Experiment time: 2025-06-12, 18:00 - 2025-06-15, 21:00

Proteomics Data Analysis

I. Objective of the experiment

Through proteomic technology, we conducted comprehensive protein expression profiling analysis on the RC48 drug-resistant human bladder transitional cell carcinoma T24-RC48 cell line (low/medium/high drug resistance group) and the T24 cell line (control group). This approach identified significantly differentially expressed proteins, revealed drug resistance-related functional pathways and key protein biomarkers, thereby providing molecular-level data support for elucidating the mechanisms of acquired drug resistance in RC48.

II. Experimental procedure

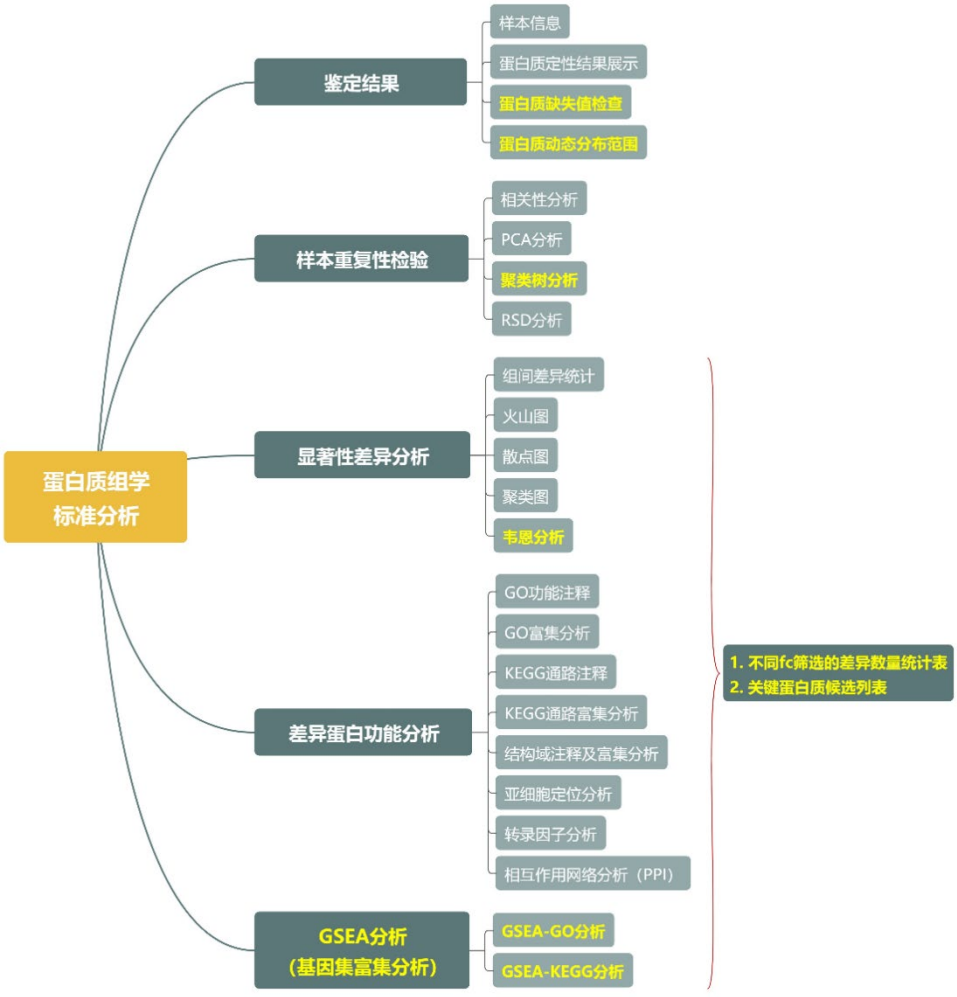

Figure 1 Proteomics information analysis technology flow

2.1 Sample preparation

1) Extract total protein by cell lysis, quantify by BCA method, and verify protein integrity by SDS-PAGE.

2) Quality control after enzymatic digestion (peptide amount ≥200ng), and data were acquired using a TimsTOF Pro mass spectrometer in Data-Dependent Acquisition (DDA) mode.

2.2 Mass spectrometry data acquisition and analysis

1) The original mass spectrometry data were searched in the database (Homo sapiens UniProt database) by MaxQuant software.

2) Label-free quantification, and screening more than 50% of non-empty data in the group for differential analysis.

3) Significance difference criteria: fold change>1.2 (up or down) and P value <0.05.

2.3 Bioinformatics analysis

1) Differential protein screening: volcano plots, scatter plots, and clustered heatmaps.

2) Functional annotation and enrichment analysis: GO, KEGG, domain, subcellular localization.

3) GSEA analysis: Based on gene set enrichment to evaluate drug-resistant functional modules.

4) Key protein screening: integrate multiple criteria including fold change, P-value, pathway enrichment results, and protein-protein interaction (PPI) network centrality.

III. Experimental results

3.1 Analysis of significant differences

In the analysis of significant differences in quantitative results, we first screened the data with at least half of the repeated experimental data in the sample group for differential comparison analysis. Proteins that meet the screening criteria of expression difference ratio greater than 1.2 times (up/downregulation) and P value (t test/significance A) less than 0.05 were regarded as differentially expressed proteins (DEPs).

Table 1 Summary of differentially expressed proteins (DEPs)

Comparison	Up-regulated proteins	Down-regulated proteins	Total DEPs
high vs control	1229	1327	2556

Experiment time: 2025-06-12, 18:00 - 2025-06-15, 21:00

Comparison	Up-regulated proteins	Down-regulated proteins	Total DEPs
low vs control	424	519	943
moderate vs control	894	1049	1943
high VS moderate	1094	1081	2175
high VS low	1260	1266	2526
moderate VS low	723	820	1543

The number of differentially expressed proteins (DEPs) for each comparison is displayed in the bar chart below.

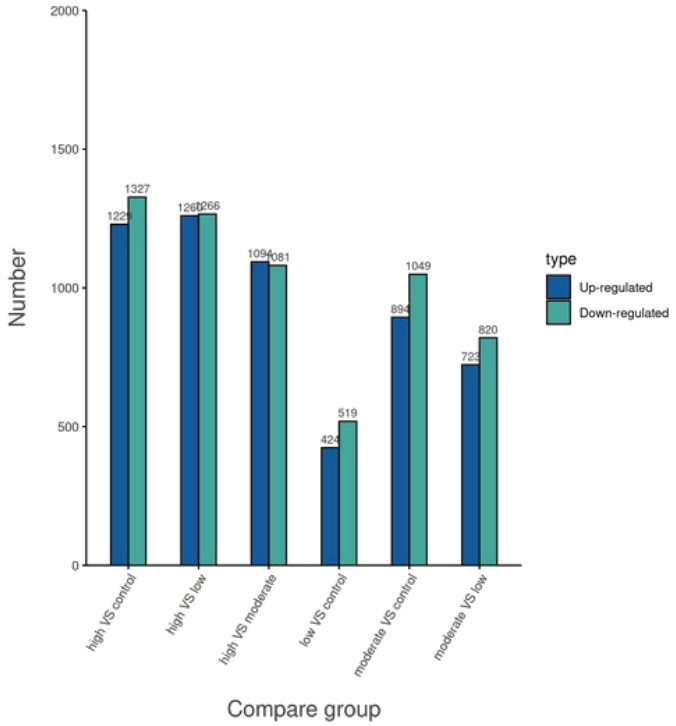

Figure 2 Statistical bar chart of differential proteins

The volcano plot provides a visual representation of differential expression distribution between comparison groups. We generated volcano plots without gene names and with only down/upregulated genes labeled (selecting the top 10 genes ranked by descending P values in both up/downregulation). The results are shown below. In the graph, blue dots represent upregulated

proteins, cyan dots indicate downregulated proteins, and gray-black dots denote no significant differences.

