Experiment time: 2025-04-07, 17:00 - 2025-04-12, 16:00

Sample Preparation for Proteome Sequencing (II)

I. Objective of the experiment

The TRIzol lysis samples of T24 wild-type cell lines and low, medium and high drug-resistant T24-RC48 cell lines were prepared. These samples were then shipped on dry ice to the sequencing company for subsequent transcriptome sequencing analysis.

II. Experimental content

2.1 Experimental design

Sample type: T24 wild-type cell line, low/medium/high drug resistance T24-RC48 cell line.

Number of repetitions: 3 samples per group.

2.2 Measurement principle

RNA-seq (RNA sequencing) converts RNA into double-stranded cDNA libraries through reverse transcription, followed by sequencing to quantify gene expression. The workflow involves: extracting total RNA from samples, enriching mRNA using oligo (dT) magnetic beads (which bind to the poly-A tails of eukaryotic mRNA), fragmenting RNA, synthesizing cDNA with random primers, constructing libraries via end repair and ligation, and performing high-throughput sequencing on platforms like Illumina. This technology captures transcript sequences at single nucleotide resolution, where sequencing depth (the number of reads mapped to a gene) correlates directly with the abundance of its transcripts, enabling full transcriptome coverage and precise identification of structural variations such as alternative splicing and fusion genes. RNA-seq eliminates the need for probe design, offering high sensitivity and cross-species applicability. It is widely used in disease mechanism research, developmental regulation analysis, and novel transcript discovery.

III. Materials and reagents

3.1 Materials

Pipette tips and 15ml centrifuge tubes were purchased from Axygen

Serological pipettes were purchased from a domestic company

Micropipettes: Eppendorf

T25 culture flasks were purchased from Corning

10-cm cell culture dishes (Corning)

Liquid nitrogen, dry ice

Human bladder transitional cell carcinoma cell line (T24)

Low, medium, and high drug-resistant cell lines T24-RC48

3.2 Reagents

PBS and trypsin were purchased from Gibco, USA

McCoy's 5A medium from Shanghai Yuanpei Biotechnology Co., Ltd., China

Fetal bovine serum (FBS, Gibco)

IV. Experimental instruments

Clean bench: Suzhou purification equipment factory

Benchtop low-speed centrifuge at room temperature: Eppendorf

V. Experimental steps

5.1 Low, medium, high drug-resistant and wild-type cell lines were seeded in 10-cm cell culture dishes until the cells grew to more than 80%.

5.2 Discard the culture medium and add 4℃ pre-cooled PBS. Shake gently for 1 minute on a flat surface and discard the PBS. Repeat the above operation three times to fully wash off the culture medium.

5.3 Aspirate the PBS completely. Then, add 1 mL of TRIzol reagent per 10 million cells ($1x10^7$) to lyse the cells directly in the culture dish. Ensure the reagent covers the cell monolayer completely.

5.4 Using a pipette, repeatedly aspirate and dispense the TRIzol lysate to homogenize it until no cell clumps are visible. The lysate should appear clear and non-viscous, indicating complete dissociation. Transfer the homogeneous lysate to a pre-chilled, RNase-free 1.5 mL microcentrifuge tube.

5.5 Label the tube, immediately freeze it in liquid nitrogen, and then store it at -80°C.

5.6 Dry ice packing, fill in the handover form of the sequencing company (with sample information

sheet), and send it to the sequencing company.

VI. Photo of experimental operation

Figure 1 Samples prepared for transcriptome sequencing

VII. Experimental results

7.1 Correlation analysis between samples

By calculating Pearson correlation coefficients (R^2) of gene expression values (FPKM) across all samples, we found that the R^2 values between samples within each group were all greater than 0.985, indicating excellent intra-group reproducibility. Notably, the correlation coefficient between the drug-resistant group and the wild-type group showed a significant decrease, suggesting that drug resistance induction caused alterations in gene expression patterns. The correlation heatmap showed clear clustering of samples within the same group and separation between different groups, which aligns with the experimental design expectations.

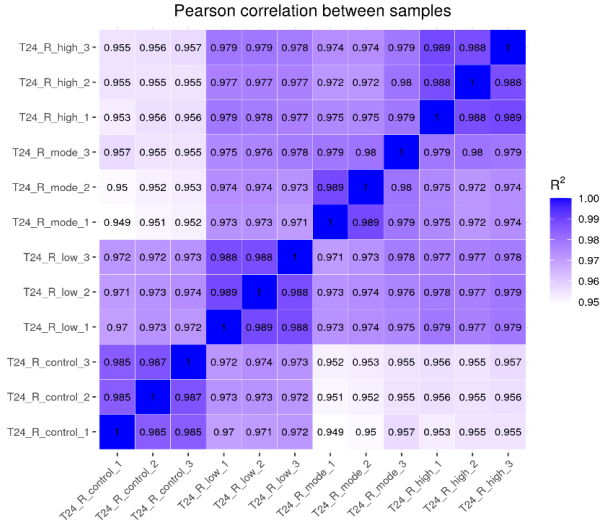

Figure 2 Inter-sample correlation

7.2 Principal component analysis (PCA)

The results of PCA and 3D-PCA based on FPKM values showed that the samples in each drug-resistant gradient group were closely clustered in the principal component space; the wild-type and drug-resistant groups were significantly separated along PC1 axis, and the low-, medium-, and high-resistance groups showed a stratified distribution along the PC2 axis.

