

2025 年国际定向进化大赛实验记录

实验时间：2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

15. 转录组学数据分析

一、实验目的

通过转录组测序技术（RNA-seq）分析人膀胱移行细胞癌 T-24 细胞及其低、中、高浓度 RC48 耐药株（T24-RC48）的基因表达谱差异，筛选与 RC48 获得性耐药相关的关键基因及信号通路，为阐明耐药机制提供分子层面的数据支持。

二、实验流程

2.1 样本处理与建库测序

样本分组：T24 细胞（对照组）、低度耐药株、中度耐药株、高度耐药株，每组设置 3 个生物学重复。

RNA 提取与质控：采用 Agilent 2100 Bioanalyzer 检测 RNA 完整性（RIN 值 >7 ）。

文库构建：使用链特异性建库（保留转录方向信息），通过 Oligo(dT)磁珠富集 mRNA，随机打断后合成双链 cDNA，连接测序接头并进行 PCR 扩增。

测序：Illumina 平台进行双端测序（PE150），单样本数据量 $\geq 6G$ 。

2.2 生物信息分析

数据质控：过滤低质量、含接头及 N 比例过高的 reads，获得 clean reads（Q20 $\geq 97\%$ ，Q30 $\geq 93\%$ ）。

参考基因组比对：使用 HISAT2 软件将 clean reads 比对至人类参考基因组（比对率 $>89\%$ ）。

基因定量：通过 featureCounts 统计基因表达量（FPKM 标准化）。

差异分析：采用 DESeq2（有生物学重复）筛选差异基因（ $|\log_2(\text{FoldChange})| \geq 1$ 且 $\text{padj} \leq 0.05$ ）。

功能富集分析：基于 GO、KEGG、Reactome、DO、DisGeNET 数据库，通过 clusterProfiler 进行通路富集（ $\text{padj} < 0.05$ ）。

