Experiment time: 2025-06-06, 18:00 - 2025-06-09, 21:00

Transcriptomics Data Analysis

I. Objective of the experiment

By analyzing the gene expression profile differences between human bladder transitional cell carcinoma T-24 cells and low, medium and high concentration RC48 drug-resistant strains (T24-RC48) through transcriptome sequencing technology (RNA-seq), this study screened the key genes and signal pathways related to RC48 acquired drug resistance, providing molecular data support for elucidating the drug resistance mechanism.

II. Experimental procedure

2.1 Sample processing and library sequencing

Sample groups: T24 cells (control group), low drug resistance strain, moderate drug resistance strain, and high drug resistance strain, with 3 biological replicates in each group.

RNA extraction and quality control: Agilent 2100 Bioanalyzer was used to detect RNA integrity (RIN value>7).

Library construction: Use chain-specific library construction (retaining the transcription direction information), enrich mRNA by Oligo (dT) magnetic beads, synthesize double-stranded cDNA after random fragmentation, and then connect sequencing adapters for PCR amplification.

Sequencing: Illumina platform was used for double-end sequencing (PE150), and the single sample data volume was greater than or equal to 6 gigabases (Gb).

2.2 Bioinformatics analysis

Data quality control: Filter low quality reads with high proportion of connectors and N to obtain clean reads (Q20≥97%, Q30≥93%).

Reference genome alignment: Clean reads were aligned to the human reference genome using HISAT2 software (alignment rate> 89%).

Gene quantification: Gene expression was calculated by featureCounts (standardized FPKM).

Differential expression analysis: DESeq2 (accounting for biological replicates) was used to identify differentially expressed genes (DEGs) ($|\log2(\text{FoldChange})| \geq 1$ and $\text{p.adj.} \leq 0.05$).

Functional enrichment analysis: Based on GO, KEGG, Reactome, DO and DisGeNET databases, pathway enrichment was performed by clusterProfiler (adjusted p-value (p.adj.) <0.05).