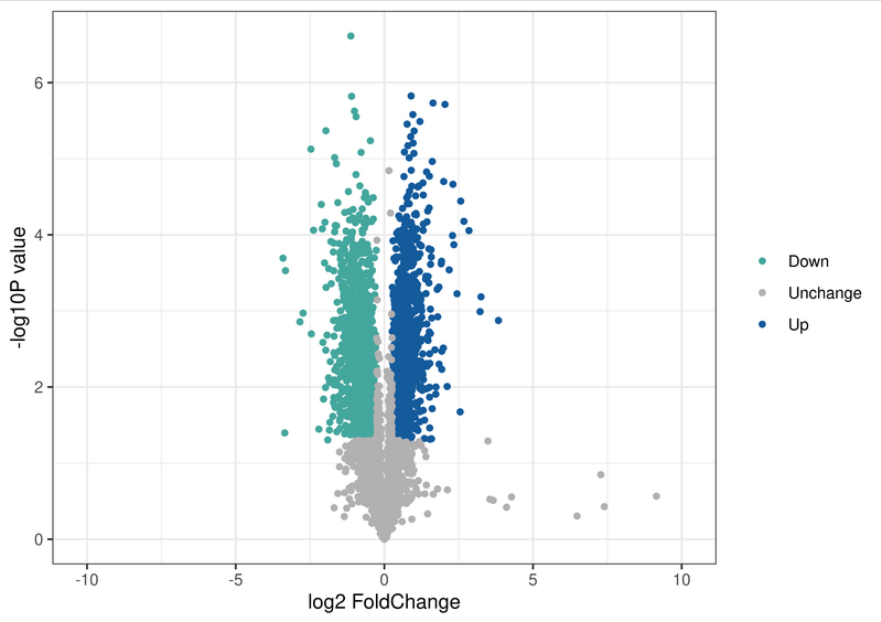

Figure 3 Volcano plot of differentially expressed proteins

Cluster analysis is a common exploratory data analysis method, whose purpose is to group and classify the data based on similarity. In the results of cluster grouping, the data patterns within groups are more similar, while the data patterns between groups are less similar.

During cluster analysis, the algorithm classifies samples and variables (in proteomics studies, variables typically refer to protein quantification data). The clustering results of samples help validate the rationality of selected target proteins, specifically determining whether changes in their expression levels can represent significant biological impacts from experimental treatments. These results enable identification of protein subgroups with distinct expression patterns within the dataset. Proteins sharing similar expression profiles may exhibit functional similarities, participate in overlapping biological pathways, or occupy adjacent regulatory positions within signaling networks. As illustrated in the figure below, the high-drug resistance group demonstrates significantly distinct protein expression patterns compared to the control group.

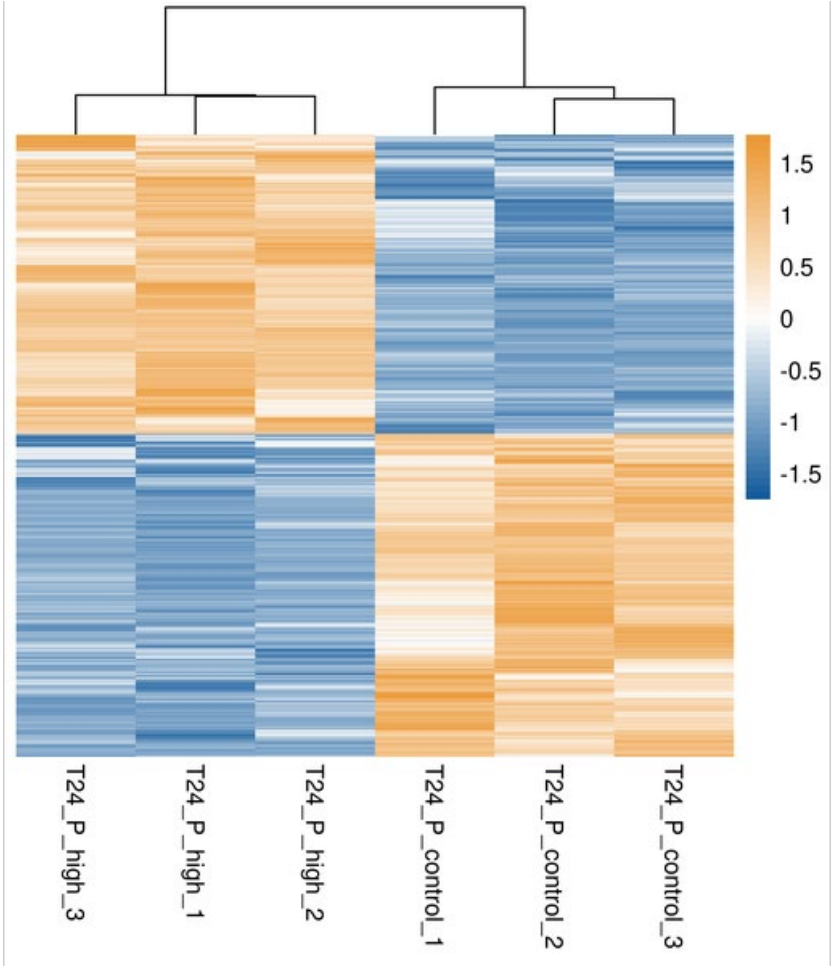

Figure 4 Clustered heatmap of differentially expressed proteins

3.2 Functional analysis of differential proteins

To elucidate the functional roles of differentially expressed proteins identified through proteomics screening, we employ Gene Ontology (GO) – a pivotal bioinformatics tool for characterizing gene and protein attributes. GO annotations are categorized into three domains: Biological Processes, Cellular Components, and Molecular Functions (Ashburner et al., 2000), which collectively explain protein biological functions from distinct perspectives. This approach enables precise identification of functional categories associated with these differentially expressed proteins.

We used Blast2Go(https://www.blast2go.com/) software (Gotz et al., 2008; Wang et al., 2014) to perform GO functional analysis on all differentially expressed proteins. Additionally, we conducted statistical analysis of the number of differential proteins at the GO secondary functional

annotation level, with the results shown in the figure.