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

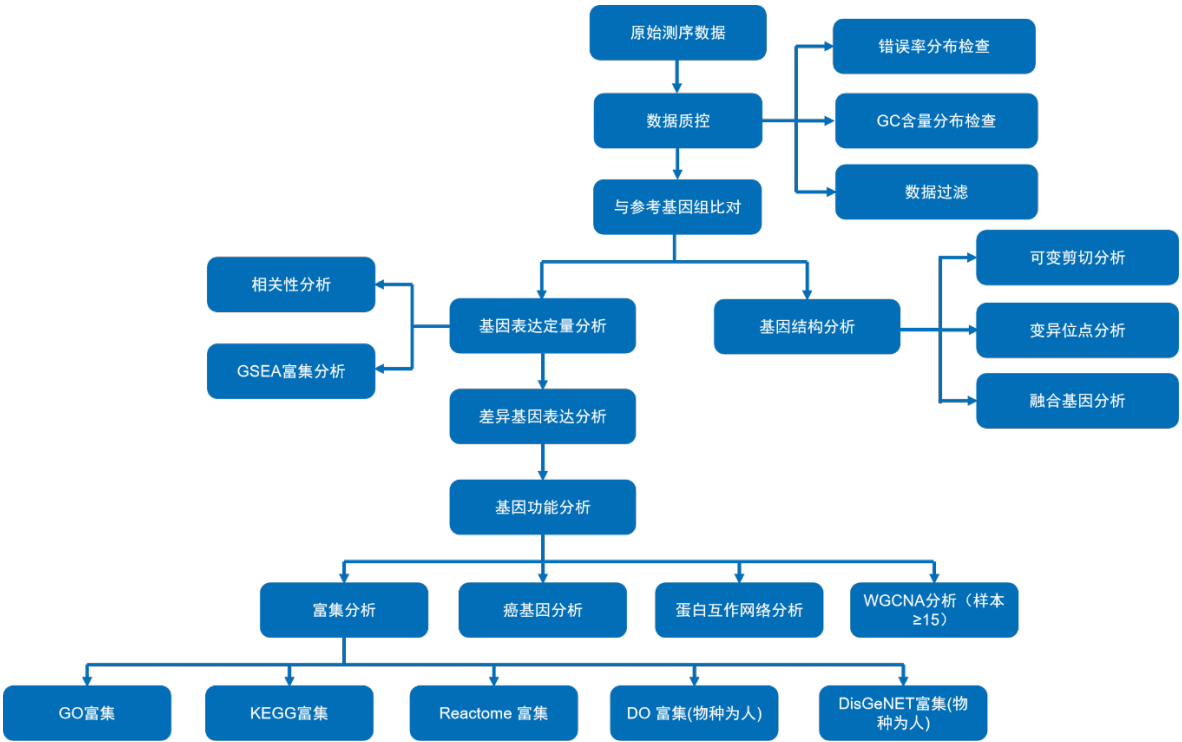


图 1 RNA-seq 信息分析技术流程

三、实验结果

3.1 差异基因分析

每个比较组合的差异基因（包括上调和下调）数目统计如下表所示：

表 1 差异基因统计结果

比较组	总差异基因数	上调基因数	下调基因数
T24_moderate vs T24_low	843	359	484
T24_high vs T24_low	768	413	355
T24_high vs T24_moderate	572	360	212
T24_low vs T24_con	948	569	379
T24_moderate vs T24_con	2218	1089	1129
T24_high vs T24_con	2059	1118	941

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

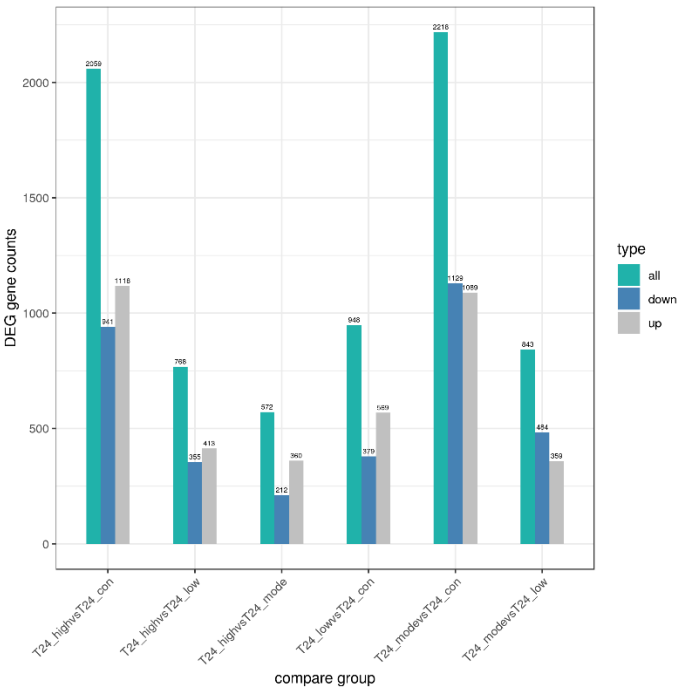


图 2 差异比较组合差异基因数目统计柱状图

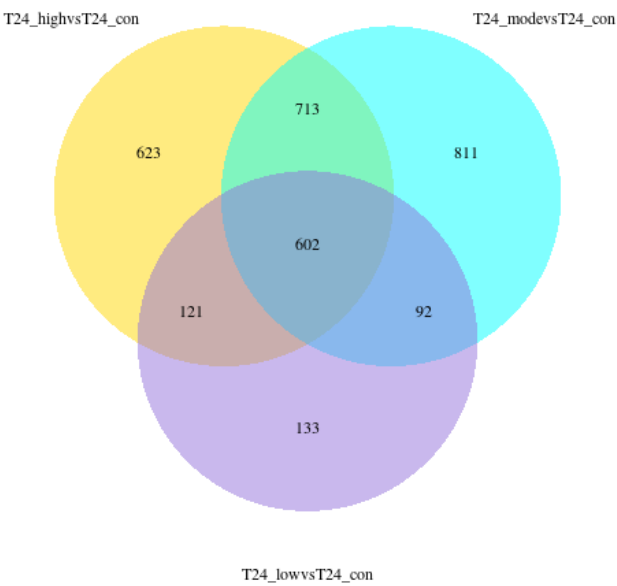


图 3 差异基因韦恩图

火山图可直观展示每个比较组合的差异基因分布情况，如下图所示。图中横坐标表示基因在处理 and 对照两组中的表达倍数变化(log2FoldChange)，纵坐标表示基因在处理 and 对照两组中表达差异的显著性水平(-log10padj 或 -log10pvalue)。为上调基因用红色点表示，下调基因用绿色点表示。