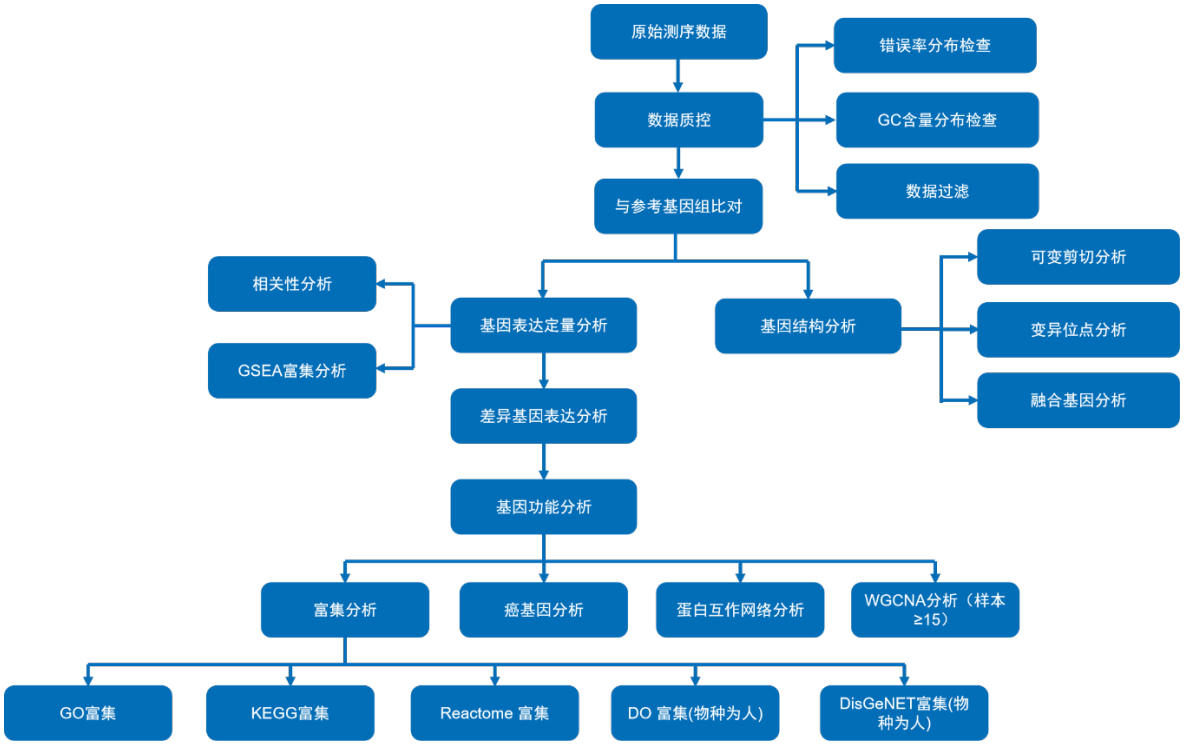

Figure 1 RNA-seq information analysis technology flow

III. Experimental results

3.1 Analysis of differential genes

The number of differential genes (including upregulation and downregulation) in each comparison is statistically shown in the following table:

Table 1 Summary of differentially expressed genes

Comparison	Total DEGs	Up-regulated	Down-regulated
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

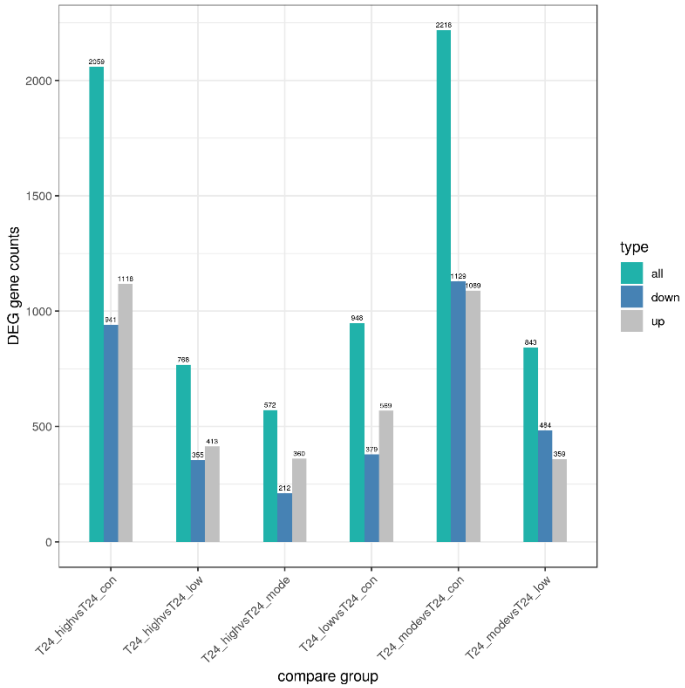

Figure 2 Bar chart of statistical number of differential genes in differential comparison combinations

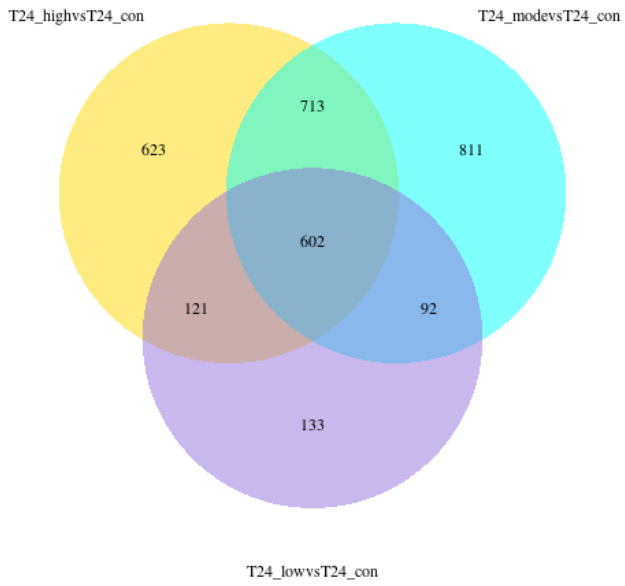

Figure 3 Venn diagram of differentially expressed genes

The volcano plot visually displays the distribution of DEGs for each comparison, as illustrated in the figure below. The x-axis shows log2-fold changes in gene expression between treatment and control groups, while the y-axis indicates significance levels (expressed as-log10padj or-log10pvalue). Up-regulated genes are represented by red dots, whereas down-regulated genes are indicated by green

dots.