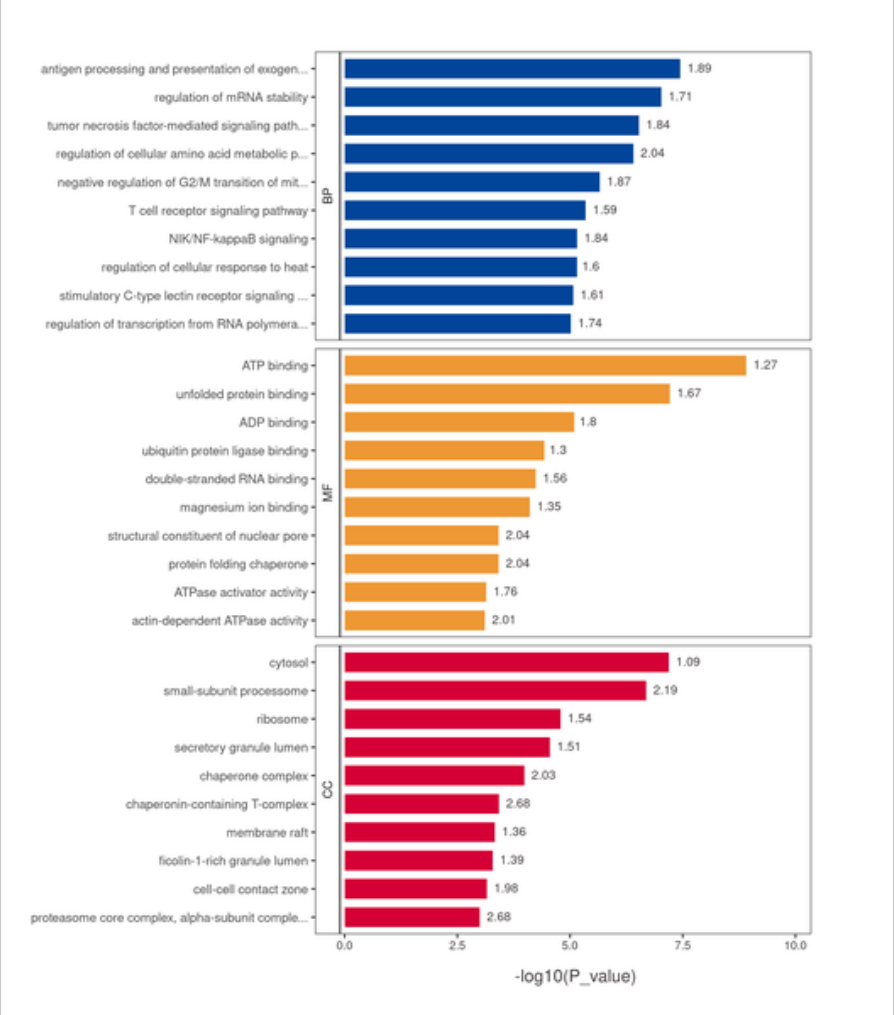

Figure 5 Level 2 statistical bar chart of GO annotation results

The top 10 enriched classification results in the three categories are shown in bubble chart:

Experiment time: 2025-06-12, 18:00 - 2025-06-15, 21:00

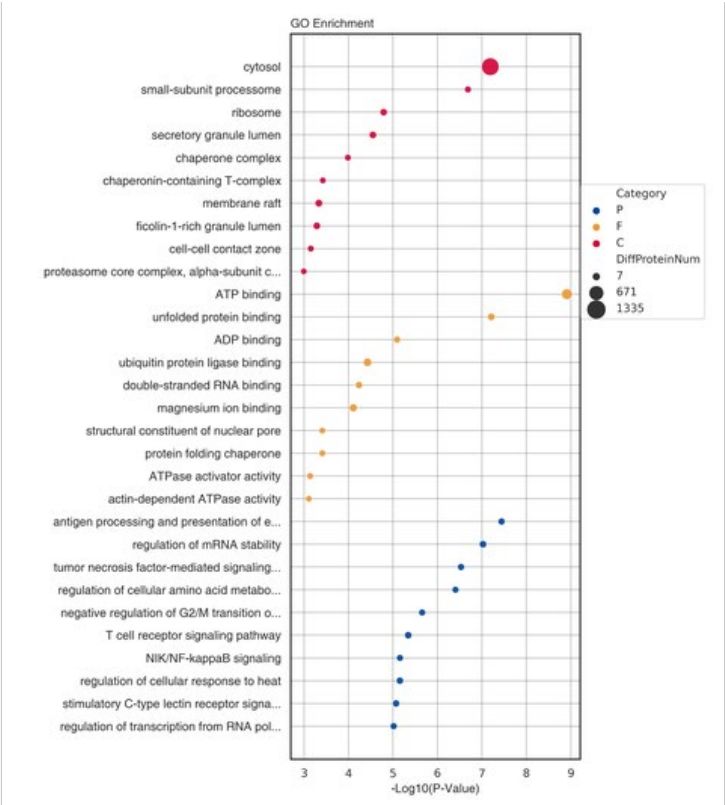

Figure 6 Enrichment statistics of GO terms (top 10)

According to Fold change, the differential proteins can be divided into two categories: up-regulated and down-regulated. In order to further understand the functions of the up-regulated and down-regulated differential proteins, we drew GO enrichment bar charts showing the up-regulated and down-regulated differential proteins, as shown in the figure.

