Sample processing:

(1) After the samples were thawed, vortex for 10 s to stir evenly;

(2) A 50 μL sample was taken and added to a numbered 1.5 ml centrifuge tube and 250 μL of 20% acetonitrile methanol extract was added;

(3) Vortex for 3 min and then centrifuge for 10 min at 12,000 r/min, 4 °C;

(4) After centrifugation, 250 μL of supernatant was pipetted into another 1.5 ml centrifuge tube and allowed to stand for 30 min at -20°C in a refrigerator;

(5) 12000 r/min and centrifuge at 4 °C for another 10 min;

(6) Take 180 μL of centrifuged supernatant and pass it through a protein precipitation plate for instrumental analysis. The processed supernatant should be stored at -20 °C for preservation.

Chromatography mass spectrometry acquisition conditions:

The chromatographic column was an ACQUITY BEH Amide column with 1.7 μm I.D., 2.1 mm I.D. and 100 mm length.The mobile phase was divided into phase A and phase B. Phase A was ultrapure water containing 2 mm ammonium acetate and 0.04% formic acid, while phase B was acetonitrile, also containing 2 mm ammonium acetate and 0.04% formic acid.The gradient elution program is set to:The volume ratio of phase A to phase B was 10:90 from the beginning to 1.2 min; then the ratio was gradually adjusted to 40:60 over 9 min; then the ratio changed to 60:40 from 10 to 11 min; and finally, the ratio returned to 10:90 from 11.01 min to 15 min.The flow rate was maintained at 0.4 ml/min throughout the process, the column temperature was set at 40 °C, and the injection volume was 2 μL.

The conditions for mass spectrometry analysis were as follows:The temperature of the electrospray ionization source was set to 550 °C. In positive ion mode, the mass spectrometry voltage was 5500 V, while in negative ion mode, the mass spectrometry voltage was adjusted to -4500 V. The temperature of the electrospray ionization source was set to 550 °C.The Curtain Gas (CUR) pressure was set to 35 psi. In the Q-Trap 6500+ mass spectrometer, each ion pair was scanned and detected with optimized declustering potential (DP) and collision energy (CE) to ensure the highest analytical sensitivity and accuracy.This step is essential to ensure the reliability and accuracy of the mass spectrometry data.

Qualitative and quantitative principles

(1) A metware database (MWDB) was constructed based on the standards, and the results of the data obtained using mass spectrometry were analyzed qualitatively.

(2) Quantitative data analysis by triple quadrupole mass spectrometry in multiple reaction monitoring (MRM) mode is an efficient and accurate method. As shown in Fig.1: In this mode, the four-stage rod first screens the precursor ions (i.e., parent ions) of the target substance, thus eliminating ions corresponding to other molecular weight substances, and achieving an initial reduction of interference. Subsequently, these precursor ions undergo induced ionization in the collision chamber and break into multiple fragment ions. These fragment ions are further subjected to fine filtration by a triple four-stage rod in order to select specific and characteristic fragment ions, thus further eliminating the interference of non-target ions and greatly improving the accuracy and reproducibility of quantification. Once the mass spectrometry analysis data for the different samples were obtained, the next step was to integrate the chromatographic peaks for all the target substances. This process accurately reflects the amount of the target substance in the

sample. Finally, quantitative analysis of the target substance can be achieved by comparing the standard curves.

Data preprocessing

Mass spectrometry data is analyzed in depth using MultiQuant 3.0.3 software, which integrally corrects the chromatographic peaks in the sample of analytes to be tested based on retention time and peak shape data from the standards.This process ensures the precision of qualitative and quantitative analysis.

standard curve

Standard solutions were prepared at different concentrations of 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000 and 20000 ng/ml. The peak intensity values of each concentration standard in the mass spectrum were analyzed and determined. The concentration rate of the external and internal standards was used as the x-axis. The peak area ratio of the external standard and internal standard was taken as the Y-axis. The standard curve of each substance was constructed.

Sample content

The integrated peak area ratios of the measured samples were calculated by the linear equations of the standard curves, and the real concentration data of the compounds in the measured samples were obtained by substituting the calculation formulas.

$$\text{material content}\ (\ \text{ng/mL}\ ) = c * V1/1000/V2$$

Note: c: concentration value (ng/ml) obtained by substituting the integrated peak area ratio in the sample into the standard curve equation; V1: total volume of extraction solution (μL); V2: volume of sample pipetted (ml).

