 富集分析 (Fig. **[5](#XENA-SKCM-GSEA-KEGG-enrichment)** , Fig. **[6](#XENA-SKCM-GSEA-plot-of-the-pathways)** ) 未找到 CAP2 所在通路。

在 MSigDB C2 基因集中可以富集到 CAP2，但与 Melanoma 无关 (Tab. **[4](#XENA-SKCM-filter-by-match-genes-g)** )。 尝试以关联分析寻找可能与 CAP2 相关的下游基因 (Tab. **[5](#XENA-SKCM-significant-correlation-analysis-data)** ), 见 Tab. **[6](#XENA-SKCM-filter-by-match-genes-f)** 以基因集 (CXCL9, CDC6, …(n = 14)., 来自于关联分析(Section: XENA\_SKCM)) 筛选到的通路。 Cell cycle, Toll-like receptor signaling pathway 可能是 CAP2 调控的下游 (或上游？) 通路，显著相关基因见 Fig. **[8](#XENA-SKCM-marked-pathway-hsa04110)** Fig. **[9](#XENA-SKCM-marked-pathway-hsa04620)**

# Reference

1. Colaprico, A. *et al.* TCGAbiolinks: An r/bioconductor package for integrative analysis of tcga data. *Nucleic Acids Research* **44**, (2015).

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

3. Wang, S. & Liu, X. The ucscxenatools r package: A toolkit for accessing genomics data from ucsc xena platform, from cancer multi-omics to single-cell rna-seq. *Journal of Open Source Software* **4**, 1627 (2019).

4. 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).

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

6. Luo, W. & Brouwer, C. Pathview: An r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics (Oxford, England)* **29**, 1830–1831 (2013).