

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T24细胞及其低、中、高浓度IC₅₀耐药株的基因表达谱差异,筛选与IC₅₀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/中/高浓度耐药株,每组设置了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)。

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DNA,连接测序接头并扩增PCR。

测序:Illumina平台进行双端测序(PE150),单样本数据量>6G。

2.2 生物信息分析

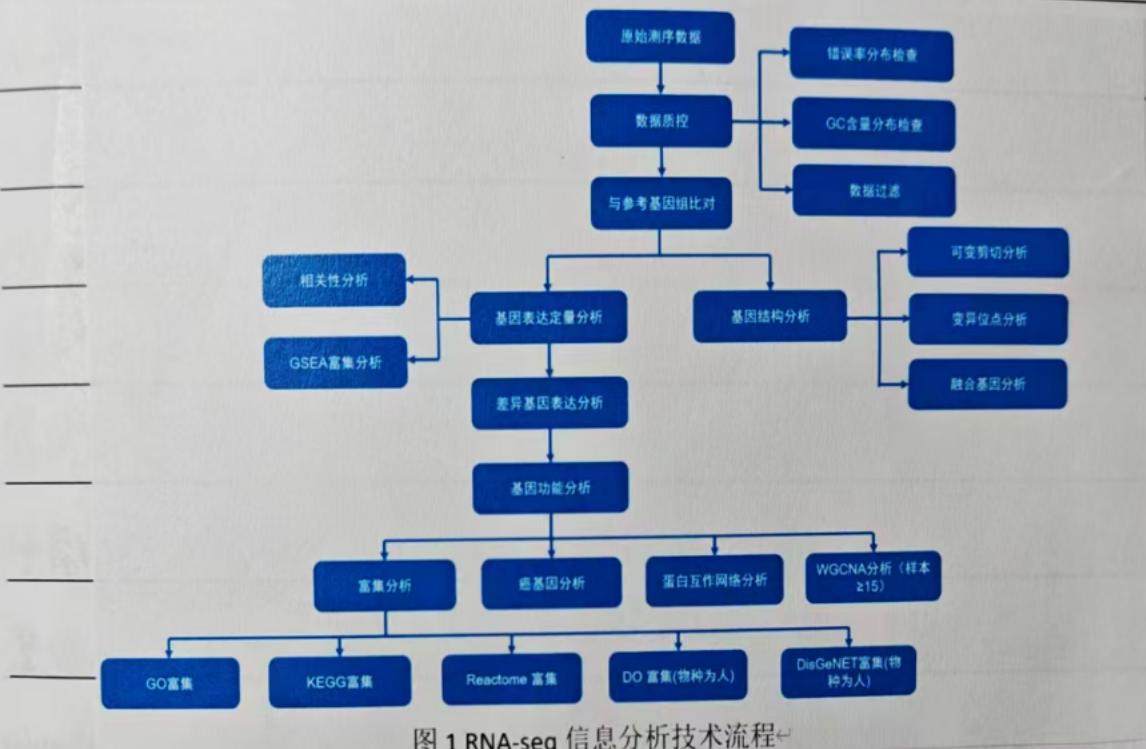
数据质控:过滤低质量、含接头及N比例过高的reads,获得clean reads(Q20>97%, Q30>93%)。

参考基因组比对:使用 HISAT2 软件将 clean read 对比至人类参考基因组(比对率>89%)。

基因定量：通过 feature Counts 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 ($|log_2(FoldChange)| > 1$ 且 $padj < 0.05$)

功能富集分析：基于 GO、KEGG、Reactome、Do、PisGeNET 数据库，通过 clusterProfiler 进行通路富集 ($padj < 0.05$)。