Sample quality control analysis

Repeatability and reliability were tested using total ion chromatogram (TIC), Pearson correlation analysis, and coefficient of variation (CV) values.

Overall sample principal component preanalysis

A principal component preanalysis was performed on all samples in order to get an initial idea of the overall differences in metabolites between the samples in each group and the degree of variability between all samples within the group.

Analysis of data results

The onboard raw data were unit variance scaling (UV) and then subjected to Principal Component Analysis (PCA) using the MetaboAnalystR package OPLSR.Anal function in the R software.

Differential Metabolite Screening and Annotation

This study was screened using univariate statistical analyses such as hypothesis testing and fold change (FC) analysis of variance, combined with PCA multivariate analysis.*P*-value < 0.05,

fold change ≥ 2 and fold change ≤ 0.5 were used as screening criteria for major metabolites. Finally, ROC curves and AUC evaluations were utilized to screen for differential metabolites for sensitivity, specificity, and differential diagnostic ability.

Results:

From the results of the total ion chromatogram (TIC) of the QC samples analyzed by mass spectrometry detection, the stability of the instrument was good throughout the process of the QC samples being analyzed (Fig. 2). Pearson's (Pearson) correlation analysis was performed on the QC samples, and the results indicated good stability of the instrument throughout the testing process and good quality of the data results (Fig. 3). In this study, the percentage of substances with CV values <0.3 was more than 80%, while the percentage of substances with CV values <0.2 was also more than 80%. These indicate the high quality of the experimental data with significant stability. (Fig. 4). The data were standardized and subjected to principal component analysis (Fig. 5), and the explanatory ratios of the first five principal components of PCA are shown in (Fig. 6). The three-dimensional plot of PCA suggests that the samples are differentiated in the distribution of the two groups of samples. After the initial screening of metabolites and selection using the screening conditions, we identified 21 metabolites with significant differences (Fig. 7) and plotted their violin maps to show their distribution (Fig. 8).

