International Directed Evolution Competition Lab Notebook

Experiment time: 2025-06-17, 18:00 - 2025-06-19, 21:00

Integrative Analysis of Transcriptomics and Proteomics

I. Objective of the experiment

1.1 Through the combined analysis of transcriptome and proteome, the correlation and differences between gene expression and protein levels were systematically investigated in a drug-resistant cell model of human bladder transitional cell carcinoma (T24-RC48).

1.2 Select genes/proteins that are significantly changed at both transcription and protein levels, explore their biological functions and signaling pathways, and reveal potential regulatory mechanisms related to drug resistance.

1.3 Verify whether the changes in transcript level directly determine the changes in protein abundance, and explore the coordinated multi-omics regulatory network underlying the drug-resistant phenotype.

II. Experimental procedures

2.1 Sample information and grouping

Transcriptome samples: T24_R_high (highly resistant), T24_R_mode (moderately resistant), T24_R_low (low-resistant), and T24_R_control (wild-type)

Protein group samples: T24_R_high (highly resistant), T24_R_mode (moderately resistant), T24_R_low (low-resistant), and T24_R_control (wild-type)

2.2 Criteria for Differential Analysis

Transcriptome: $|\log2FC| \geq 1.0$, $pAdj < 0.05$

Protein group: $|\log2FC| \geq 0.263$, $p\_value < 0.05$

2.3 Joint analysis process

Overall results comparison: Venn diagram analysis, gene expression distribution, correlation analysis (linear fit R^2 value to assess transcription and protein level consistency).

Comparison of differential results: statistical analysis of differential genes/proteins, Venn analysis, and clustered heatmaps.

2.4 Data mining analysis

GO functional enrichment: Enriched GO terms (Biological Process, Molecular Function, Cellular Component) were identified and visualized using Venn diagrams, heatmaps, bubble

plots, and bar charts.

KEGG pathway enrichment: a total of pathway screening combined with pathway and network maps to analyze key signaling pathways.

Screening of key genes/proteins: Identify core drug-resistant related molecules based on co-expression trends (such as UP-UP, DOWN-DOWN, etc.) and PPI network analysis.

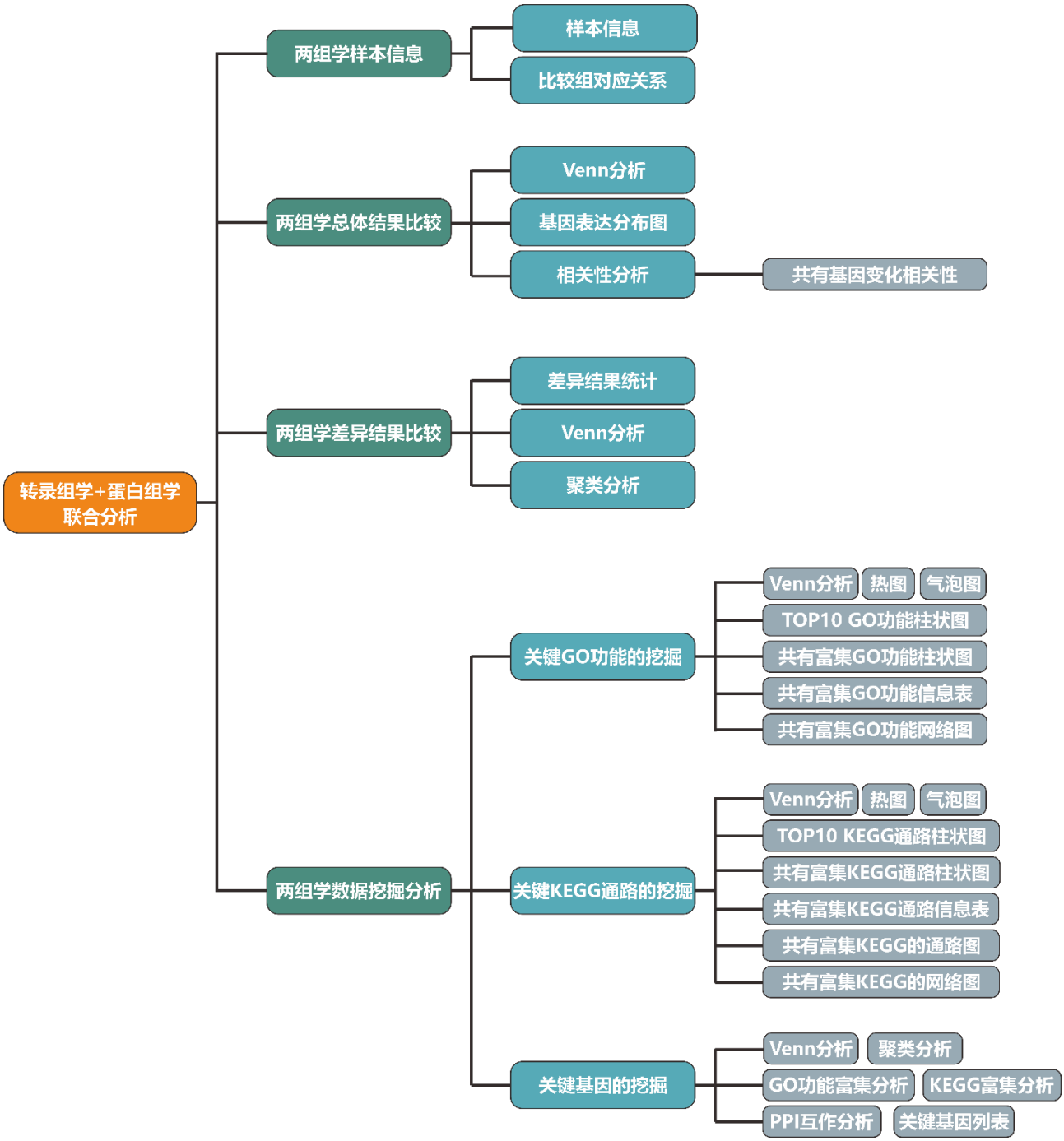

Figure 1 Joint analysis process and method

III. Experimental results

3.1 Comparison of the overall results of the two groups

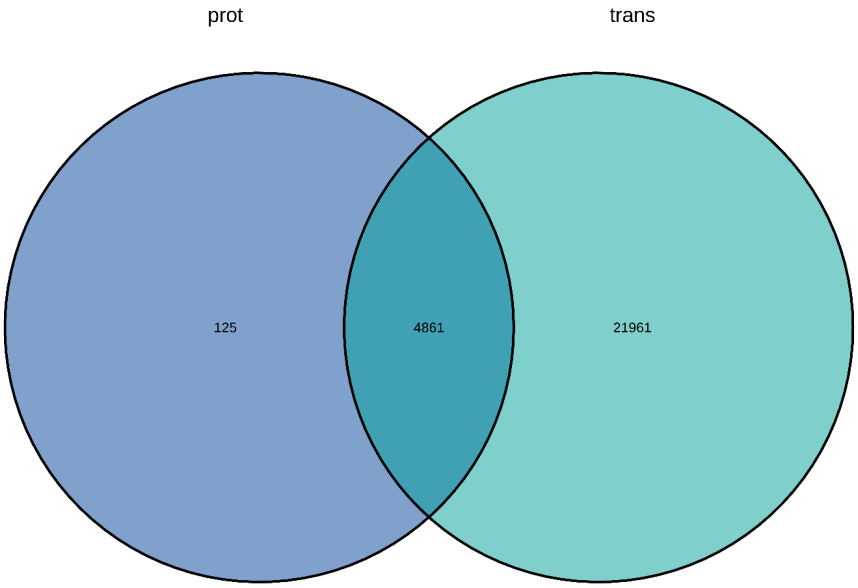

Figure 2 Venn diagram showing the overlap of genes detected in the transcriptomics and proteomics datasets

The gene expression distribution map (Figure 3) shows that the transcriptome data covered a wider range of gene expression levels (blue distribution), whereas the proteomics data were concentrated in the medium-to-low expression range (gray distribution).