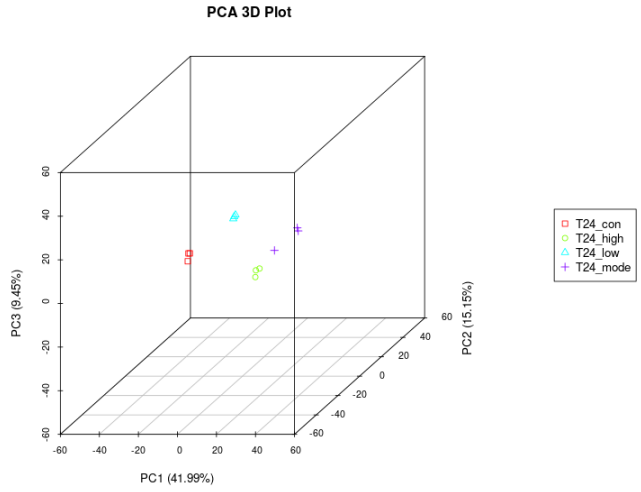

Figure 3 3D results of principal component analysis

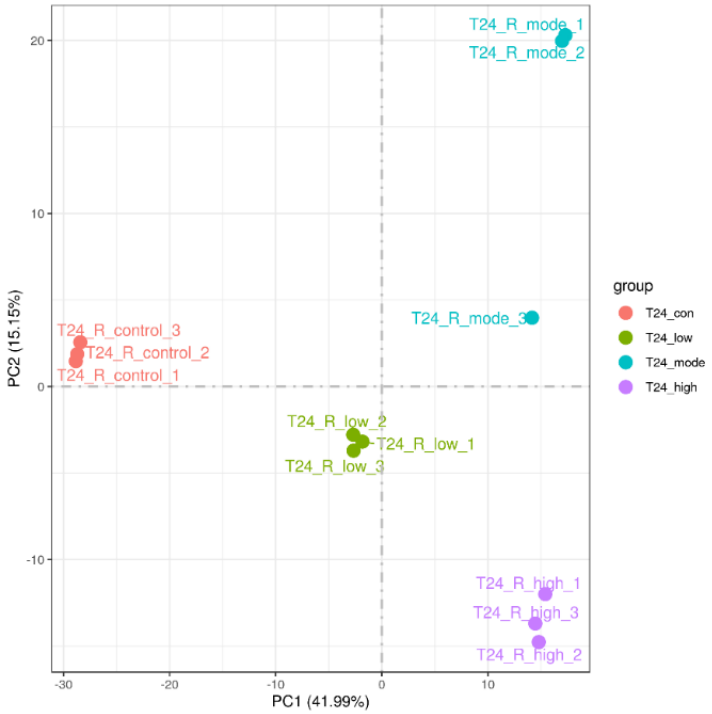

Figure 4 Results of principal component analysis (horizontal axis is the first principal component, vertical axis is the second principal component)

7.3 Differential gene overlap analysis

Through comparative analysis of differential genes across three drug resistance gradient groups using a triple Venn diagram, we identified 50 core shared genes in the high vs medium, high vs low, and medium vs low comparisons, revealing fundamental regulatory networks underlying drug resistance development. The overlapping regions between pairwise comparisons demonstrated distinct developmental phases: 178 genes shared between high vs medium and high vs low groups, 241 shared between high vs medium and medium vs low groups, and 361 shared between high vs low and medium vs low groups. These findings collectively demonstrate that drug resistance progression is accompanied by dynamic gene expression evolution.