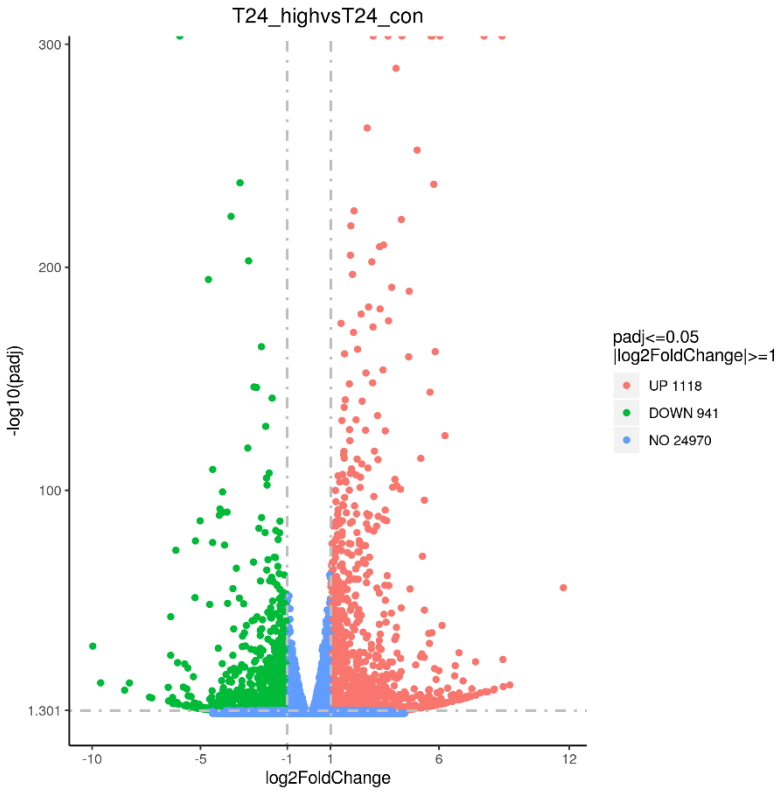

Figure 4 Volcano plot of the distribution of up-regulated genes (red) and down-regulated genes (green) in highly drug-resistant strains compared with wild-type

3.2 Functional enrichment analysis

3.2.1 GO enrichment analysis

Gene Ontology (GO) is a standardized functional classification system that provides a dynamically updated standardized vocabulary. It describes the attributes of genes and gene products in organisms through three aspects: biological processes (BP), molecular functions (MF), and cellular components (CC). Using the R package clusterprofiler for differential expression gene GO enrichment analysis, the top 30 most significant terms are selected from the results to generate bar charts. If fewer than 30 terms are identified, all terms are plotted as shown in the figure below. The x-axis represents GO terms, while the y-axis shows the significance level of GO term enrichment (higher values indicate higher significance). The numerical values on the bars indicate the number of DEGs enriched in each term, with different colors indicating BP, CC, and MF subcategories.

Experiment time: 2025-06-06, 18:00 - 2025-06-09, 21:00

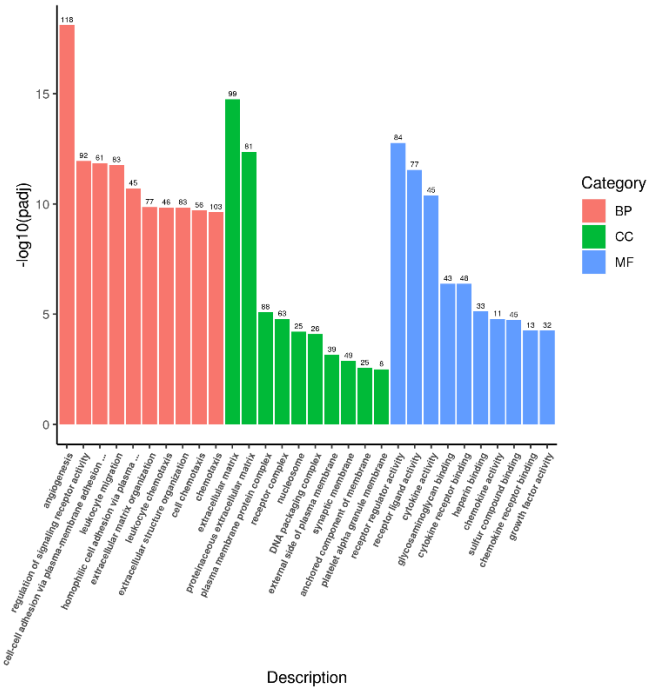

Figure 5 GO enrichment analysis (T24_high vs T24_con)