Experiment time: 2025-06-17, 18:00 - 2025-06-19, 21:00

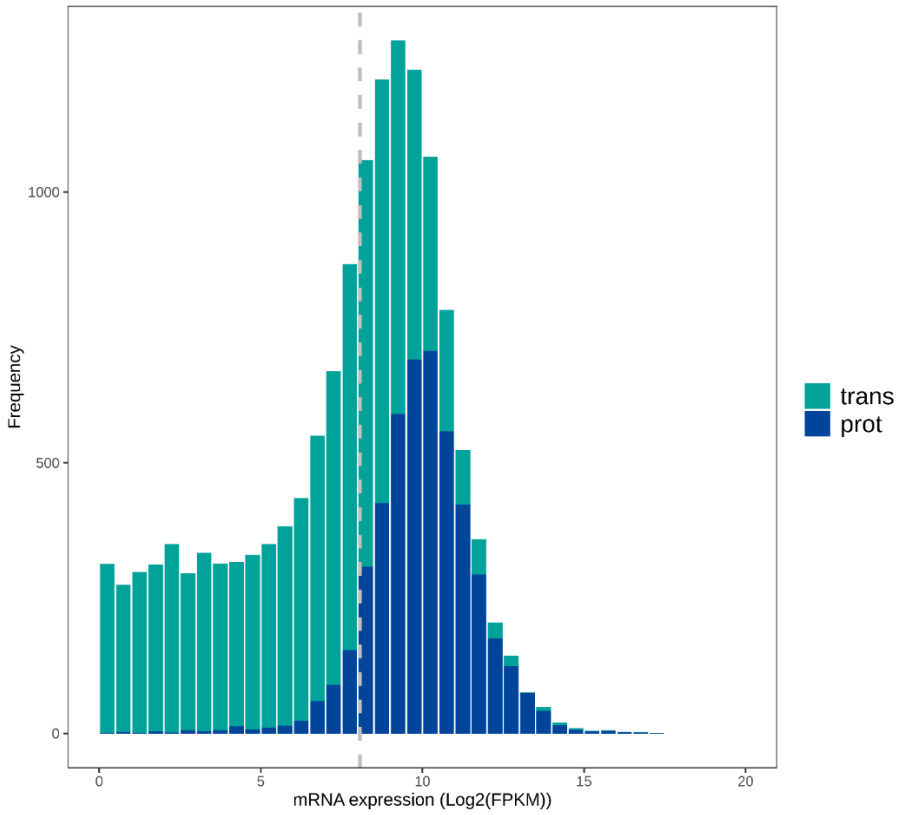

Figure 3 Gene expression distribution

Correlation analysis involves examining two or more interrelated variables to measure their degree of association. Given that "how consistent gene expression is at the transcriptional and protein levels" remains a key focus for researchers, we employ linear fitting (linear equation y=kx + b) to analyze shared genes between transcriptomics and proteomics. This approach helps us assess overall correlations in both transcriptional and protein expression levels between experimental and control groups under identical physiological conditions. The correlation coefficient is represented by R, where R^2 serves as the determination coefficient that evaluates how well experimental data aligns with the fitted function. A higher R^2 value (closer to 1) indicates stronger alignment, suggesting stronger correlations between shared genes at transcriptional and protein expression levels; conversely, a lower R^2 value (closer to 0) suggests weaker correlations.

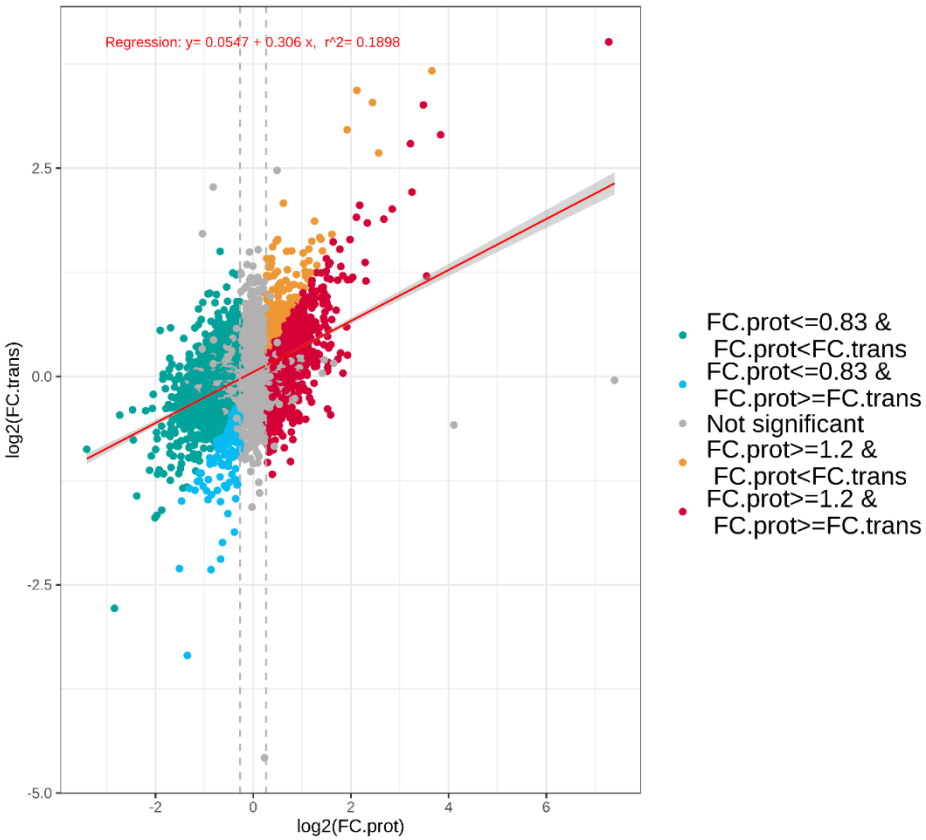

Figure 4 Gene variation correlation diagram

3.2 Comparison of Differential Expression Results Between the Two Omics

We conducted comparative analysis between intergroup differential gene expression results at the transcriptome level and those at the proteome level. The screening criteria for transcriptomic differential genes in this study were: |log2FC| ≥1.0 with p_adj <0.05; while for proteomic differential proteins, the criteria were: |log2FC| ≥0.263 with p-value <0.05. The statistical results of the two-omics differences are summarized in the table below:

Table 1 Summary of differentially expressed genes (DEGs) and proteins (DEPs)

Comparison	transcriptomics			proteomics		
	UP	DOWN	TOTAL	UP	DOWN	TOTAL
uniteT24_high_vs_uniteT24_con	1118	941	2059	1229	1327	2556
uniteT24_high_vs_uniteT24_low	413	355	768	1260	1266	2526
uniteT24_high_vs_uniteT24_mode	360	212	572	1094	1081	2175

uniteT24_mode_vs_uniteT24_low	359	484	843	723	820	1543
uniteT24_mode_vs_uniteT24_con	1089	1129	2218	894	1049	1943
uniteT24_low_vs_uniteT24_con	569	379	948	424	519	943