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

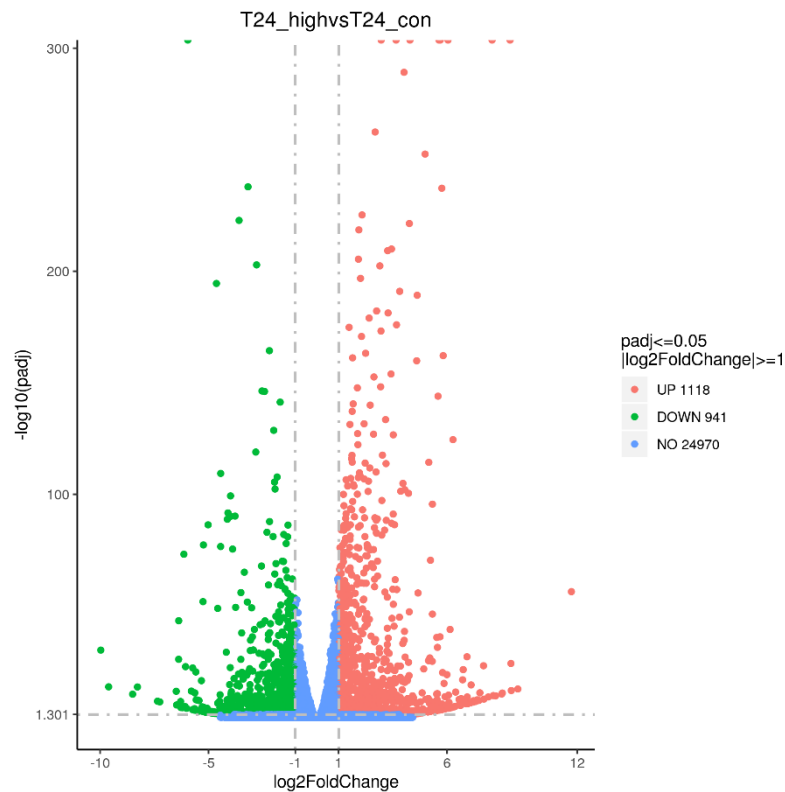


图 4 高耐药株相对于野生型上调基因（红色）和下调基因（绿色）分布火山图

3.2 功能富集分析

3.2.1 GO 富集分析

基因本体（Gene Ontology, GO）是一个标准化的功能分类体系，提供了一套动态更新的标准词汇表，并以以下三个方面描述生物体中基因和基因产物的属性，包括参与的生物过程（Biological Process, BP），分子功能（Molecular Function, MF）和细胞组分（Cellular Component, CC）。利用 R 包 clusterprofiler 对差异表达基因进行 GO 富集分析，从 GO 富集分析结果中，选取最显著的 30 个 Term 绘制柱状图进行展示，若不足 30 个，则绘制所有 Term，如下图所示。图中横坐标为 GO Term，纵坐标为 GO Term 富集的显著性水平，数值越高越显著，柱子上的数值代表富集该 term 的差异基因个数，不同颜色分别代表 BP，CC，MF 三个 GO 子类。

2025 年国际定向进化大赛实验记录

实验时间: 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

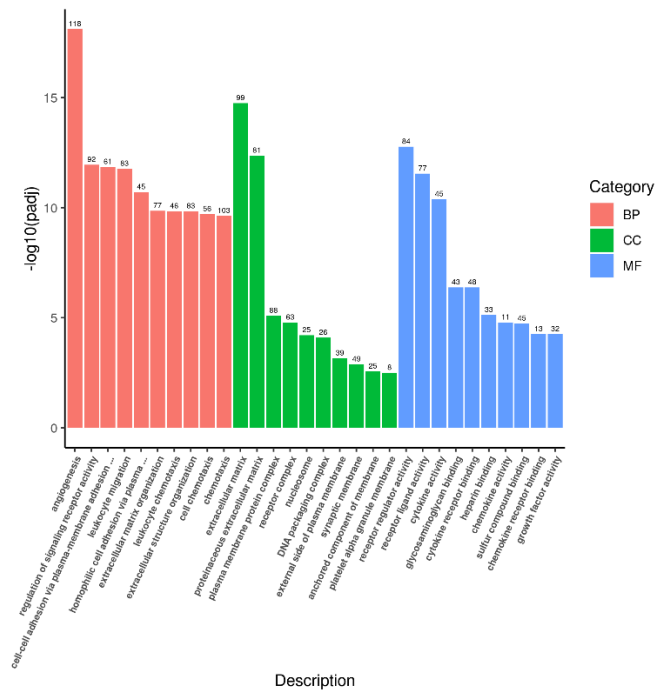


图 5 GO 富集分析柱状图 (T24_high vs T24_con)