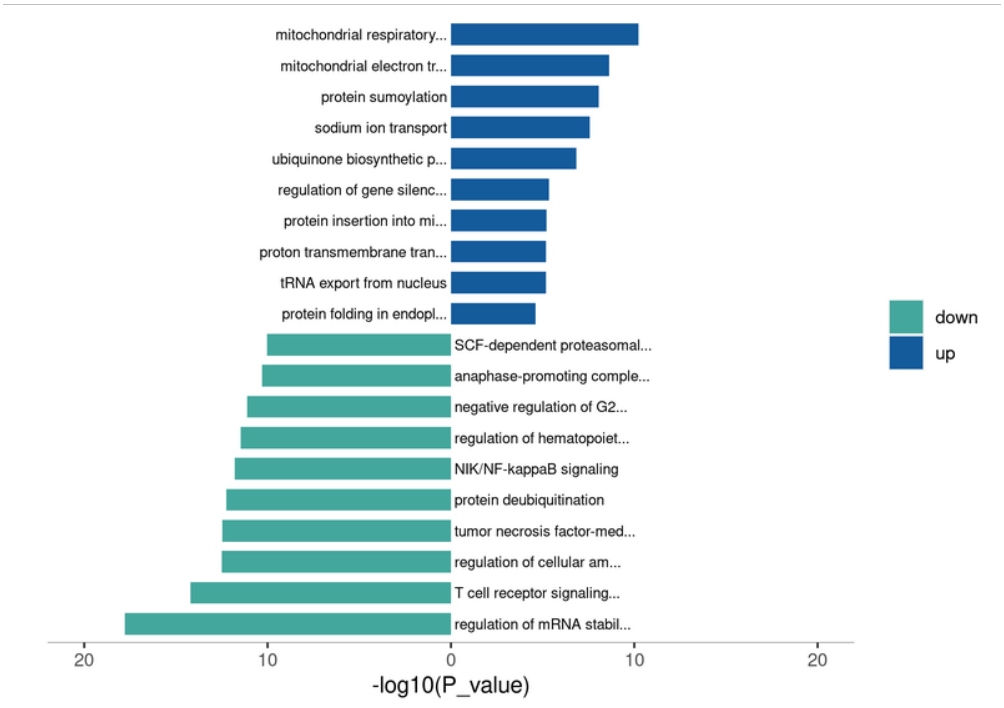

Figure 7 GO enrichment analysis (BP) for up and down regulated proteins

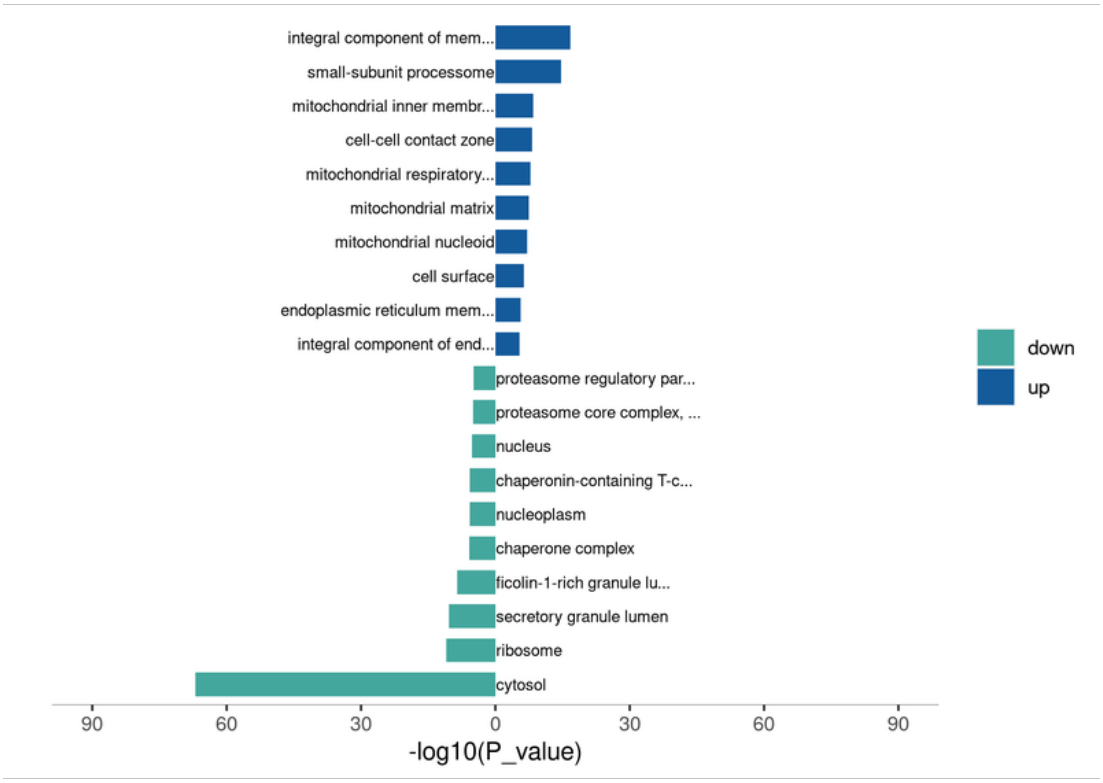

Figure 8 GO enrichment analysis (CC) for up and down regulated proteins

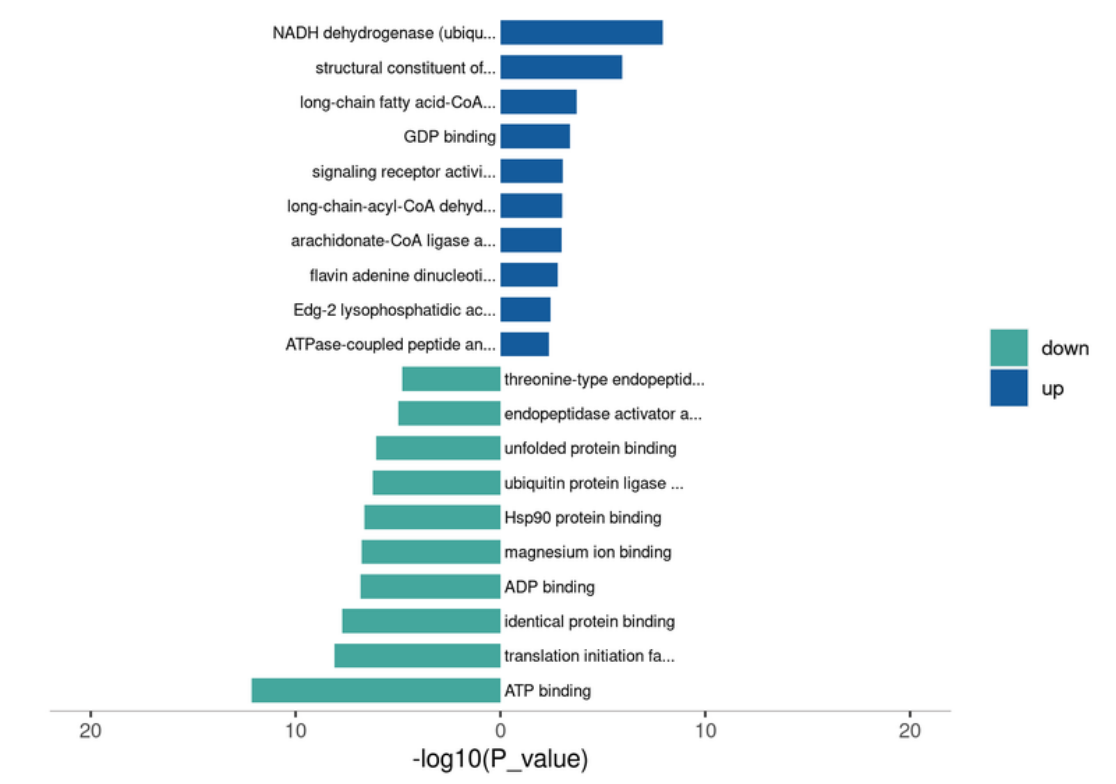

Figure 9 GO enrichment analysis (MF) for up and down regulated proteins

3.3 Pathway Analysis of Differentially Expressed Proteins

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a leading database for pathway research, developed by researchers through analyzing vast literature to visually represent metabolic pathways and their interactions (Kanehisa et al., 2012). The KEGG database contains pathway information covering metabolism, genetic processing, environmental regulation, cellular processes, biological systems, human diseases, and drug development. For more details, see: http://www.kegg.jp/. By annotating significantly differentially expressed proteins with KEGG pathways, researchers can identify potential metabolic or signaling pathways involved. This approach reveals protein changes from the cell surface to the nucleus, uncovers biological events and factors in these processes, and highlights potential biological consequences when specific pathways are disrupted or altered.

Based on the KEGG annotation results, we ranked the TOP20 KEGG pathway annotation results based on the number of differential proteins in the pathway, as shown in FIG.