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

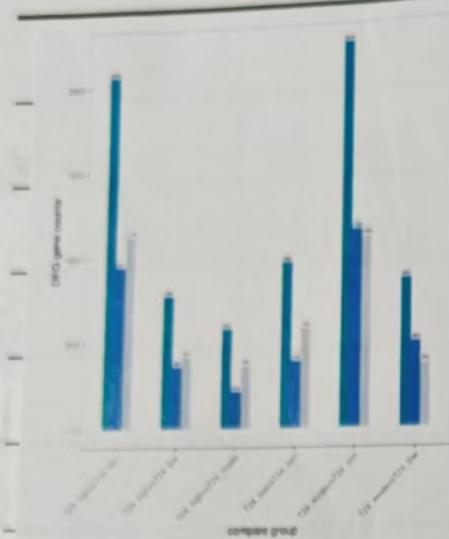


图 2 差异比较组合差异基因数目统计柱状图

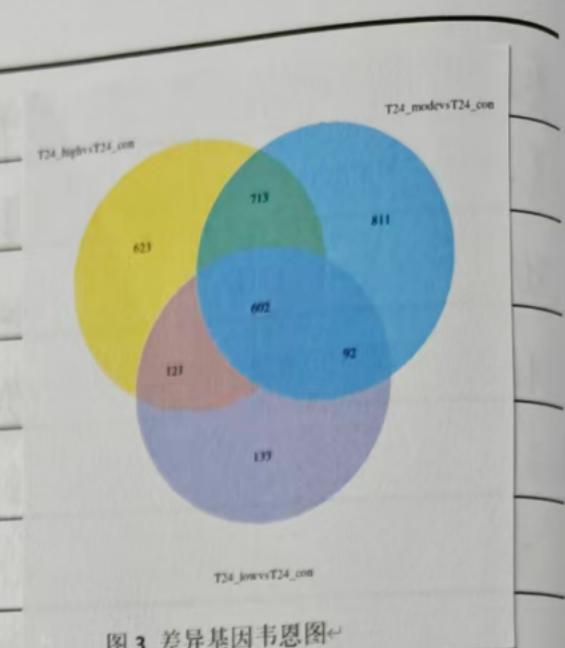


图 3 差异基因韦恩图

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和对照两组中的表达倍数变化($\log_2(\text{FoldChange})$)，纵坐标表示基因在处理和对照两组中表达差异的显著性水平($-\log_{10}(\text{padj})$ 或 $-\log_{10}(\text{pvalue})$)，
为上调基因用红色点表示，下调基因用绿色点表示。