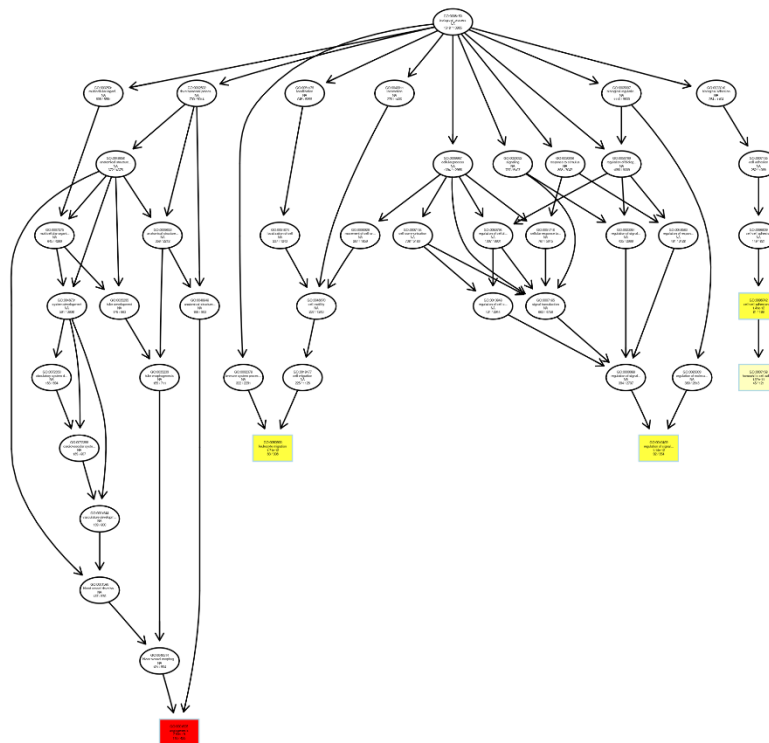


图 6 GO 富集分析 DAG 图 (T24_high vs T24_con)

3.2.2 KEGG 通路富集

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

KEGG 通路富集分析方法与 GO 富集分析相似，即以 KEGG 通路为单位，以参考基因组为背景，通过 Fisher 精确检验（Fisher’s Exact Test），来分析计算各个通路基因富集度的显著性水平，从而确定受到显著影响的代谢和信号转导途径。分析结果如下：