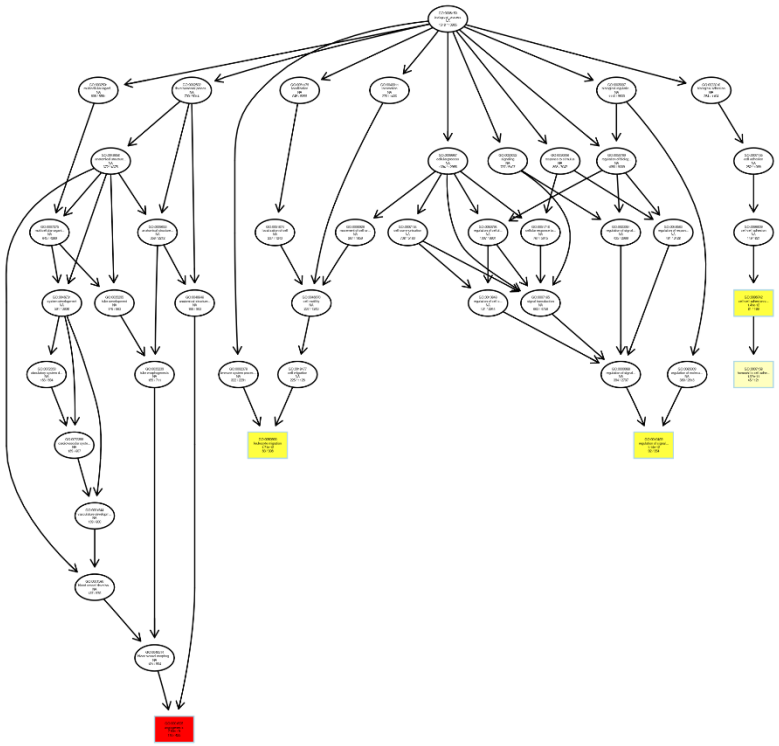

Figure 6 Directed acyclic graph (DAG) of significant GO terms (T24_high vs T24_con)

3.2.2 KEGG pathway enrichment

The KEGG pathway enrichment analysis method is similar to GO enrichment analysis. Using KEGG pathways as the unit and the reference genome as the background, we employ Fisher's exact test to calculate the significance levels of gene enrichment in each pathway. This approach helps identify significantly affected metabolic and signaling pathways. The analysis results are as follows:

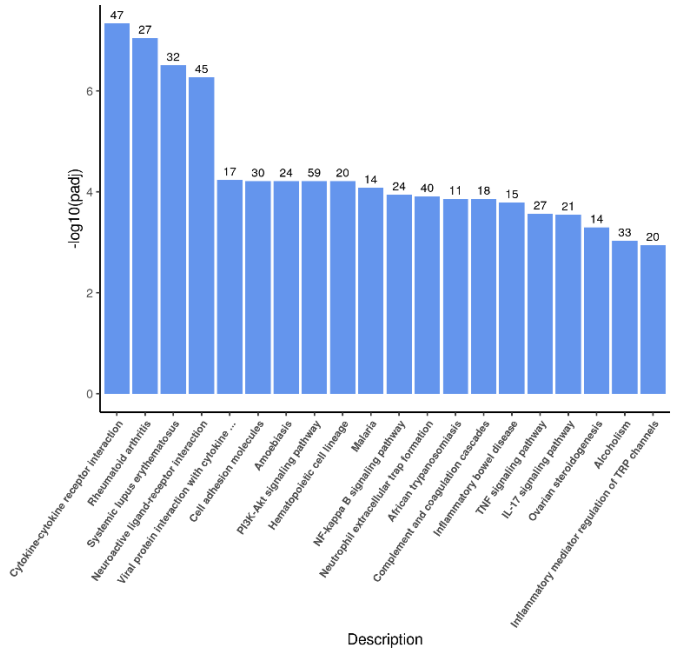

Figure 7 KEGG enrichment analysis (T24_high vs T24_con)