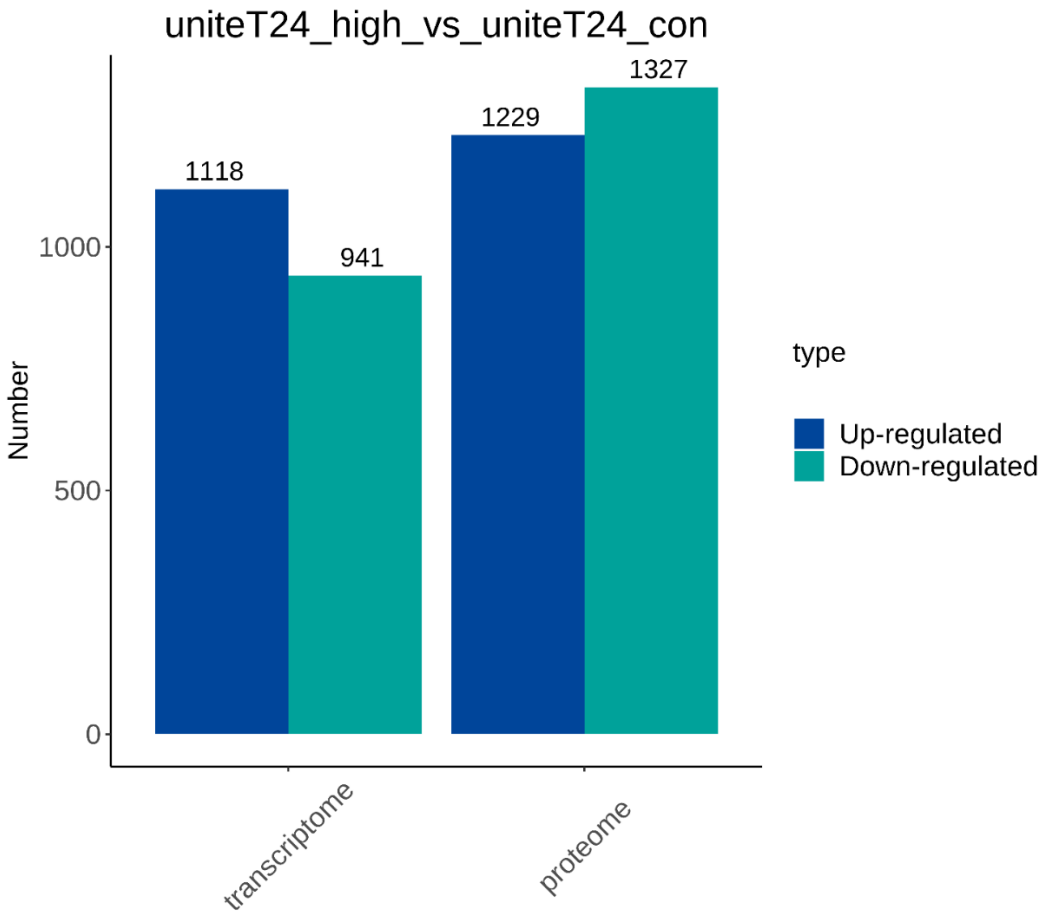

Figure 5 Bar chart of difference results

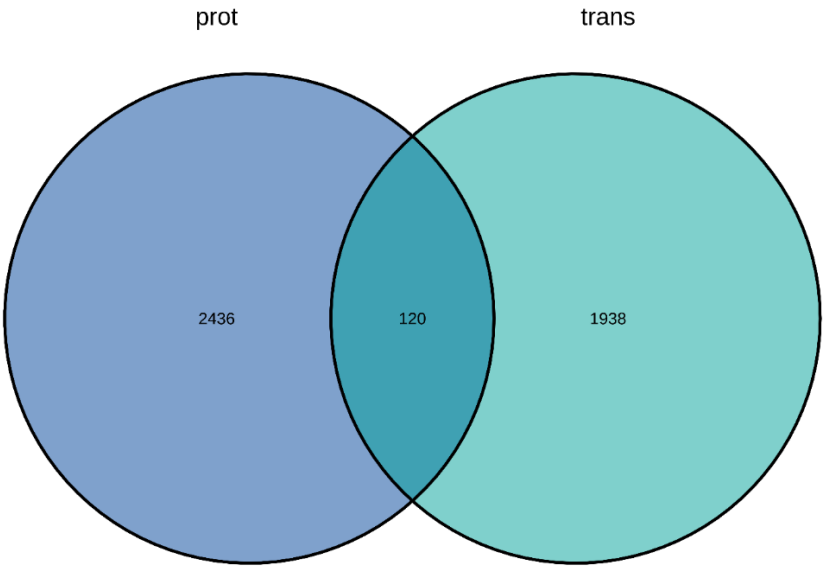

Figure 6 Venn analysis of differential results

The clustered heat map analysis of the FC values (log2FC) of the common differential genes obtained in the above Venn analysis in the two omics is shown as follows:

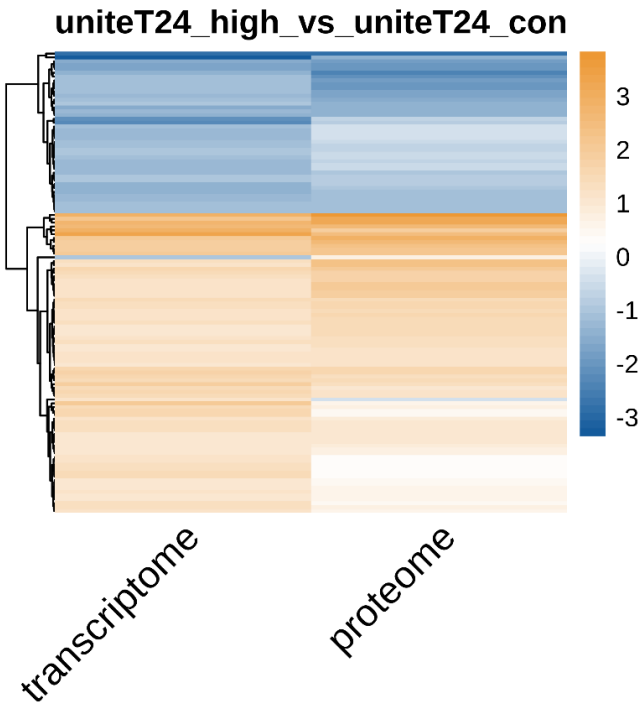

Figure 7 Cluster analysis heat map

3.3 Integrated Mining Analysis of the Two Omics Datasets

3.3.1 Key GO functional enrichment

Gene ontology (GO) is a systematic approach for annotating gene and protein attributes across species. Its objectives include: 1) maintaining and expanding a standardized vocabulary for gene and protein descriptions; 2) annotating genes and proteins while assimilating and disseminating annotation data; 3) providing user-friendly tools for data access; 4) enabling GO-based programming analysis using experimental data, such as gene enrichment analysis. The framework primarily consists of three branches: cellular components, molecular functions, and biological processes. This study compares GO functional enrichment results from differential mRNA and protein analyses in two omics datasets to identify key GO functions and gain a comprehensive understanding of the biological mechanisms underlying regulatory interactions between differentially expressed molecules.

The Venn analysis of the significantly enriched GO functions (transcriptomics: p-value <0.05; proteomics: p-value <0.05) for the two groups of differential molecules is shown in the figure below:

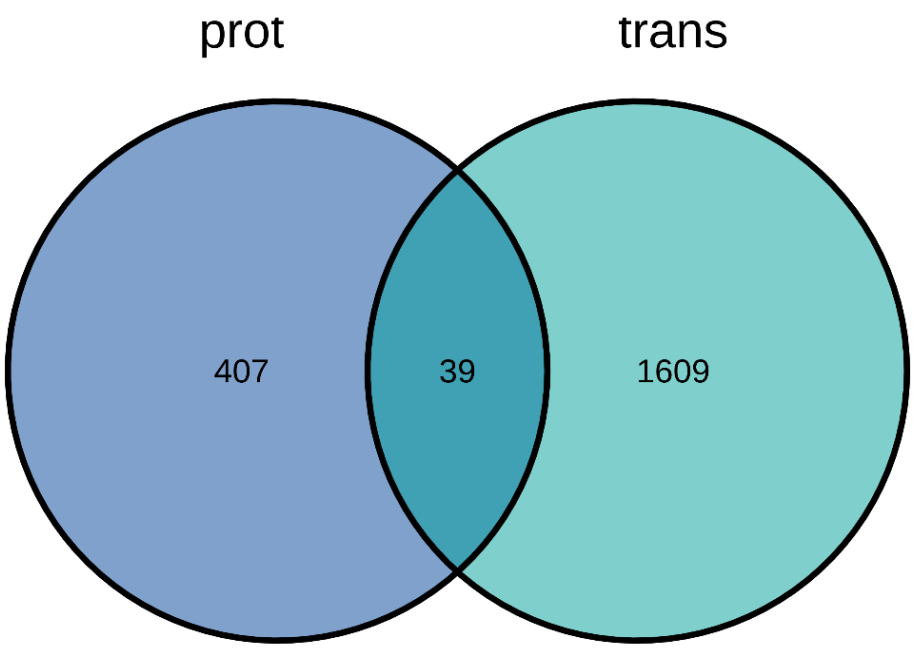

Figure 8 GO functional enrichment Venn diagram

The p-value values of the two sets of GO functional enrichment results (p-value <0.05) were used for cluster heat map analysis. The results are shown as follows:

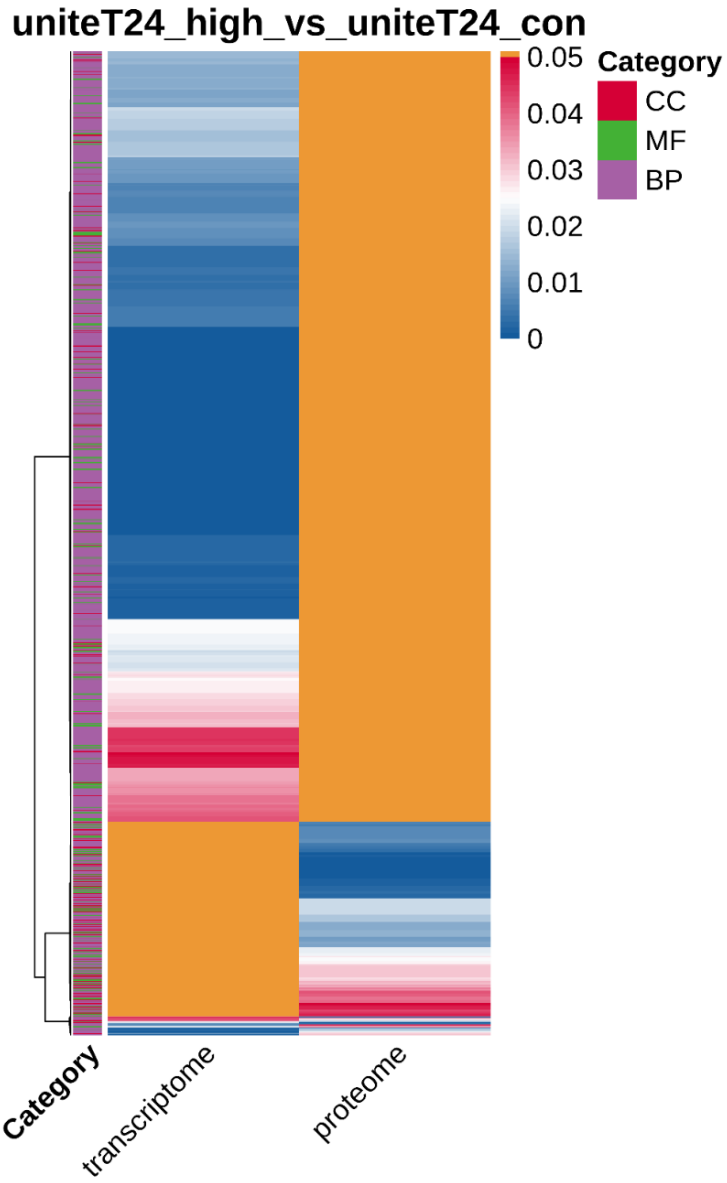

Figure 9 GO functional heat map

The bubble diagram not only visually shows the GO functions that are significantly enriched in each omics, but also visually reflects the commonness and specificity of the GO functions enriched in the two omics.

Based on the GO functions enriched by two omics differences, we ranked the p values in order

from small to large, and selected the top 10 pathways with the smallest p values of each omics group, that is, a total of up to 20 pathways were presented in the form of bubble diagram as follows.

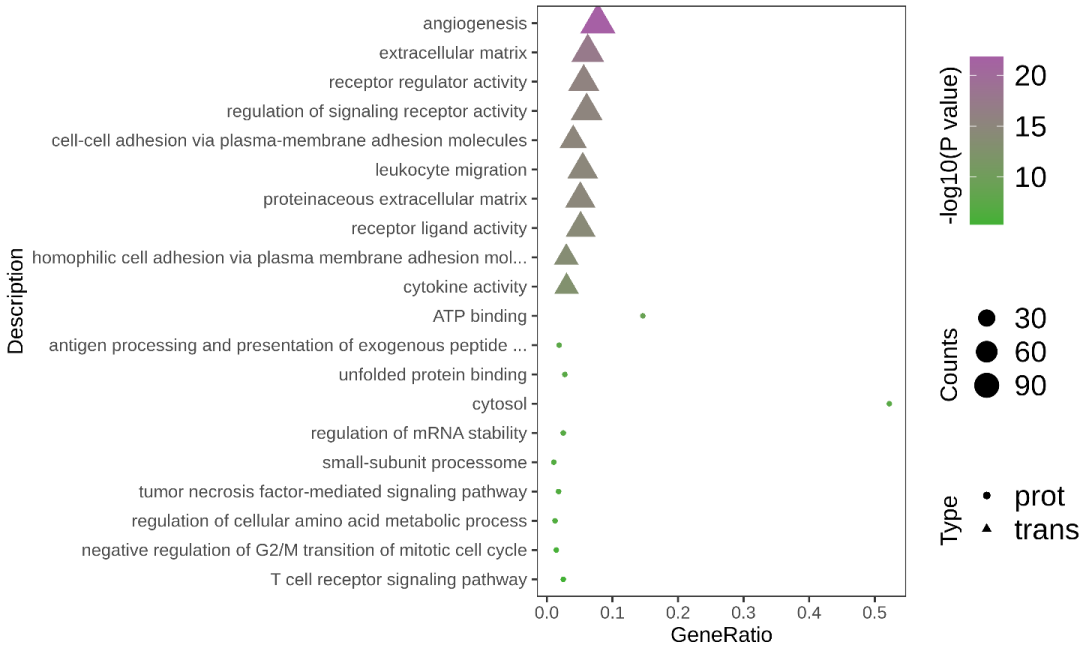

Figure 10 GO functional bubble diagram

Which GO functions are enriched with the highest number of differential molecules (mRNA or protein)? We conducted a summation analysis of differential mRNA counts and protein counts enriched by each GO function. The top 10 GO functions with the highest total differential molecule counts (differential mRNA count + differential protein count) were prioritized for presentation, helping researchers identify which GO functions are most enriched in the project's differential molecules. When fewer than 10 GO functions are present, we display the number of GO functions and their respective counts.

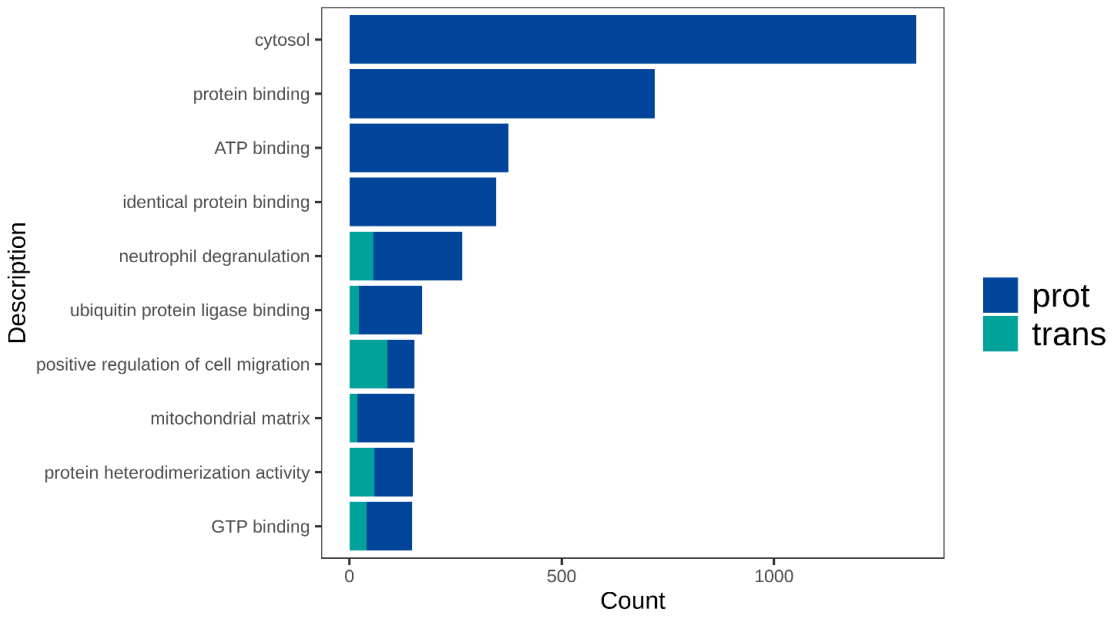

Figure 11 TOP10 GO function bar chart

The shared GO functions across the two omics are also of interest to researchers. We present these shared GO functions through a bar chart. By default, the TOP10 shared GO functions are prioritized (ranked by the ascending order of enrichment significance p-values for shared pathways across both omics, with the total ranking sum of both omics combined, and the top 10 with the smallest total sum selected as the TOP10 shared enriched GO functions based on p-value ranking). The bars display the cumulative sum of differential mRNA and protein counts.