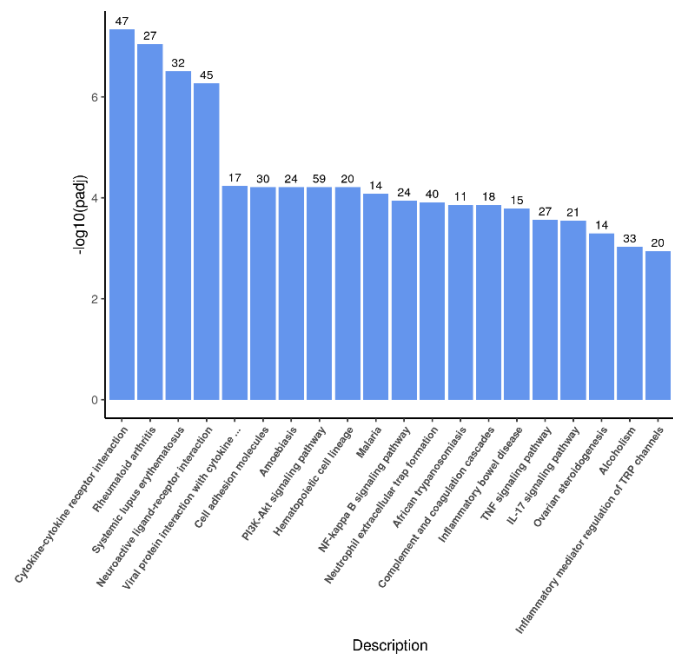


图 7 KEGG 富集分析柱状图（T24_high vs T24_con）

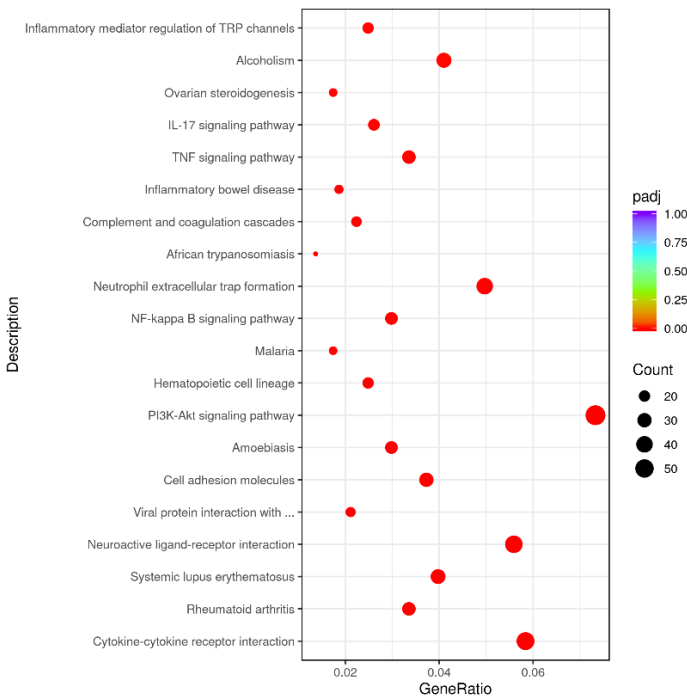


图 8 KEGG 富集分析散点图（T24_high vs T24_con）

3.2.3 Reactome 通路富集：

Reactome 数据库汇集了人类等模式物种各项反应及生物学通路。Reactome 通路富集以

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

padj 小于 0.05 作为显著性富集的阈值，从 Reactome 富集分析结果中，选取最显著的 20 个 Reactome 通路绘制柱状图进行展示，若不足 20 个，则绘制所有通路，如下图所示。图中横坐标为 Reactome 通路，纵坐标为通路富集的显著性水平，数值越高越显著,柱子上的数值代表富集到该通路的差异基因个数。