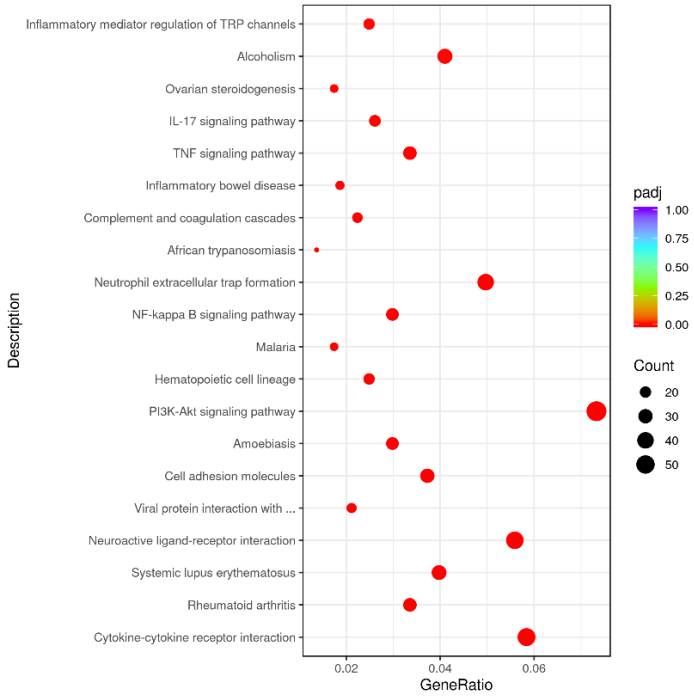

Figure 8 KEGG enrichment analysis scatter plot (T24_high vs T24_con)

3.2.3 Reactome pathway enrichment

Experiment time: 2025-06-06, 18:00 - 2025-06-09, 21:00

The Reactome database compiles biological pathways and reactions from model species including humans. Using a p.adj. value below 0.05 as the threshold for significant enrichment, we selected the top 20 most significantly enriched Reactome pathways for visualization through bar charts. When fewer than 20 pathways are identified, all pathways are plotted. As shown in the figure below, the x-axis represents Reactome pathways, while the y-axis indicates the significance level of pathway enrichment (higher values indicate greater significance). The numbers on each column represent the number of differentially expressed genes within that pathway.