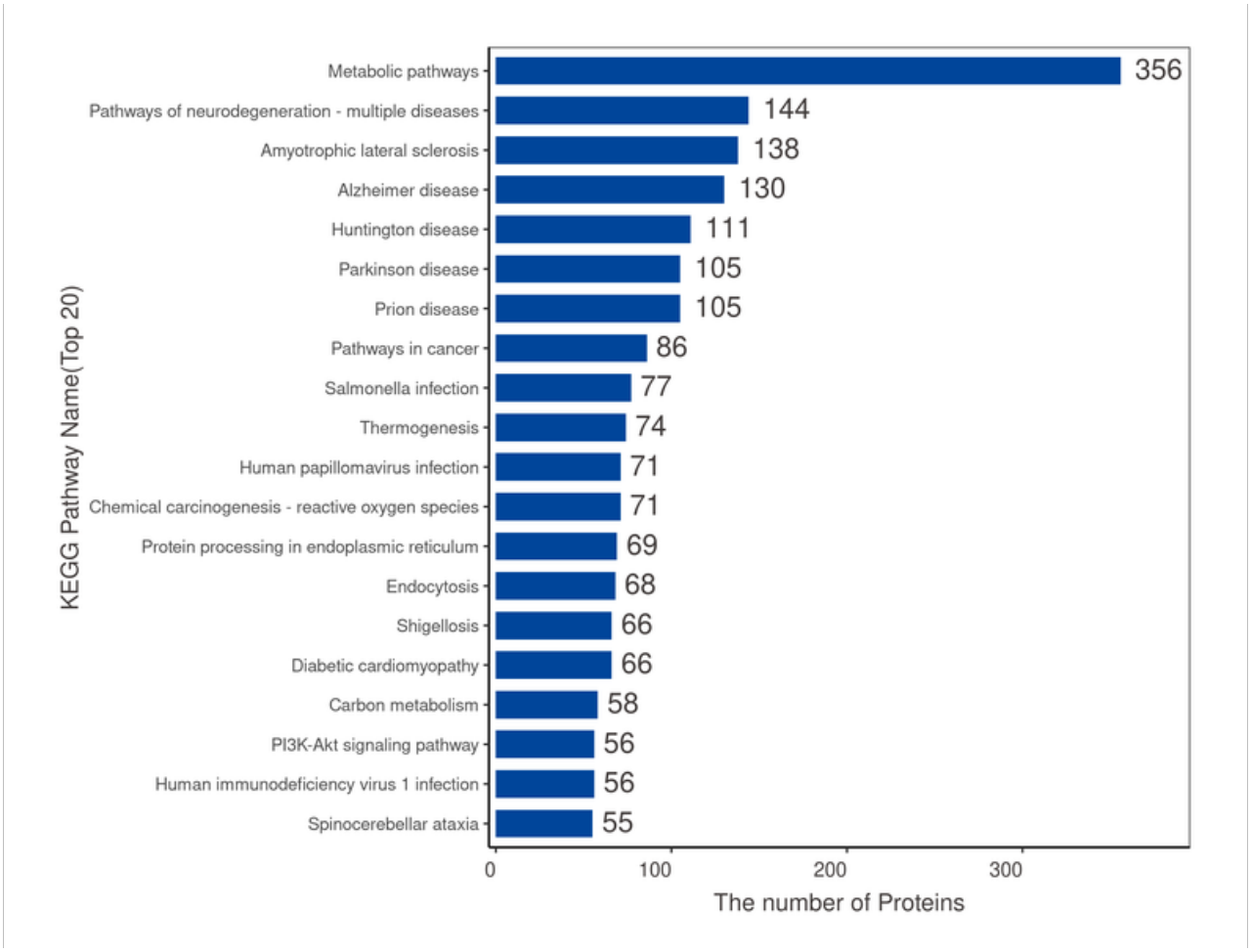

Figure 10 KEGG pathway annotation results bar chart (top 20)

The KEGG pathway enrichment analysis method is similar to GO enrichment analysis. Using KEGG pathways as the unit and all qualitative proteins as the background, it employs Fisher's exact test to calculate the significance levels of protein enrichment in each pathway, thereby identifying significantly affected metabolic and signaling pathways. The results of KEGG enrichment analysis for differential proteins are presented through circle diagrams, bar charts, and bubble plots.

Experiment time: 2025-06-12, 18:00 - 2025-06-15, 21:00

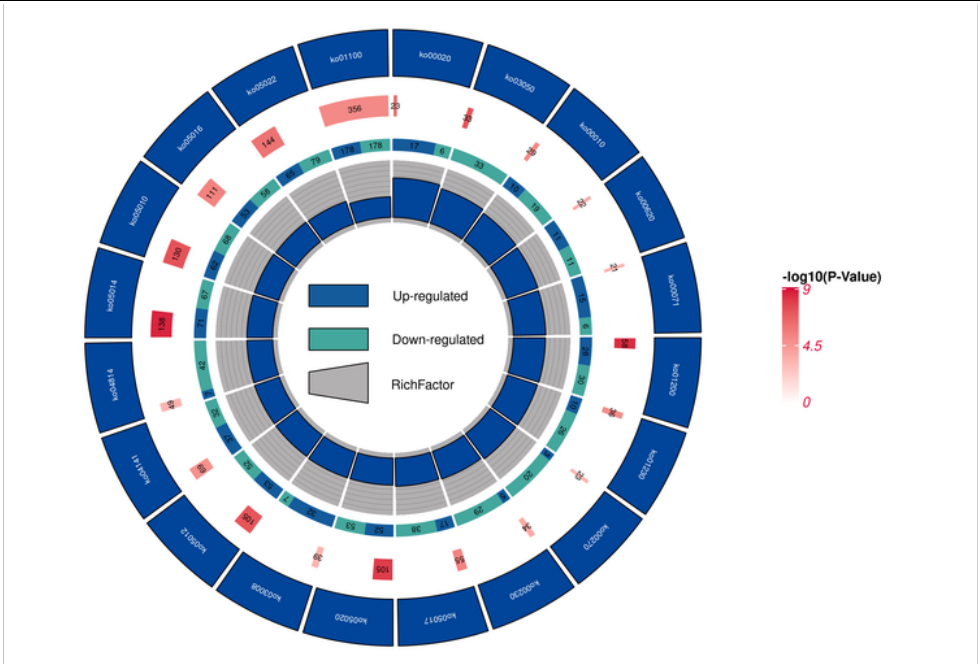

Figure 11 Top 20 enrichment statistics of KEGG pathways

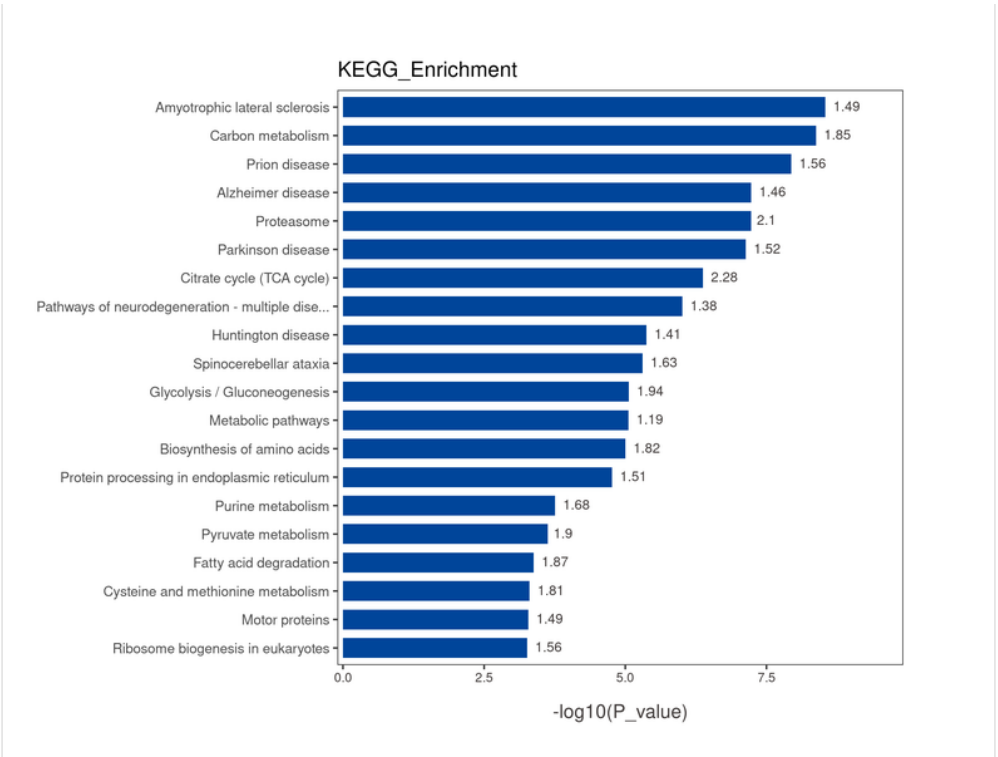

Figure 12 Enrichment statistics bar chart of KEGG pathways (top 20)