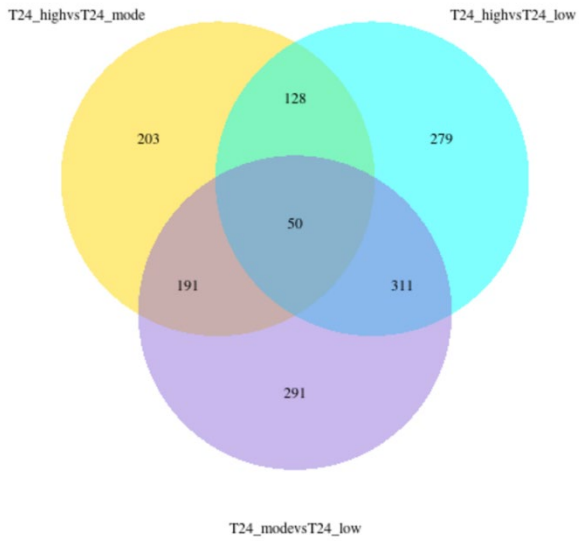

Figure 5 Venn Diagram of Differentially Expressed Genes

7.4 Gene expression distribution assessment

The box plot analysis based on log2 (FPKM+1) showed that the median FPKM value range of all samples was 1-2, and the interquartile range (IQR) was less than 3, indicating that the data standardization was effective. The proportion of outliers in each sample was relatively small, which met the quality requirements of high-throughput sequencing data.