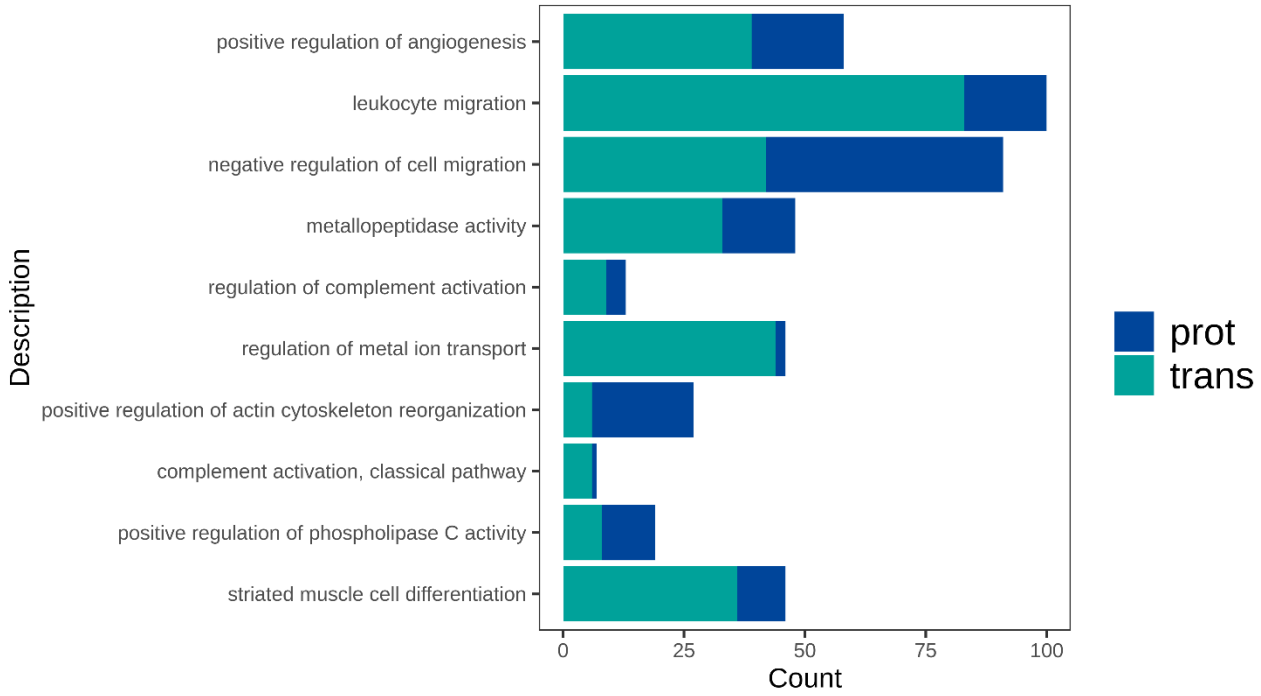

Figure 12 Bar chart of shared enriched GO functions

3.3.2 Key KEGG pathway enrichment

The Kyoto Encyclopedia of Genes and Genomes,http://www.kegg.jp/ (KEGG) database integrates genomic information with metabolite functional data, systematically mapping pathways across metabolic processes, genetic regulation, environmental interactions, cellular mechanisms, biological systems, human diseases, and pharmaceutical development. By conducting pathway integration analysis on differentially expressed molecules identified through transcriptomic and proteomic studies, researchers can investigate biological mechanisms at the level of information flow. This cross-verification approach enables identification of critical signaling pathways from massive datasets, ultimately facilitating the construction of comprehensive regulatory networks governing organism functions.

The Venn analysis was performed on the significantly enriched KEGG pathways of the two groups of differential molecules (transcriptomics: p-value <0.05; proteomics: p-value <0.05).

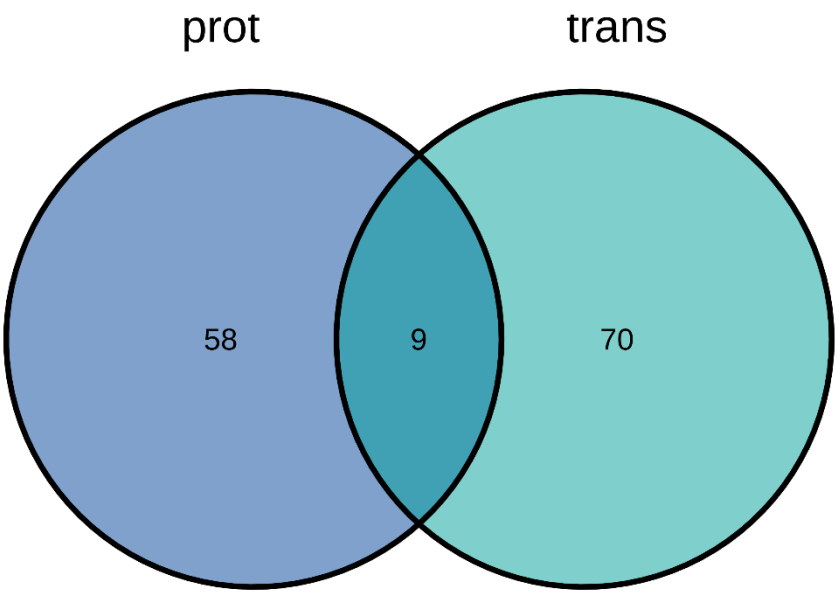

Figure 13 KEGG pathway enrichment Venn diagram

The p-value values of the two sets of KEGG pathway enrichment results (p-value <0.05) were used for cluster heat map analysis. The results are shown as follows:

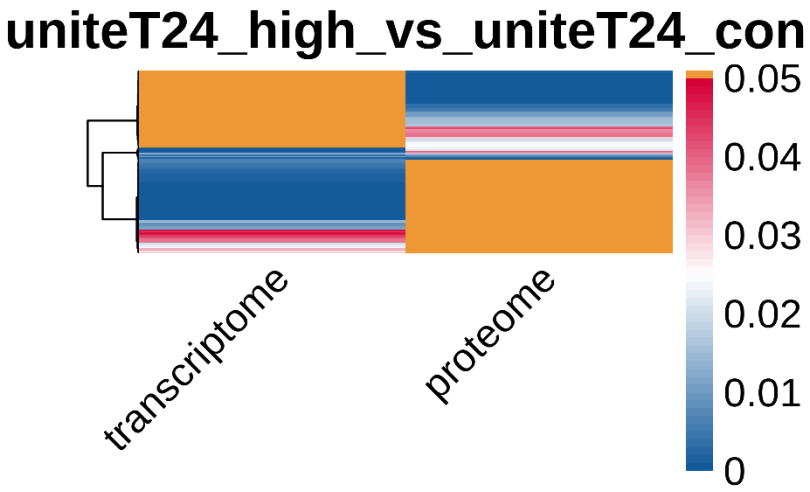

Figure 14 KEGG pathway heat map

The bubble plot not only visually shows the KEGG pathways that are significantly enriched in each omics, but also visually reflects the commonness and specificity of the omics KEGG pathway enrichment results.

Based on the KEGG pathways enriched by differential molecular groups, we ranked the p-values in descending order of significance and selected the top 10 pathways with the smallest p-values in each group, that is, a total of up to 20 pathways were presented in the form of bubble diagram as follows.