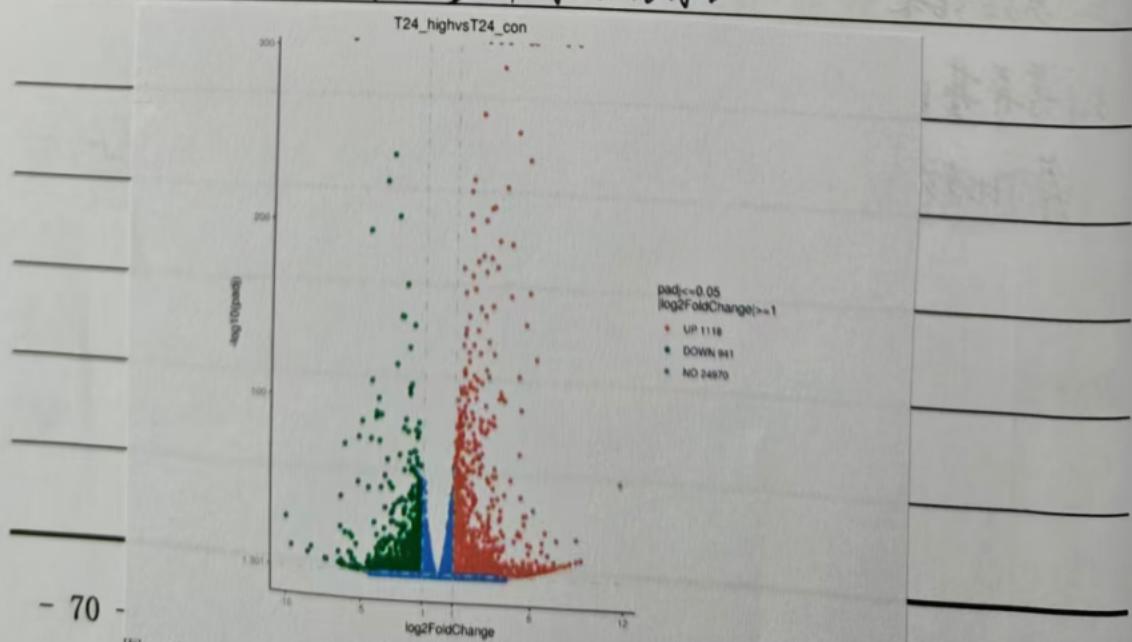
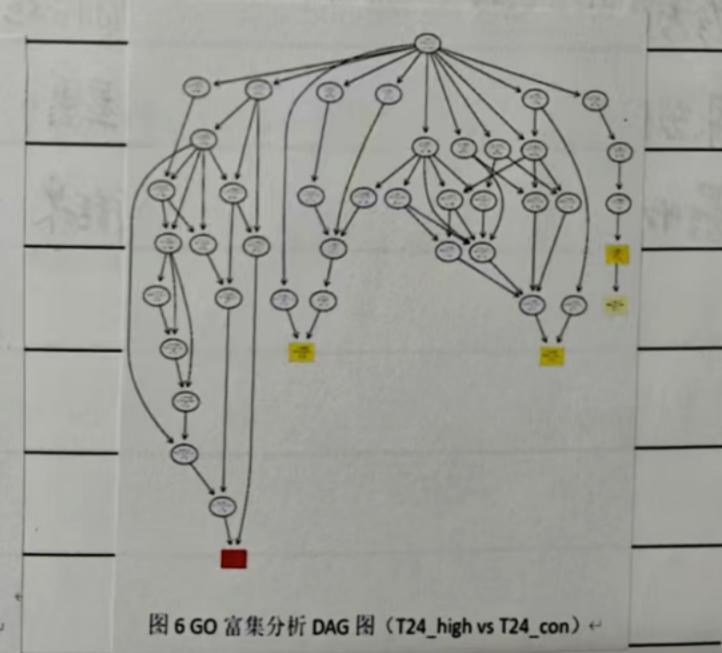
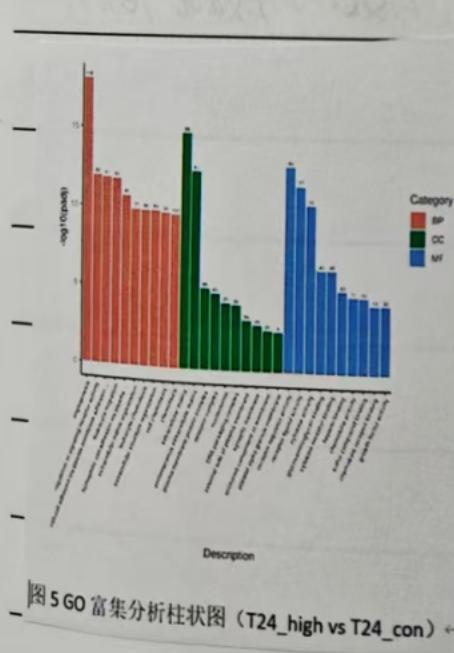


图 4 高耐药株相对于野生型上调基因（红色）和下调基因（绿色）分布火山图

3.2 功能富集分析

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化词汇表,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
的差异基因数,不同颜色分别代表BP, CC, MF三个GO子类。