Experiment time: 2025-04-07, 17:00 - 2025-04-12, 16:00

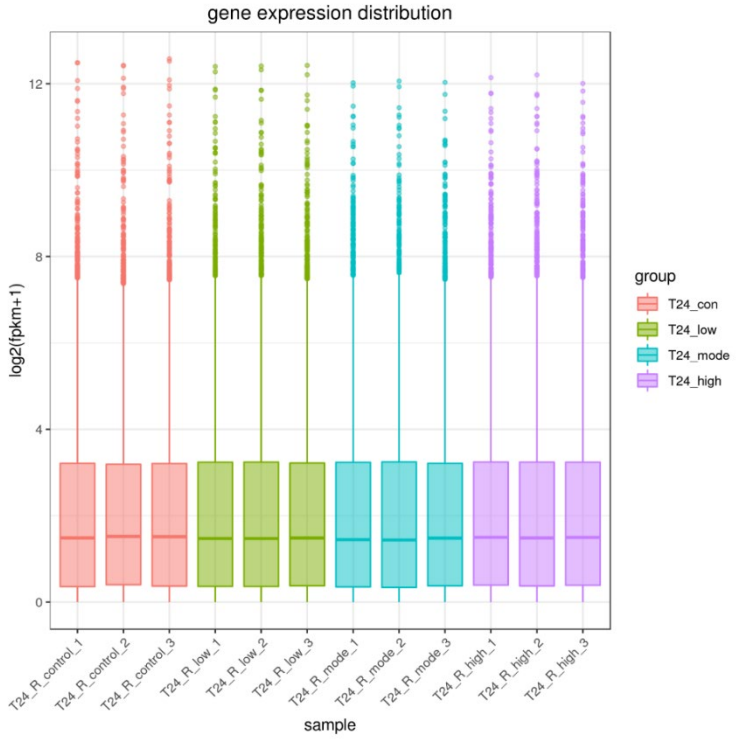

Figure 6 Gene expression distribution

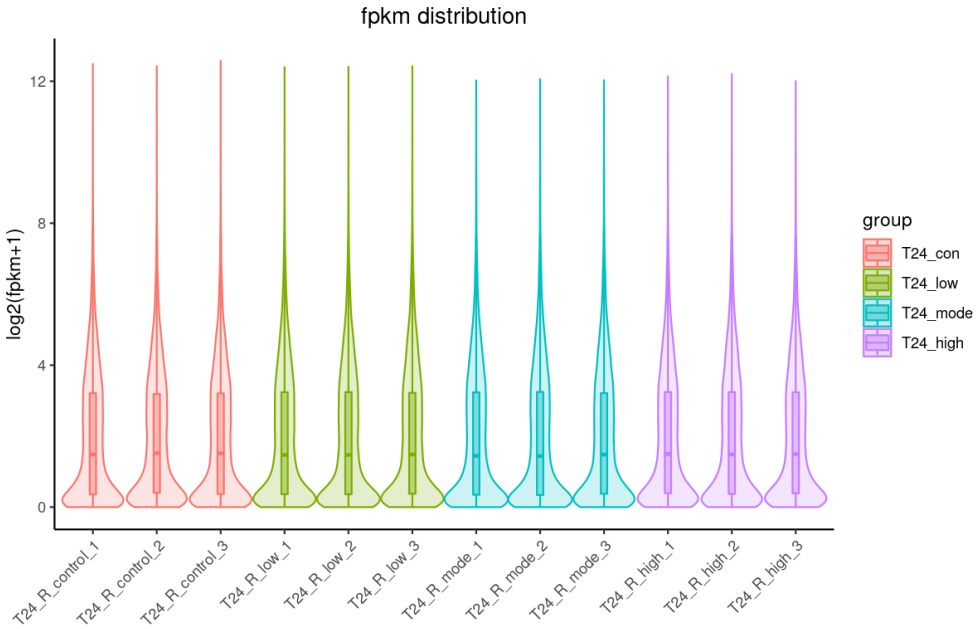

Figure 7 FPKM distribution

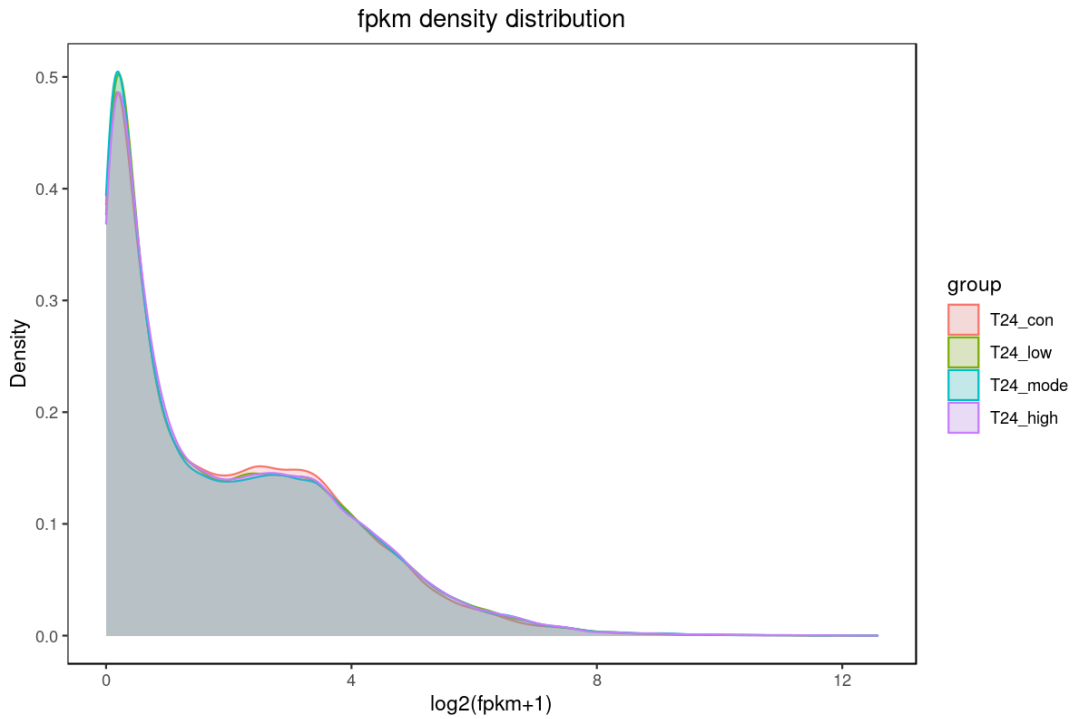

Figure 8 FPKM density distribution

VIII. Results analysis

The correlation of gene expression levels among samples serves as a crucial indicator for evaluating experimental reliability and the rationality of sample selection. A correlation coefficient closer to 1 indicates a higher degree of similarity in expression patterns between samples. The ENCODE Project recommends that the Pearson's correlation coefficient squared (R^2) should exceed 0.92 (ENCODE Project Consortium, 2004). In this study, the R^2 values between samples within each group all exceeded 0.985, demonstrating excellent intra-group reproducibility. Notably, the correlation coefficient between the drug-resistant group and the wild-type group showed a significant decrease. These results confirm the reproducibility of the experimental procedures and highlight the high level of standardization in key steps such as cell culture and TRIzol-based cell lysis.

Principal Component Analysis (PCA) is widely used to evaluate inter-group differences and intra-group sample redundancy. By employing linear algebraic calculations, PCA reduces dimensionality and extracts principal components from tens of thousands of genetic variables. Under ideal conditions, PCA plots should show samples across groups dispersed while intra-group samples cluster together. In this experiment, samples within each drug resistance gradient group

clustered closely in the principal component space: wild-type versus drug-resistant groups were significantly separated along PC1 axis, while low-, medium-, and high-drug resistance groups formed stratified distributions along PC2 axis. This pattern indicates a strong correlation between drug resistance variations and changes in principal components.

The Venn diagram reveals overlapping differential genes across experimental comparisons, enabling identification of shared or unique differences. When a single comparison exists, we generate a Venn diagram showing co-expressed genes between the treatment and control groups. The total number of differential genes in the circle represents the combined count, while overlapping regions indicate shared genes. Our study identified 50 core shared genes across three comparisons (high vs medium, high vs low, medium vs low), which may constitute key regulatory hubs driving drug resistance development. The pairwise overlaps revealed distinct phases: 178 genes shared between the high-vs-medium and high-vs-low comparisons; 241 genes shared between the high-vs-medium and medium-vs-low comparisons; and 361 genes shared between the high-vs-low and medium-vs-low comparisons. These patterns demonstrate dynamic evolution of gene expression during drug resistance progression, suggesting that different resistance stages might rely on specific molecular mechanisms.