Experiment time: 2025-06-12, 18:00 - 2025-06-15, 21:00

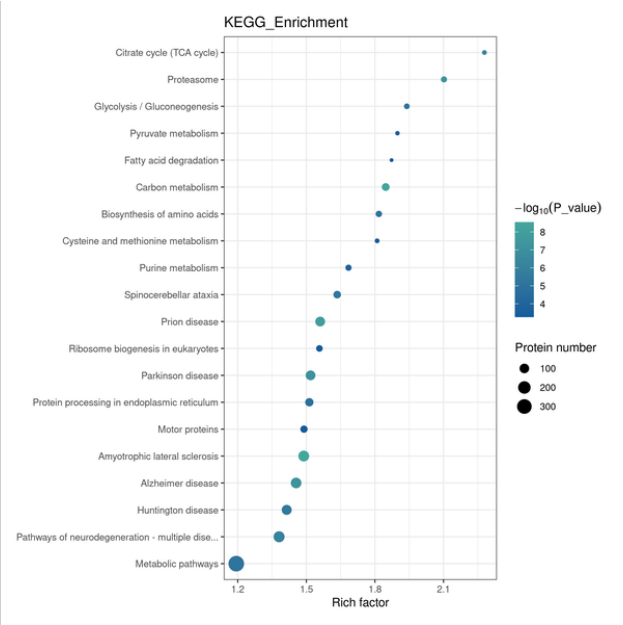

Figure 13 Enrichment statistics bubble plot of KEGG pathways (top 20)

According to Fold change, the differential proteins can be divided into upregulation and downregulation. In order to further understand the metabolic and signal transduction pathways involved in the upregulation and downregulation of the differential proteins, we plotted the KEGG enrichment bar chart of the upregulation and downregulation respectively, as shown in the figure below.

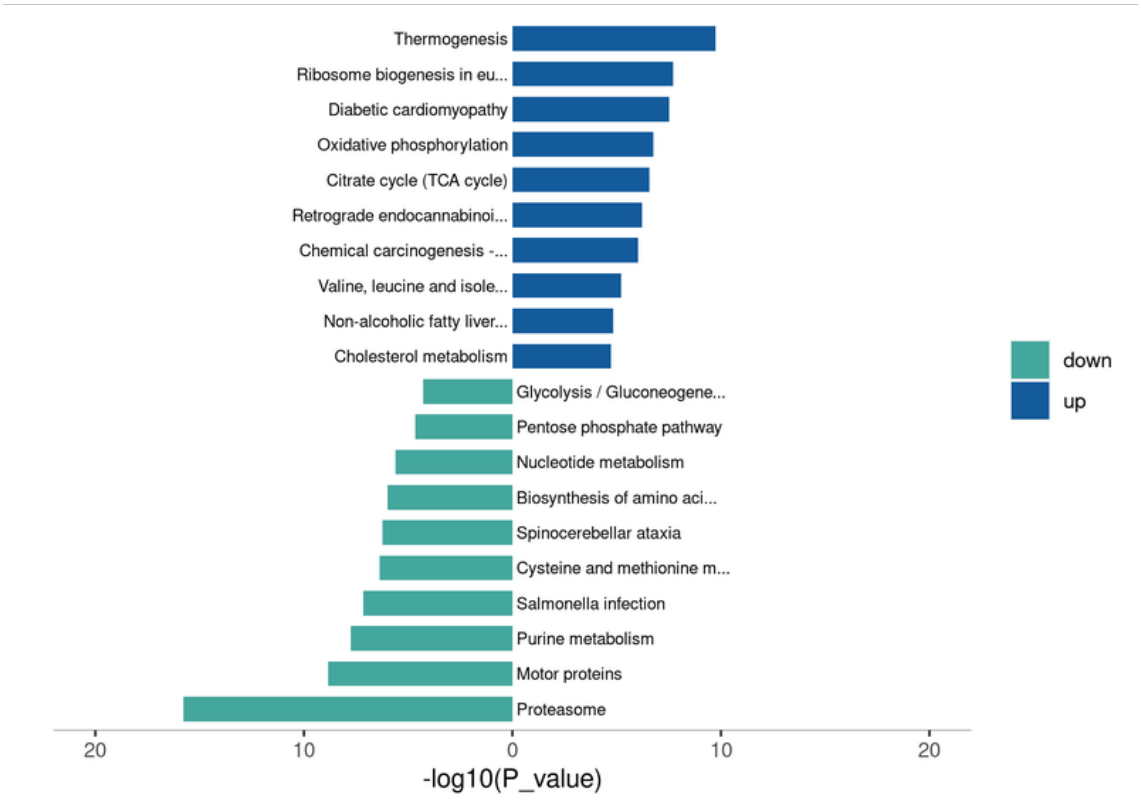

Figure 14 KEGG pathway enrichment analysis for up- and down-regulated DEPs

3.4 Transcription factor analysis

Transcription factors (TFs) are protein molecules that specifically bind to specific sequences upstream of the 5'-end of genes, ensuring the targeted gene is expressed at precise intensity in specific time and space. Given their critical role, these proteins require annotation and in-depth analysis. The AnimalTFDB (Animal Transcription Factor Database) and PlantTFDB (Plant Transcription Factor Database) databases contain information on transcription factors and their families in animals and plants, respectively. These databases can predict whether target proteins are TFs and identify their respective families. Annotation results for differentially expressed proteins are provided in the appendix. The statistical analysis of the top 10 transcription factor families is shown in the figure below.