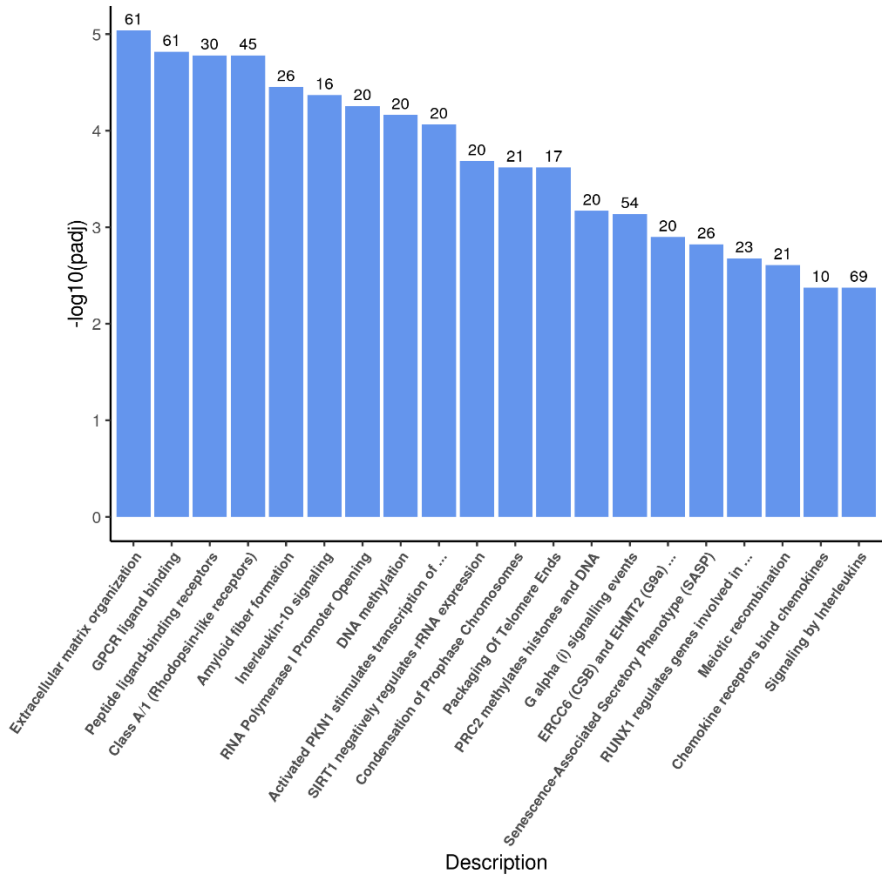

Figure 9 Reactome pathway enrichment analysis (T24_high vs T24_con)

Based on Reactome enrichment analysis results, we selected the top 20 most significant Reactome pathways to plot as a scatter diagram. If fewer than 20 pathways are identified, all pathways are plotted (as shown in the figure below). The x-axis displays the ratio of differential genes annotated to Reactome pathways to the total number of differential genes, while the y-axis represents the

Reactome pathways themselves. Dot sizes indicate the number of genes annotated to each pathway, with colors ranging from red to purple reflecting the significance level of the enrichment.

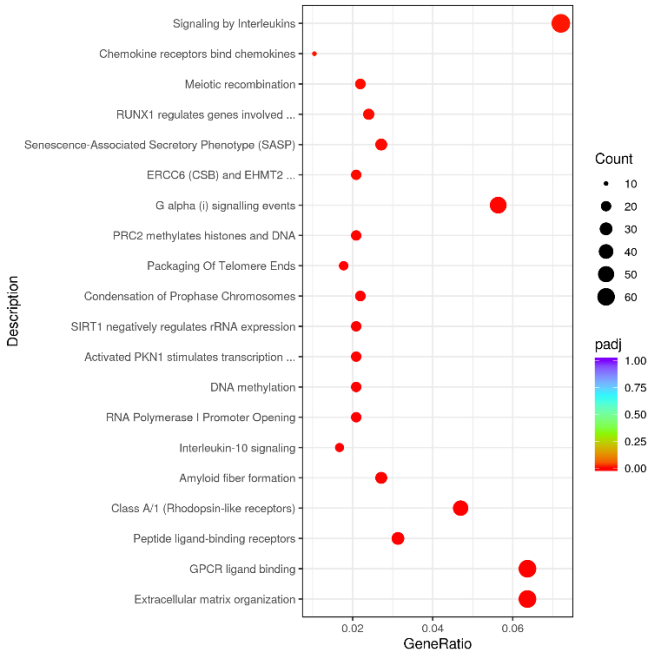

Figure 10 Reactome pathway enrichment analysis scatter plot (T24_high vs T24_con)

3.2.4 Disease-related Enrichment (DO/DisGeNET):

The Disease Ontology (DO) is a database that describes the functional relationships between human genes and diseases. DO enrichment analysis uses a p.adj. value below 0.05 as the threshold for significant enrichment. The top 20 most significant DO pathways are plotted in a bar chart, while fewer than 20 pathways are displayed as all pathways, as shown in the figure below. The x-axis represents DO pathways, and the y-axis shows the significance level of pathway enrichment (higher values indicate higher significance). The numbers on the bars represent the number of differential genes enriched in that pathway.

Experiment time: 2025-06-06, 18:00 - 2025-06-09, 21:00

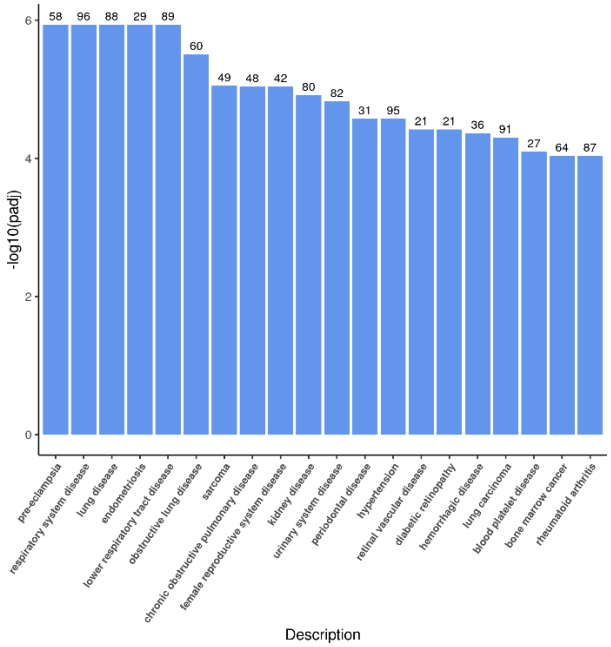

Figure 11 DO pathway enrichment analysis (T24_high vs T24_con)