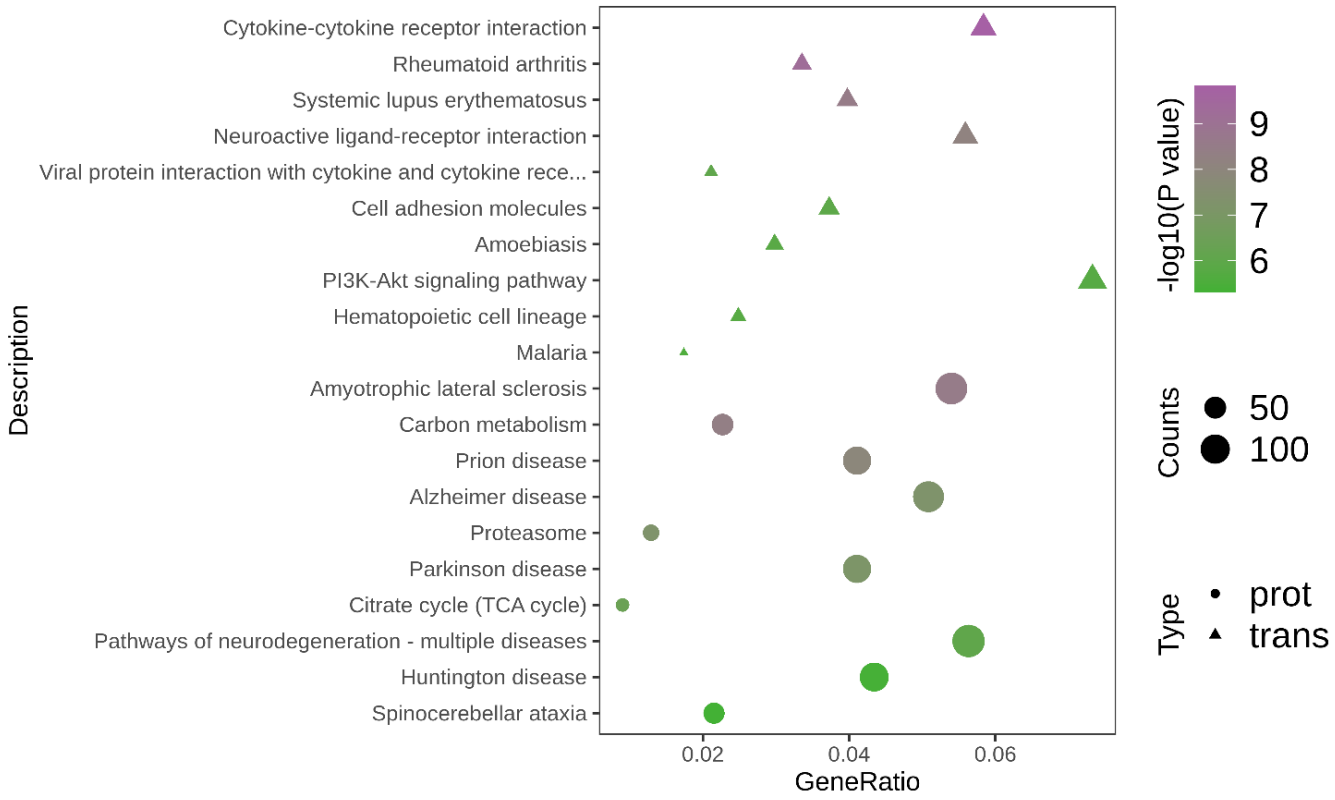

Figure 15 KEGG pathway bubble diagram

Which KEGG pathways exhibit the highest abundance of differential molecules (mRNA or protein)? We conducted a summative analysis of differential mRNA counts and protein counts across all KEGG pathways. The top 10 pathways with the highest total differential molecule counts (differential mRNA count + differential protein count) were prioritized for presentation, helping researchers identify key KEGG pathways associated with the project's differential molecular annotations. For pathways containing fewer than 10 molecules, we displayed the number of

molecules per pathway. The results are shown in the figure below:

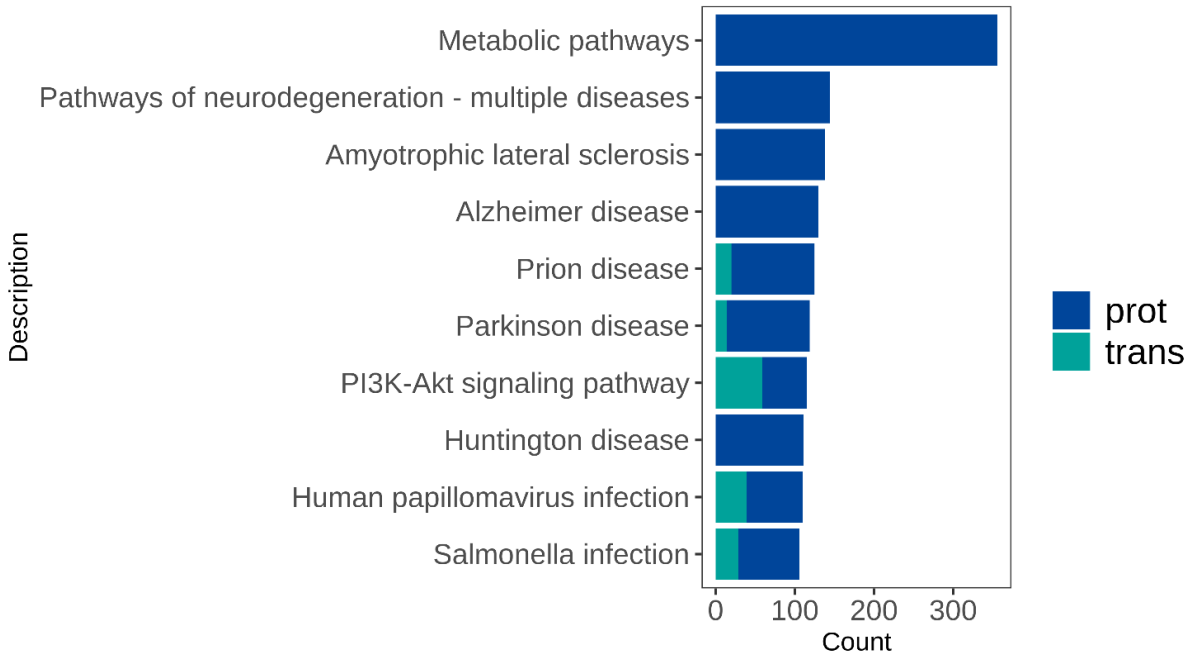

Figure 16 Top 10 KEGG pathway bar chart

The shared KEGG pathways across the two omics are also of significant interest to researchers. We present these pathways through a bar chart for visual clarity. By default, the TOP10 pathways are prioritized in the shared KEGG pathways (ranked by p-value significance in ascending order across both omics, with cumulative ranking scores calculated). The top 10 pathways are selected based on the minimum cumulative sum (i.e., pathways ranked by p-value rank that exhibit shared enrichment). The bar chart displays the cumulative sum of differential mRNA and protein counts across the shared KEGG pathways.

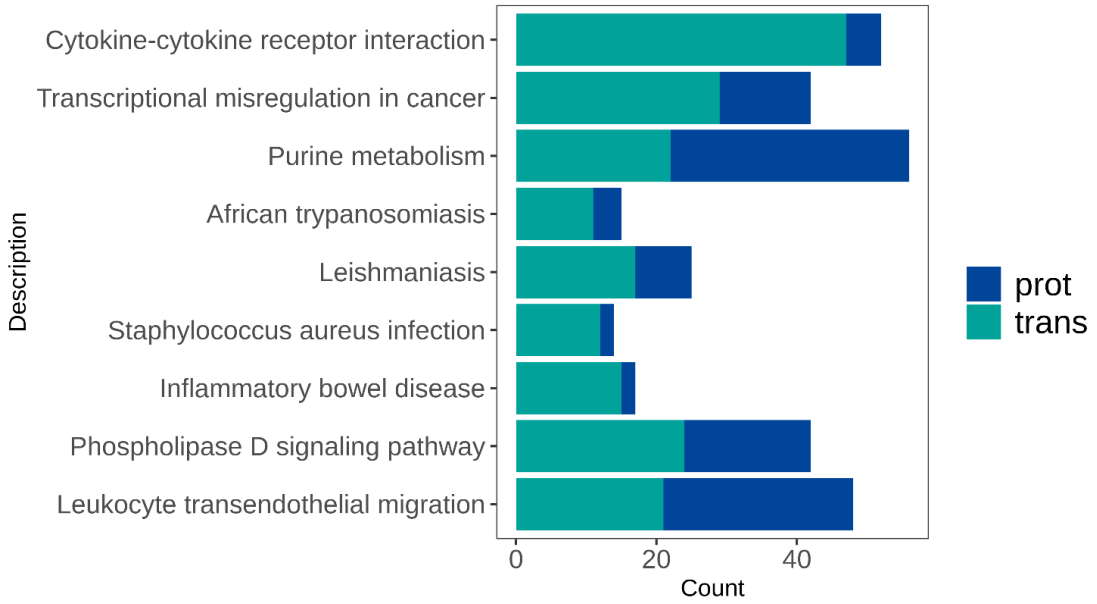

Figure 17 Bar chart of shared enriched KEGG pathways

3.3.3 Key gene analysis

Genes showing shared trends in transcriptomic and proteomic analyses typically regulate the same signaling pathways, with those common pathways across omics datasets playing pivotal roles. This project identifies key mRNA or protein targets through comprehensive analysis of critical genes, aiming to uncover deeper biological mechanisms.

The difference molecules of the two groups were analyzed by Venn to identify the difference molecules with the same or opposite trend of change.