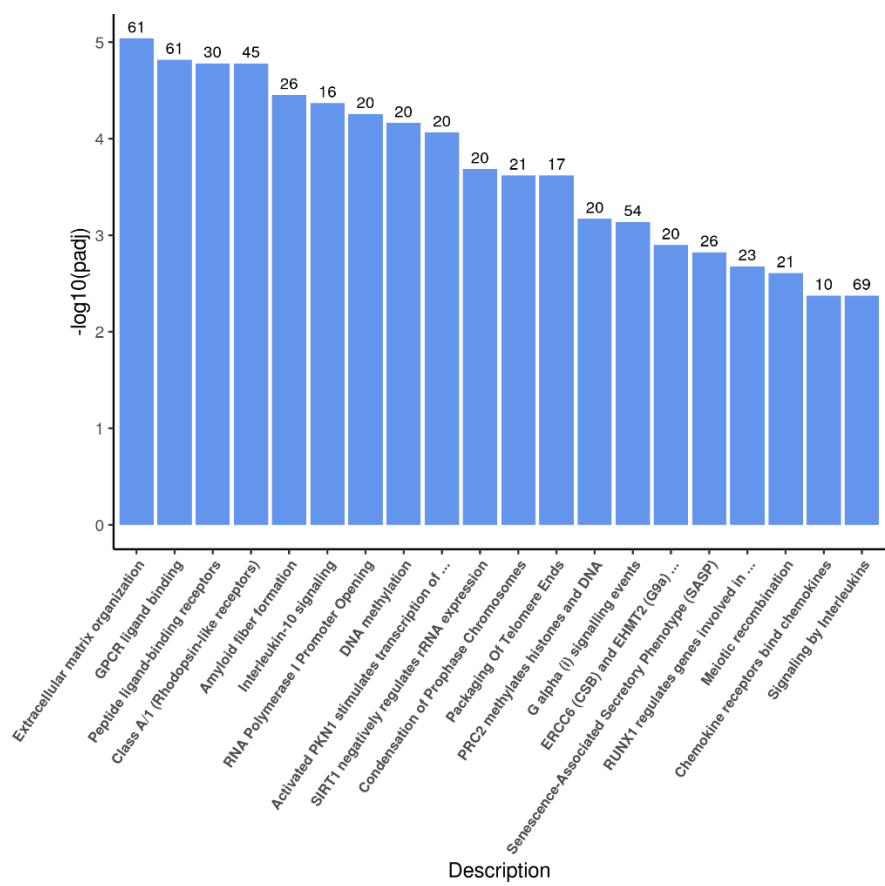


图 9 Reactome 通路富集分析柱状图（T24_high vs T24_con）

从 Reactome 富集分析结果中,选取最显著的 20 个 Reactome 通路绘制散点图进行展示，若不足 20 个，则绘制所有通路，如下图所示。图中横坐标为注释到 Reactome 通路上的差异基因数与差异基因总数的比值，纵坐标为 Reactome 通路，点的大小代表注释到 Reactome 通路上的基因数，颜色从红到紫代表富集的显著性大小。

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

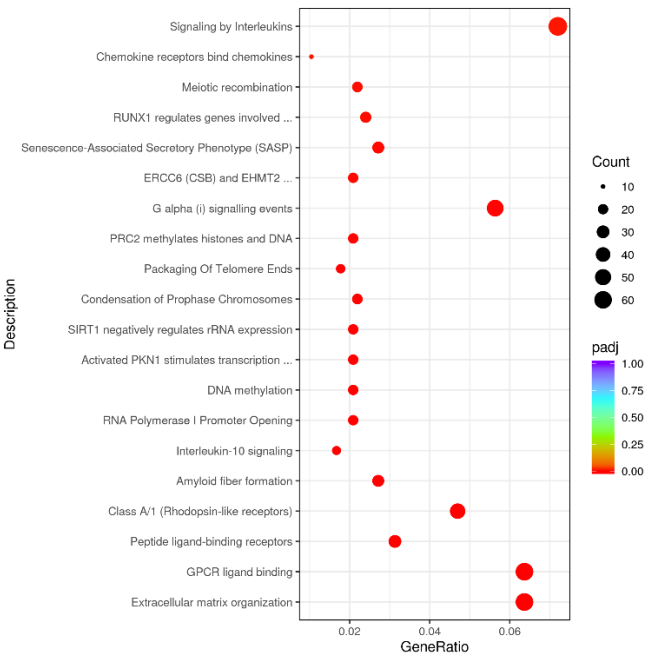


图 10 Reactome 通路富集分析散点图（T24_high vs T24_con）

3.2.4 疾病相关富集（DO/DisGeNET）:

DO(Disease Ontology)是描述人类基因功能与疾病相关的数据库。DO 富集以 padj 小于 0.05 作为显著性富集的阈值，从 DO 富集分析结果中，选取最显著的 20 个 DO 通路绘制柱状图进行展示，若不足 20 个，则绘制所有通路，如下图所示。图中横坐标为 DO 通路，纵坐标为通路富集的显著性水平，数值越高越显著,柱子上的数值代表富集到该通路的差异基因个数。