Based on the DO enrichment analysis results, we selected the top 20 most significant DO pathways to plot as a scatter diagram. If fewer than 20 pathways are identified, all pathways are plotted, as shown in the figure below. The x-axis displays the ratio of annotated differential genes to total differential genes within each DO pathway, while the y-axis represents the DO pathways themselves. Dot sizes indicate the number of genes annotated to each pathway, with colors ranging from red to purple reflecting the significance level of enrichment.

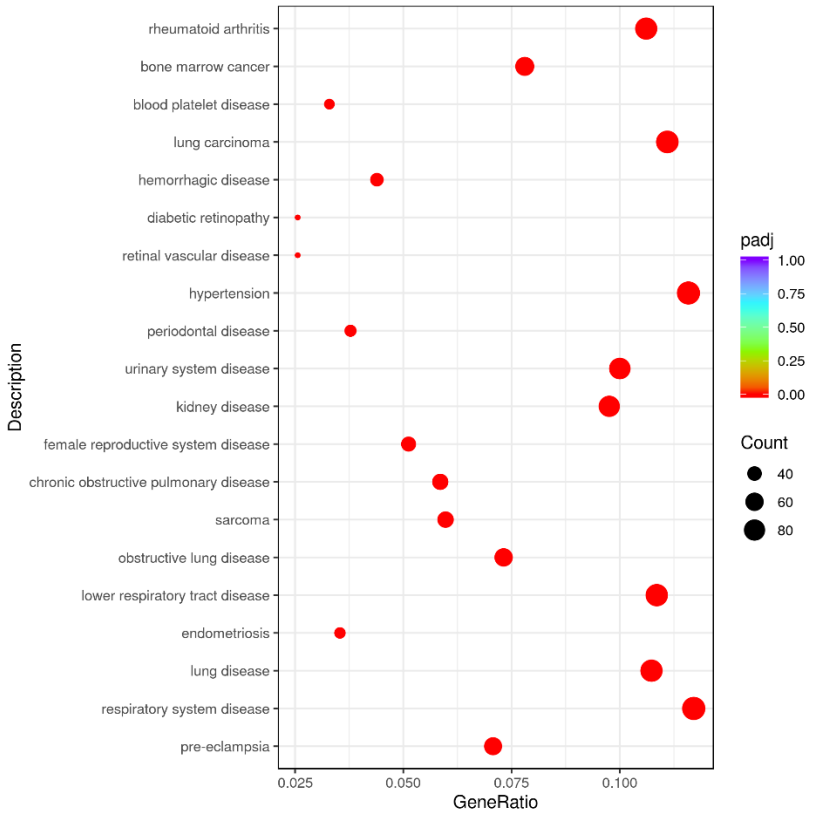

Figure 12 DO pathway enrichment analysis scatter plot (T24_high vs T24_con)

The DisGeNET database integrates human disease-related genes. DisGeNET enrichment analysis uses a p.adj. value below 0.05 as the threshold for significant enrichment. From the enrichment analysis results, the top 20 most significant terms are selected to plot a bar chart. If fewer than 20 terms are identified, all pathways are plotted, as shown in the figure below. The x-axis represents DisGeNET pathways, while the y-axis shows the significance level of pathway enrichment (higher values indicate higher significance). The numbers on the bars represent the number of differential genes enriched to that pathway.