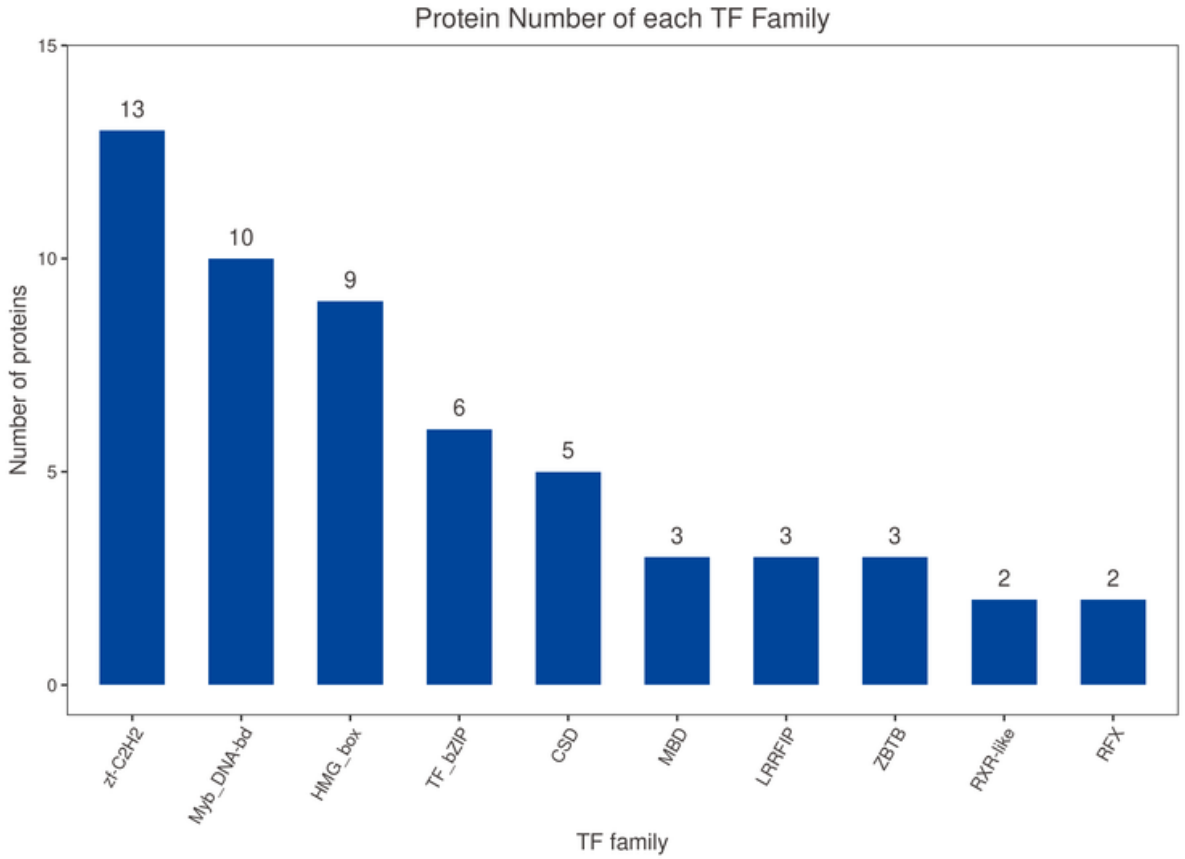

Figure 15 Statistics of the top 10 TF families in quantity

3.5 GSEA analysis

Traditional enrichment analysis focuses on comparing gene expression differences between two groups, concentrating on a small number of significantly upregulated or downregulated genes. This approach has certain limitations: (1) Unreasonable screening parameters may miss genes with significant biological importance that show no statistically significant differences; (2) When the number of differentially expressed proteins is small, traditional methods may yield minimal or even no results; (3) It struggles to answer critical questions like "If a pathway enriched by traditional methods contains both upregulated and downregulated genes, what does this indicate overall? Is it suppressed or activated?" Gene Set Enrichment Analysis (GSEA), which analyzes all gene expression levels rather than just differential proteins, effectively addresses these shortcomings. Additionally, GSEA can determine whether a pathway is activated or inhibited in specific sample groups.

The basic idea of GSEA is that there is no need to specify a clear threshold for differential

genes. Instead, it ranks all genes according to their differential expression degree in the two groups of samples, and then calculates the enrichment degree and significance of the pre-set gene set at the top or end of this ranking table.