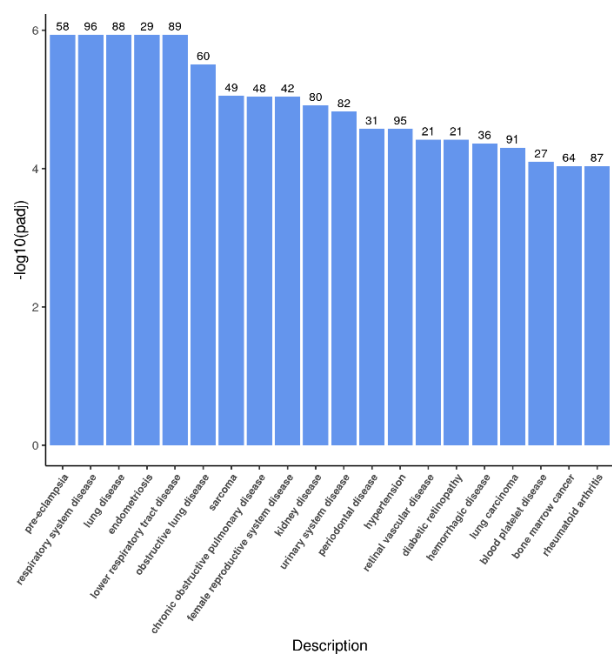


图 11 DO 通路富集分析柱状图（T24_high vs T24_con）

从 DO 富集分析结果中，选取最显著的 20 个 DO 通路绘制散点图进行展示，若不足 20

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

个，则绘制所有通路，如下图所示。图中横坐标为注释到 DO 通路上的差异基因数与差异基因总数的比值，纵坐标为 DO 通路，点的大小代表注释到 DO 通路上的基因数，颜色从红到紫代表富集的显著性大小。

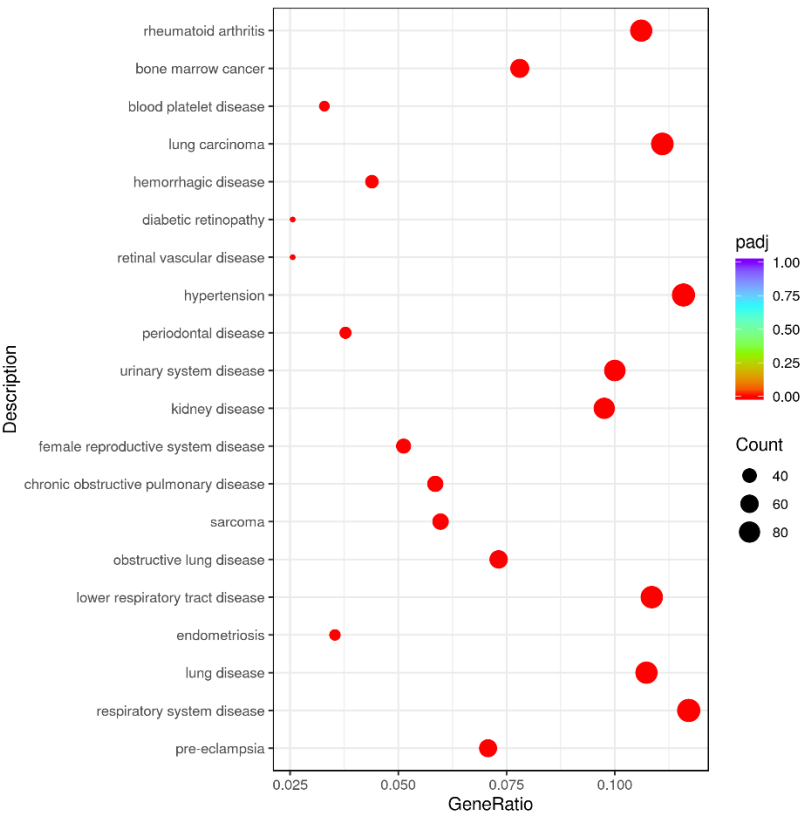


图 12 DO 通路富集分析散点图（T24_high vs T24_con）

DisGeNET 数据库整合了人类疾病相关基因。DisGeNET 富集以 padj 小于 0.05 作为显著性富集的阈值，从 DisGeNET 富集分析结果中，选取最显著的 20 个 Term 绘制柱状图进行展示，若不足 20 个，则绘制所有通路，如下图所示。图中横坐标为 DisGeNET 通路，纵坐标为通路富集的显著性水平，数值越高越显著，柱子上的数值代表富集到该通路的差异基因个数。