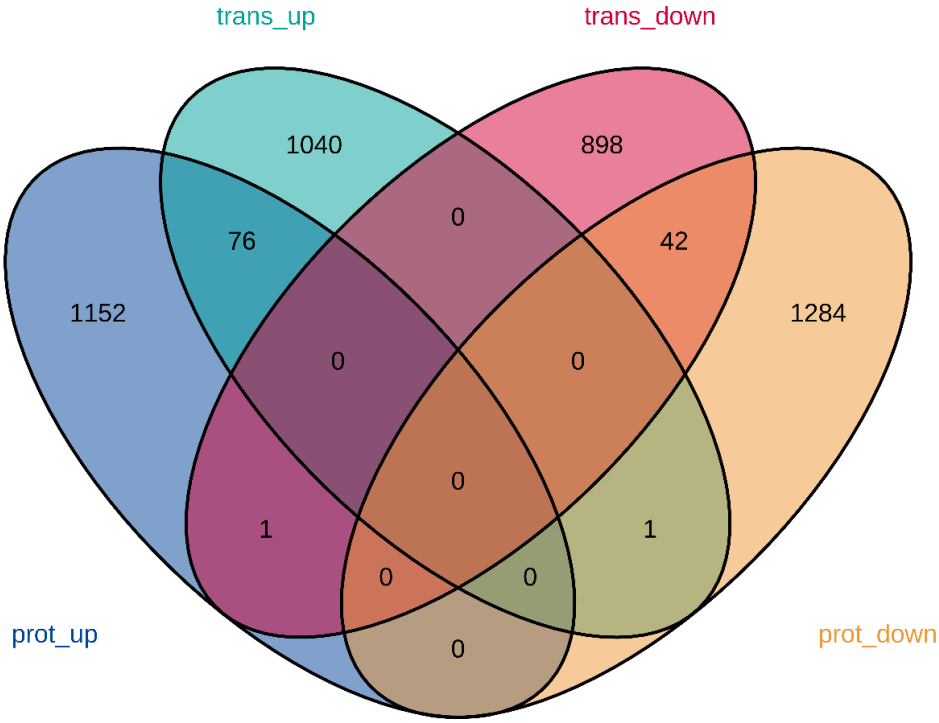

Figure 18 Venn diagram of differential gene variation trend

Meanwhile, we present a four-quadrant diagram to illustrate the shared differential molecules in the two omics. Genes upregulated in both transcriptomics and proteomics are labeled as UP-UP; those upregulated in transcriptomics but downregulated in proteomics are labeled as UP-DOWN; genes downregulated in transcriptomics but upregulated in proteomics are labeled as DOWN-UP; and genes downregulated in both omics are labeled as DOWN-DOWN. The four-quadrant diagram is illustrated below:

Experiment time: 2025-06-17, 18:00 - 2025-06-19, 21:00

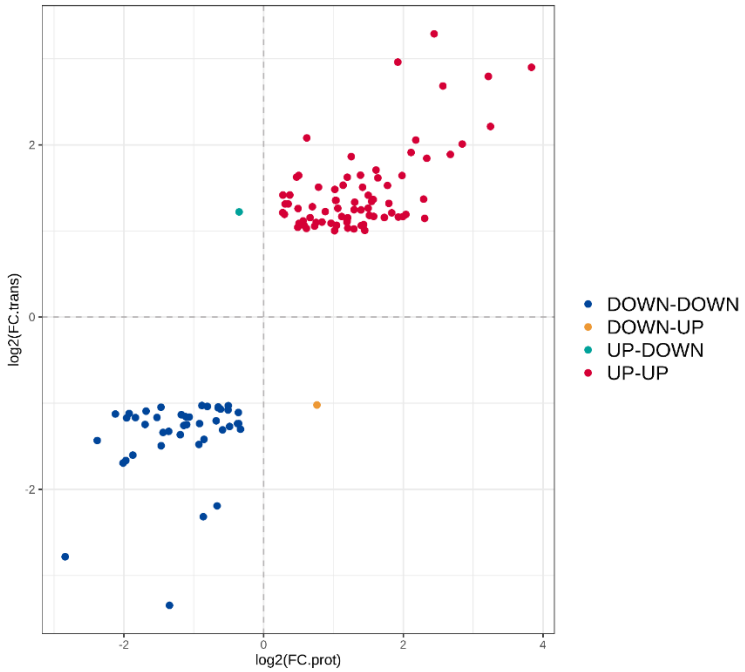

Figure 19 Four-quadrant scatter plot of common differentially expressed molecules

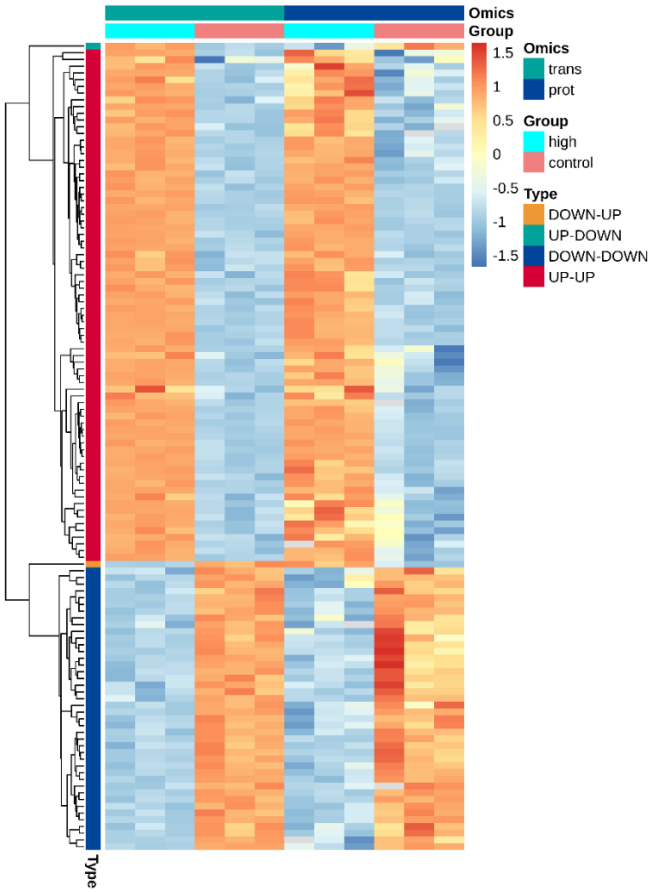

Figure 20 Cluster analysis heat map

Experiment time: 2025-06-17, 18:00 - 2025-06-19, 21:00

GO functional enrichment analysis was performed on the above four types of genes: Up-Up, Down-Down, Up-Down and Down-Up to speculate their possible biological functions. The results were displayed in bar chart and bubble chart as follows:

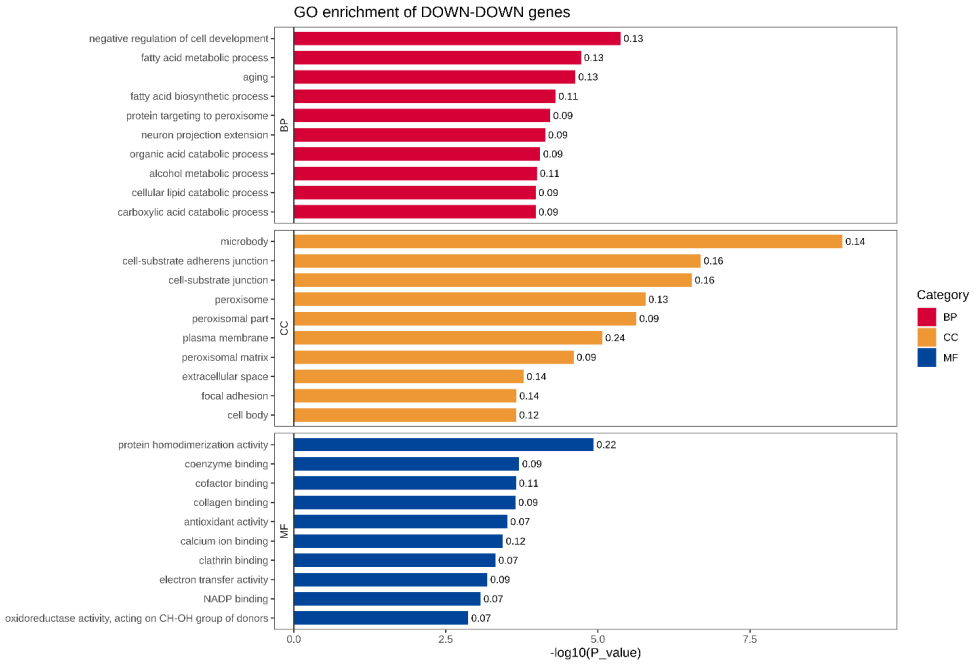

Figure 21 GO functional enrichment bar chart

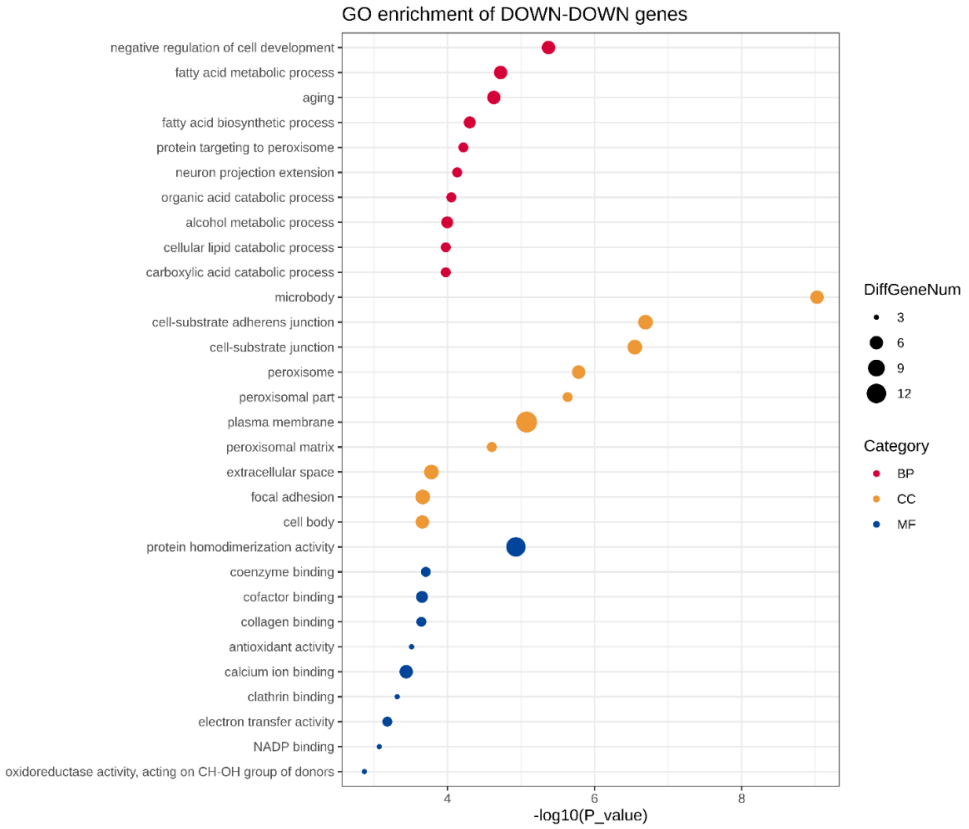

Figure 22 GO functional enrichment bubble diagram

KEGG pathway enrichment analysis was performed on the above four types of genes: Up-Up, Down-Down, Up-Down and Down-Up to speculate the possible biological pathways involved. The results were displayed in bar chart and bubble chart as follows:

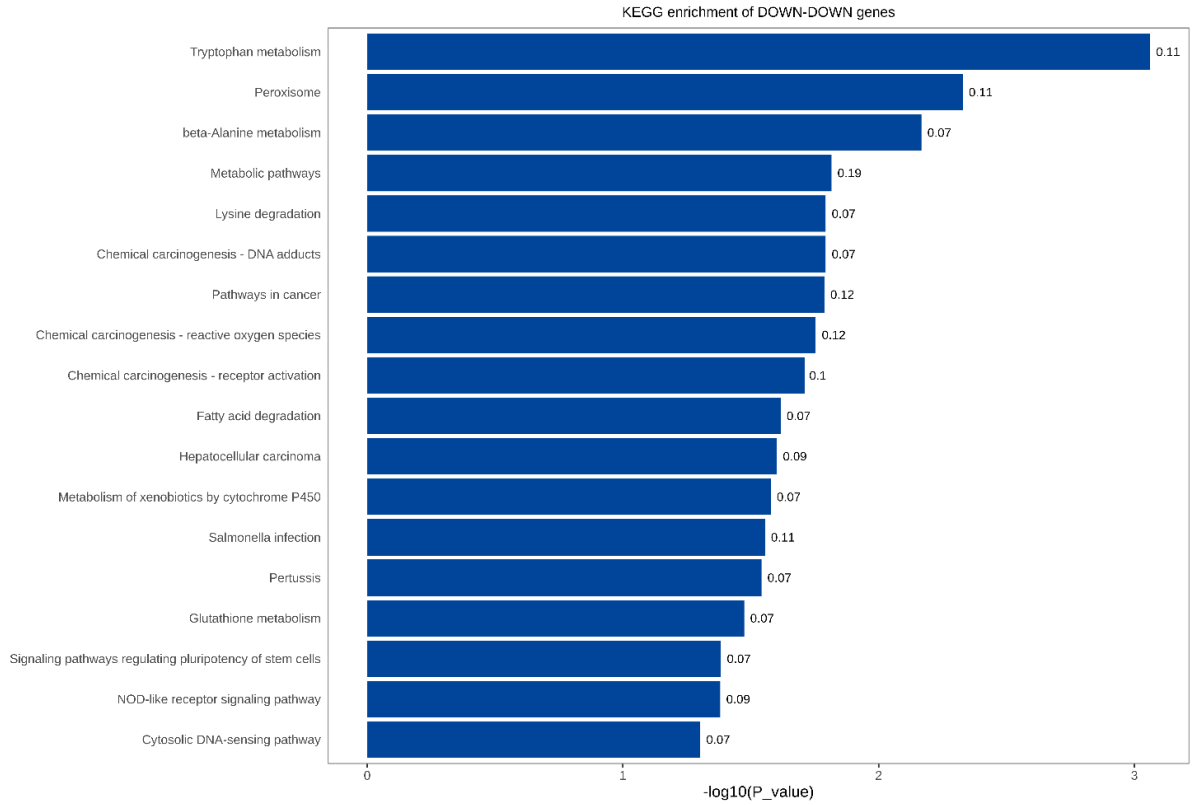

Figure 23 KEGG pathway enrichment bar chart

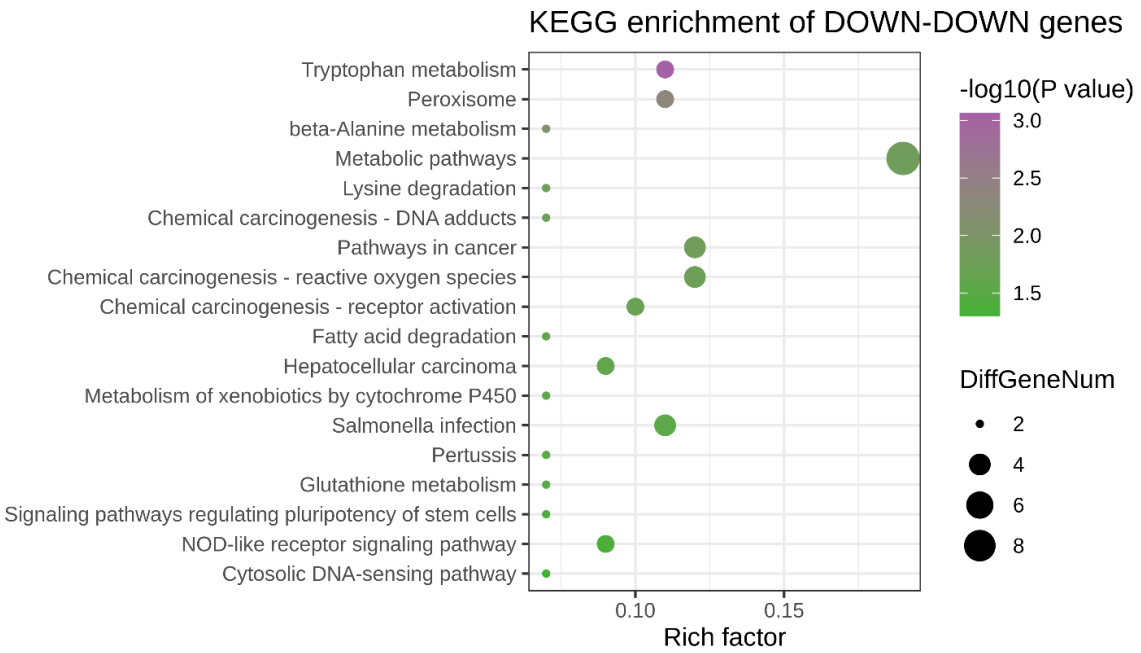

Figure 24 KEGG pathway enrichment bubble diagram

IV. Results analysis

This study systematically elucidated the molecular regulatory features of the human bladder transitional cell carcinoma T24-RC48 drug-resistant model through integrated transcriptomic and proteomic data. The results revealed significant differences in gene expression across the two omics, with non-synchronous trends observed between transcriptional and protein-level changes. This suggests that the drug resistance mechanism may involve complex processes such as post-transcriptional regulation and protein translation modifications. Functional enrichment analysis further uncovered synergistic interactions between drug-resistant-related molecules and critical biological processes (e.g., cell migration regulation, metabolic reprogramming) and signaling pathways (e.g., PI3K-Akt, cytokine interactions), indicating that the drug-resistant phenotype results from dynamic regulation of multidimensional molecular events. These findings provide crucial insights for in-depth analysis of RC48's drug resistance mechanism, while the integrated omics approach establishes a data foundation for subsequent studies on drug resistance mechanisms, demonstrating both scientific value and potential clinical applications.