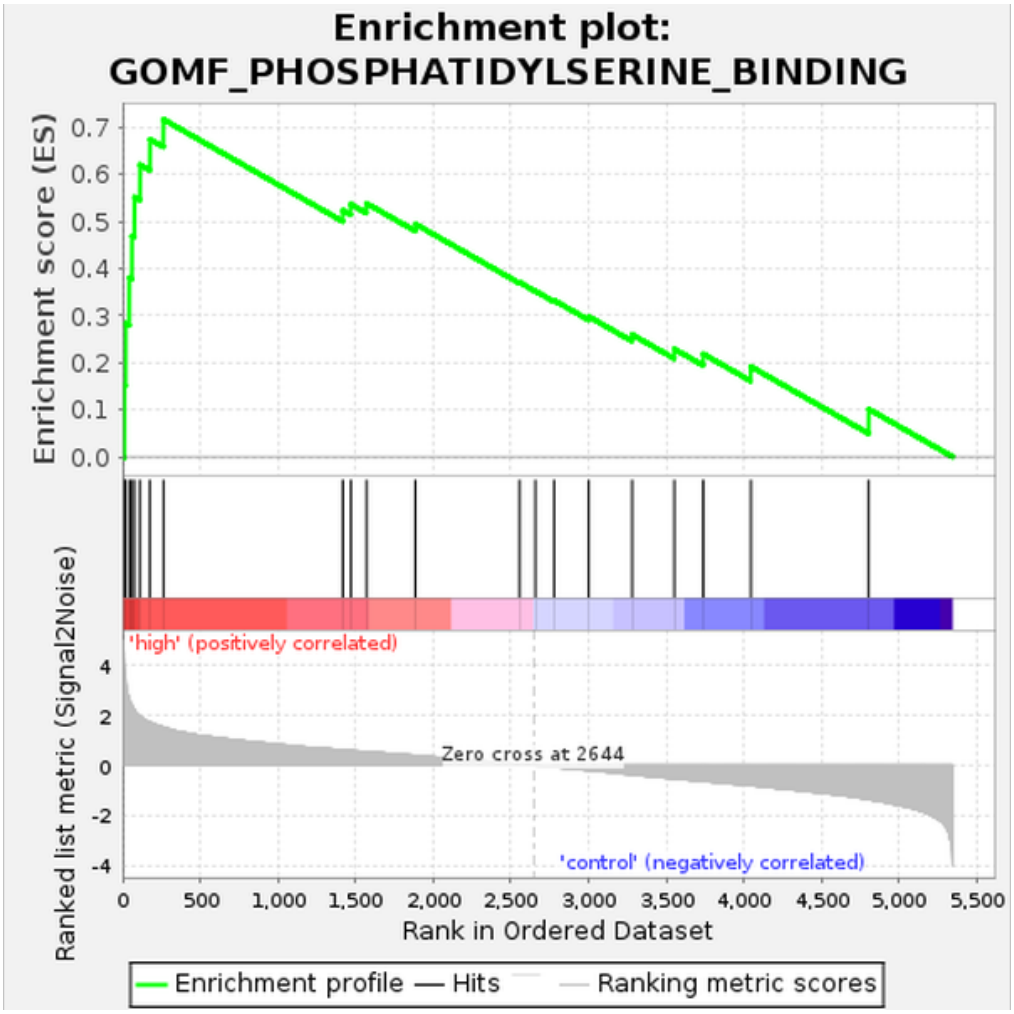

Figure 16 GSEA-GO enrichment score curve

Experiment time: 2025-06-12, 18:00 - 2025-06-15, 21:00

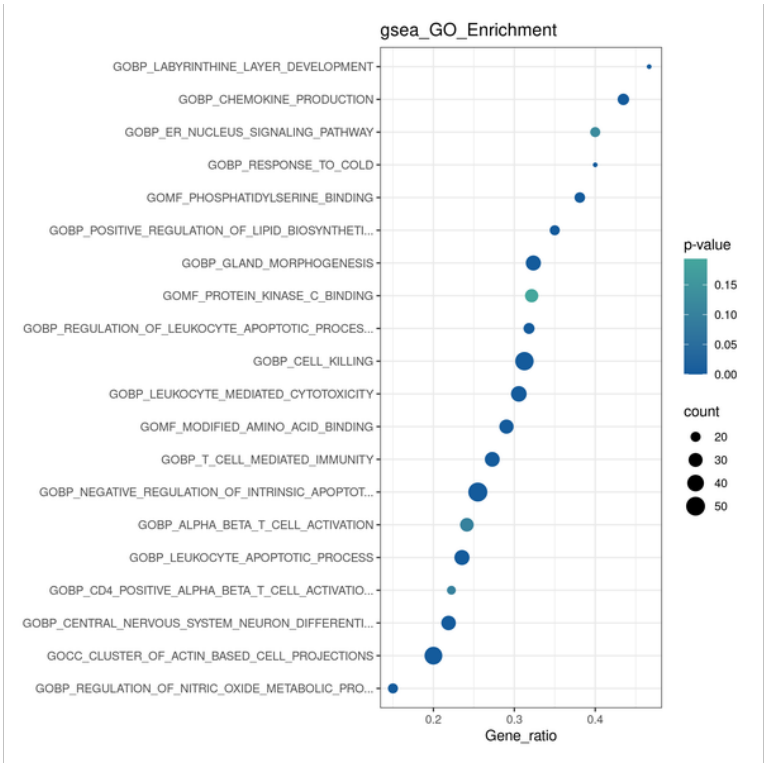

Figure 17 GSEA-GO enrichment bubble plot

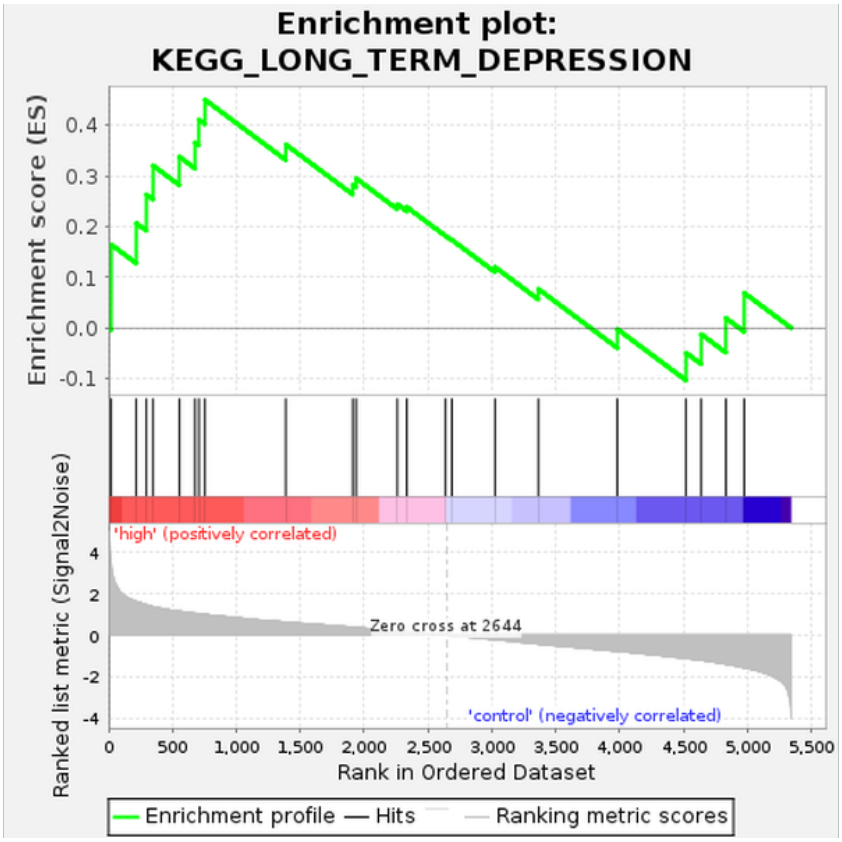

Figure 18 GSEA-KEGG enrichment score curve

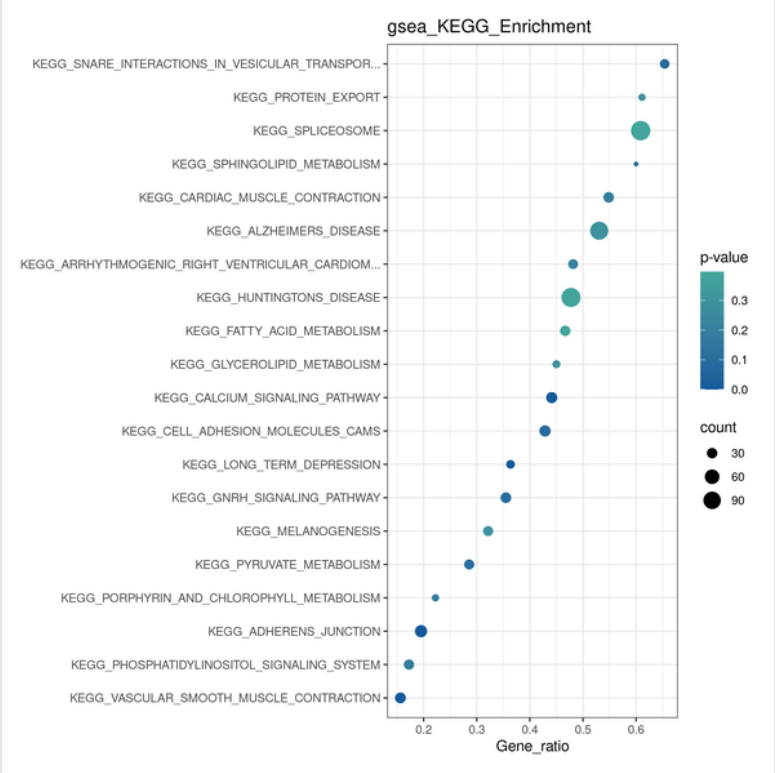

Figure 19 GSEA-KEGG enrichment bubble plot

3.6 List of key protein candidates

The top 20 core candidate proteins were selected based on the following criteria: 1) Significance of difference: minimum P-value ($P<0.001$) and maximum fold change ($|\log2FC|>3$); 2) Functional relevance: involvement in drug resistance-related pathways (e.g., metabolic reprogramming, protein ubiquitination); 3) Network centrality: degree>50 in PPI analysis.

Table 2 Top 20 candidate proteins associated with RC48 resistance.

Protein IDs	Gene Name	FC	p-value	regulation
P07204	THBD	5.593411	0.004579	UP
P05120	SERPINB2	5.383147	3.71E-05	UP
O76061	STC2	4.651881	0.000699	UP
P34741	SDC2	4.094121	0.000481	UP
P26022	PTX3	4.02582	0.001416	UP
Q86X29	LSR	3.967194	6.92E-05	UP
Q8IVT2	MISP	3.891363	0.000355	UP
P05362	ICAM1	3.452982	0.000328	UP
O00622	CCN1	3.328667	0.000932	UP
P17275	JUNB	3.257302	1.05E-05	UP
Q9NX18	SDHAF2	3.234066	0.012645	UP
P00749	PLAU	3.094785	0.001926	UP
P20591	MX1	3.078541	5.48E-05	UP
Q01201	RELB	2.950152	0.001589	UP
P20592	MX2	2.844591	0.008315	UP
Q13753	LAMC2	2.832641	0.000728	UP
Q9Y4K1	CRYBG1	2.779093	0.002373	UP
Q03405	PLAUR	2.712975	0.000375	UP
Q14574	DSC3	2.641861	0.002393	UP
P15407	FOSL1	2.590803	0.000367	UP

IV. Results analysis

This study systematically analyzed the molecular characteristics of T24-RC48 drug-resistant cells using proteomics technology. The research revealed that the drug-resistant phenotype is closely associated with metabolic reprogramming, protein homeostasis regulation, and abnormal activation of key signaling pathways. A variety of potential drug resistance-related protein biomarkers were identified, providing preliminary insights into the multi-level regulatory network governing cellular adaptation to drug stress. These findings offer crucial theoretical foundations and experimental directions for further investigation into the mechanisms of RC48 drug resistance and the development of targeted reversal strategies.