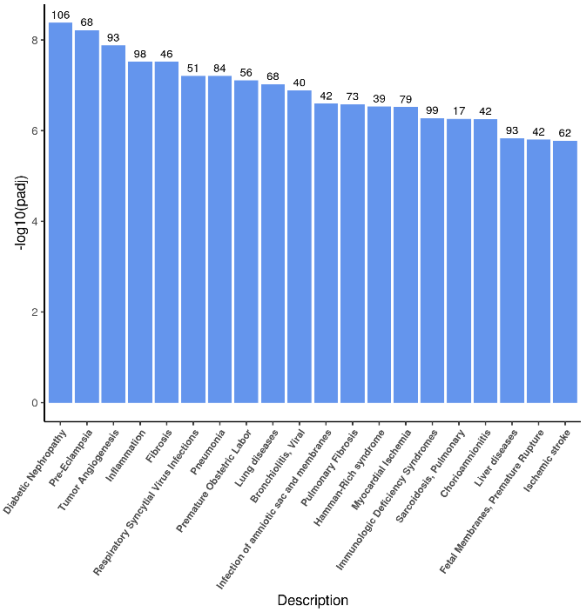

Figure 13 DisGeNET pathway enrichment analysis (T24_high vs T24_con)

Based on the enrichment analysis results from DisGeNET, we selected the top 20 most significant terms to plot as a scatter diagram. If fewer than 20 terms are identified, all pathways are plotted, as shown in the figure below. The x-axis displays the ratio of differential genes annotated to DisGeNET pathways to the total number of differential genes, while the y-axis shows the DisGeNET pathways themselves. Dot size represents the number of genes annotated to each pathway, with colors ranging from red to purple representing increasing significance levels of enrichment.

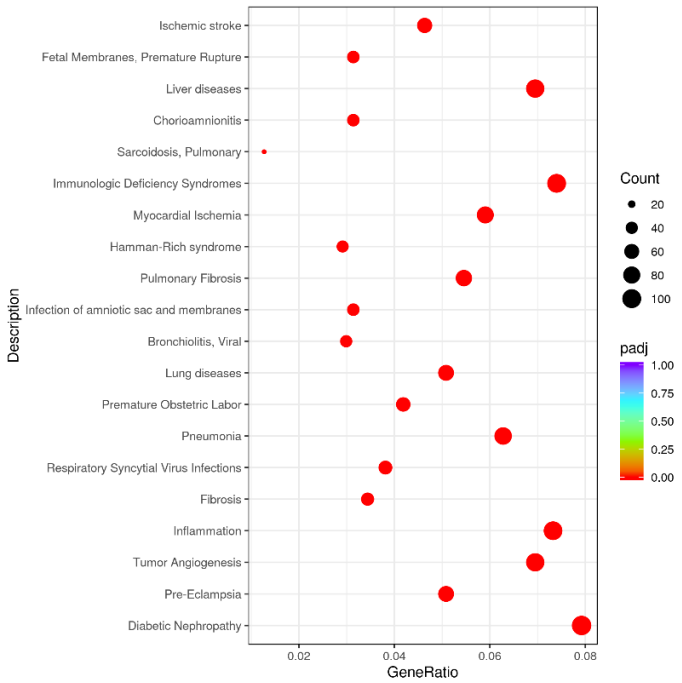

Figure 14 DisGeNET pathway enrichment analysis scatter plot (T24_high vs T24_con)

IV. Results analysis

Transcriptomic analysis revealed that the drug resistance mechanisms of T24-RC48 strains are closely associated with cholesterol metabolism pathways, PI3K-Akt signaling pathways, and extracellular matrix (ECM) remodeling. Specifically, differential genes were significantly enriched in steroid/cholesterol synthesis pathways (GO, KEGG, Reactome), suggesting that drug-resistant cells may enhance survival through lipid metabolism reprogramming. The activation of PI3K-Akt signaling pathways might promote resistance by inhibiting apoptosis. Meanwhile, up-regulated ECM remodeling-related genes (e.g., collagen genes) could reduce drug sensitivity by enhancing cell adhesion or forming physical barriers. Additionally, multiple potential resistance markers were identified, including tumor resistance-associated oncogenes (e.g., DIP2A, NRIP1) and fusion gene events (e.g., PRR16-CTD-2334019.1), which require further validation through functional experiments. Experimental reliability was supported by high correlation between biological replicates (R^2>0.8) and intra-group consistency indicated by principal component analysis. In summary, cholesterol metabolism, ECM remodeling, and PI3K-Akt pathways may be key molecular mechanisms mediating RC48 drug resistance. Subsequent studies should integrate proteomics and functional research to further elucidate regulatory networks of candidate genes, providing theoretical foundations for reversing drug resistance strategies.