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。

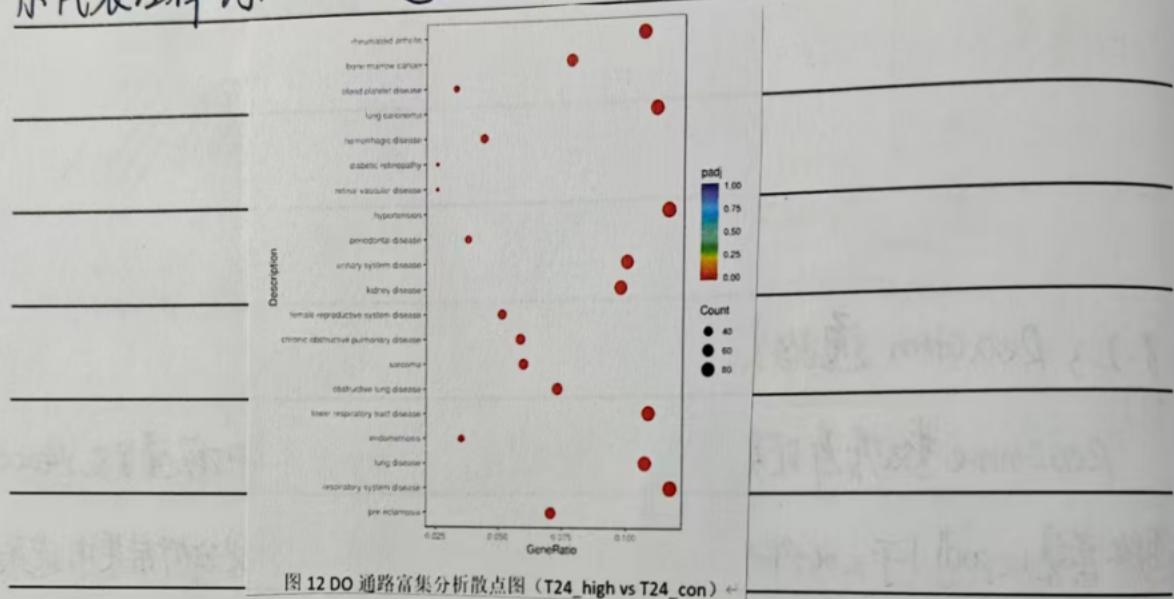


图 12 DO 通路富集分析散点图 (T24_high vs T24_con)

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。

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个，则绘制所有通路，所下图所示。图中横坐标为注释到DO通路上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通路上基因数，颜色从红到紫代表富集的显著性大小。

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，所下图所示。图中横坐标为DisGeNET通路，纵坐标为通路富集的显著水平，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。

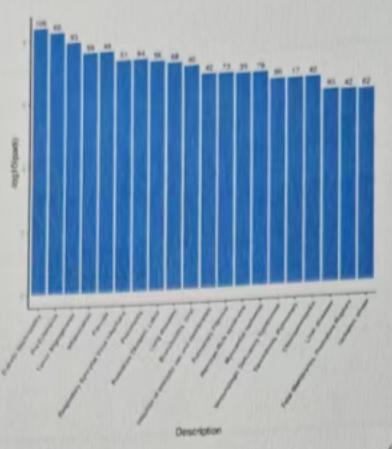


图 13 DisGeNET 通路富集分析柱状图 (T24_high vs T24_con)

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 GO 通路上的基因数，颜色从红到紫代表富集的显著性大小。

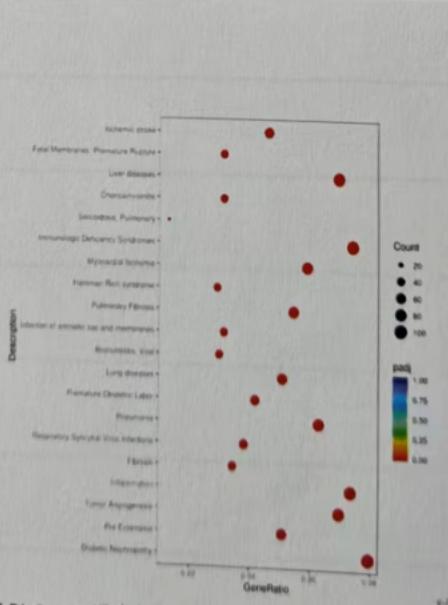


图 14 DisGeNET 通路富集分析散点图 (T24_high vs T24_con)

四、结果分析

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重构相关基因的上调, 可能通过增强细胞系部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(BCL2A, NRP1)及融合基因件(如PRR16-CTD-2334019.D), 需通过功能实验进一步验证其作用。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。