2025 年国际定向进化大赛实验记录

实验时间： 2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

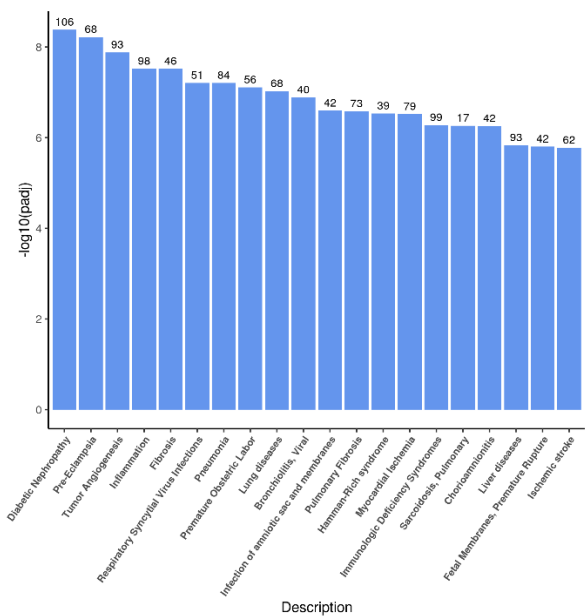


图 13 DisGeNET 通路富集分析柱状图（T24_high vs T24_con）

从 DisGeNET 富集分析结果中，选取最显著的 20 个 Term 绘制散点图进行展示，若不足 20 个，则绘制所有通路，如下图所示。图中横坐标为注释到 DisGeNET 通路上的差异基因数与差异基因总数的比值，纵坐标为 DisGeNET 通路，点的大小代表注释到 DO 通路上的基因数，颜色从红到紫代表富集的显著性大小。

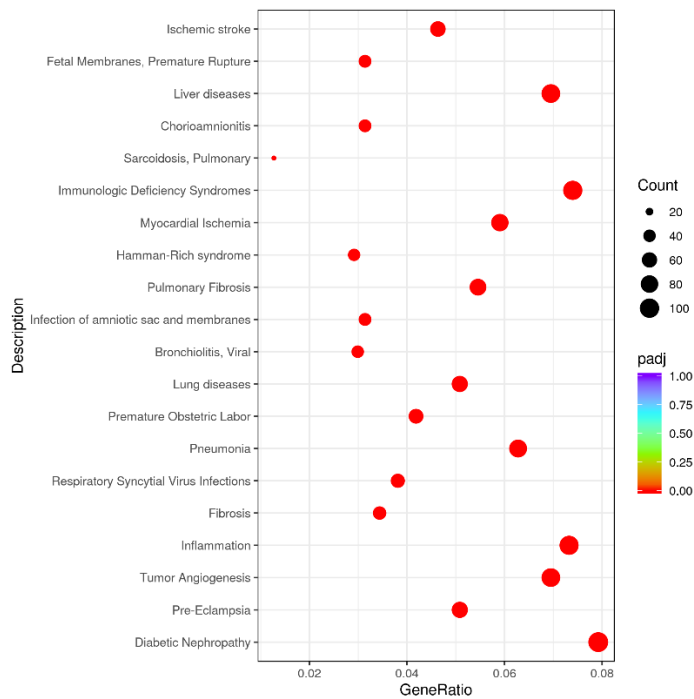


图 14 DisGeNET 通路富集分析散点图（T24_high vs T24_con）

四、结果分析

2025 年国际定向进化大赛实验记录

实验时间：2025 年 6 月 6 日 18:00 — 2025 年 6 月 9 日 21:00

转录组学分析显示，T24-RC48 耐药株的耐药机制与胆固醇代谢通路、PI3K-Akt 信号通路及细胞外基质 (ECM) 重构密切相关。具体而言，差异基因在固醇/胆固醇合成通路 (GO、KEGG、Reactome) 中显著富集，提示耐药细胞可能通过脂质代谢重编程增强生存能力；PI3K-Akt 信号通路的激活则可能通过抑制细胞凋亡促进耐药性；同时，ECM 重构相关基因 (如胶原蛋白基因) 的上调，可能通过增强细胞黏附或形成物理屏障降低药物敏感性。此外，研究发现多个潜在耐药标志物，包括与肿瘤耐药相关的癌基因 (如 DIP2A、NRIP1) 及融合基因事件 (如 PRR16-CTD-2334019.1)，需通过功能实验进一步验证其作用。实验可靠性方面，生物学重复间高相关性 ($R^2 > 0.8$) 及主成分分析显示的组内一致性，均支持数据的稳健性。综上，胆固醇代谢、ECM 重构及 PI3K-Akt 通路可能是介导 RC48 耐药的关键分子机制，后续需结合蛋白质组学及功能研究深入解析候选基因的调控网络，为逆转耐药策略提供理论依据。