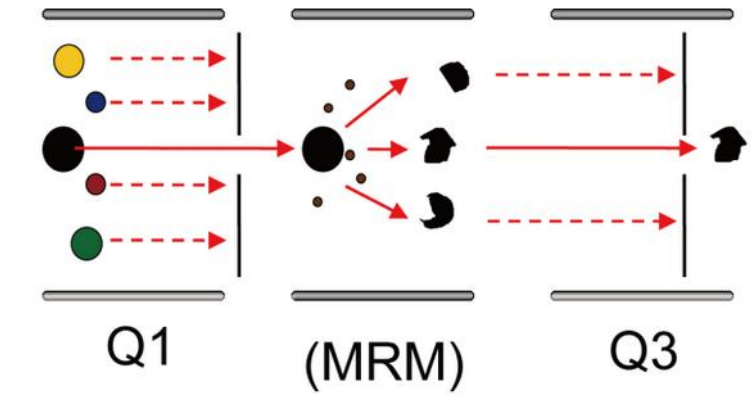

Fig.1 MRM Flowchart

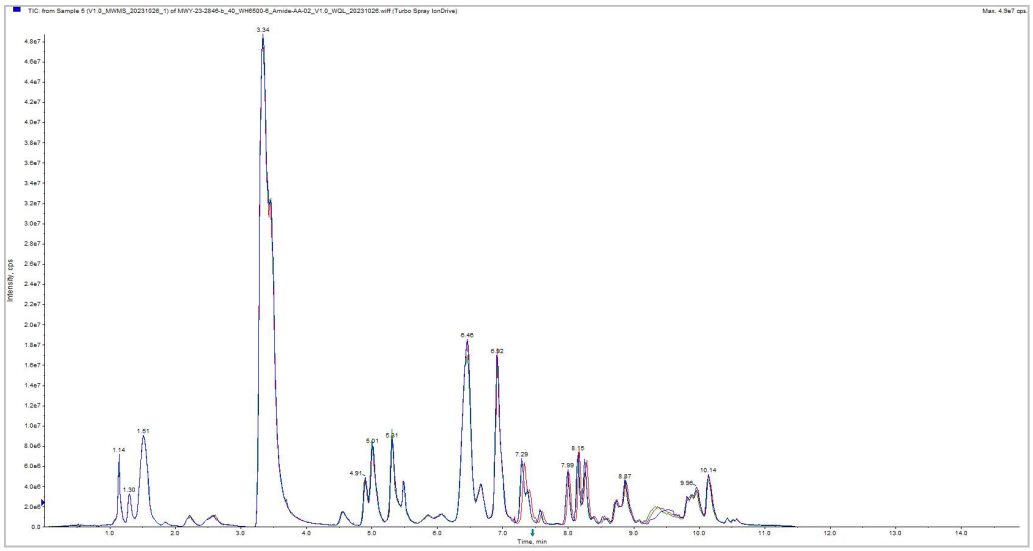

Fig.2 TIC overlap diagram. The retention times and peak intensities are consistent, indicating that the mass spectrometry has good signal stability when the same sample is detected at different times.

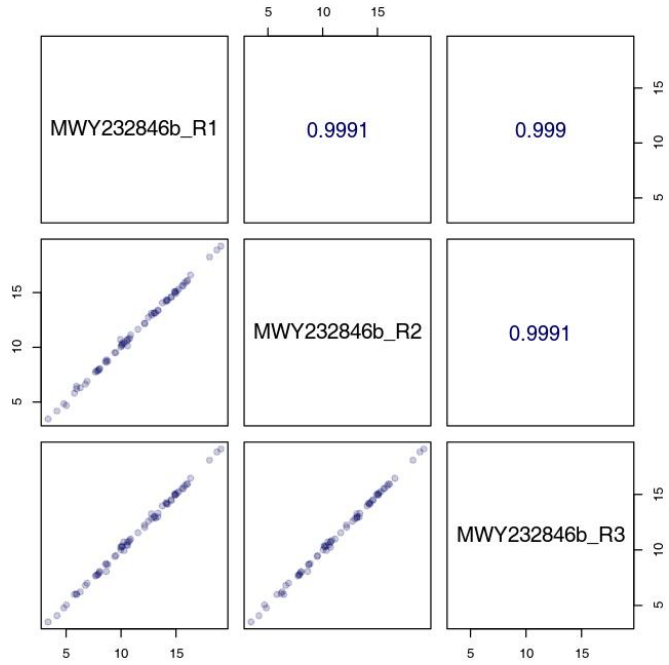

Fig.3 QC sample correlation analysis graph. The diagonal square represents the name of the QC sample; the lower-left diagonal square is the correlation scatter plot of the corresponding QC sample, the horizontal and vertical coordinates are the metabolite contents, and each point in the plot represents a metabolite; the upper-right diagonal square is the correlation coefficient of the corresponding QC sample.

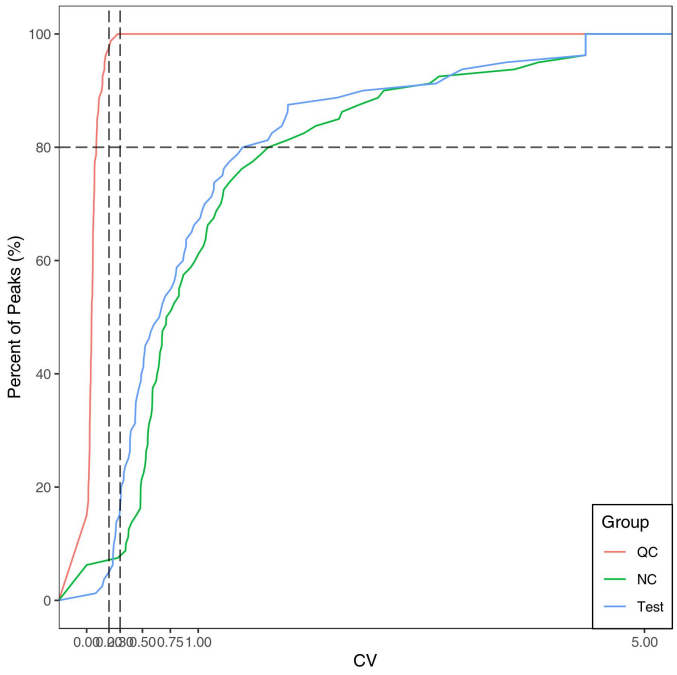

Fig. 4 Distribution of CVs in each group of samples. The horizontal coordinate represents the CV value, the vertical coordinate indicates the proportion of the number of substances smaller than the corresponding CV value to the total number of substances, different colors represent different grouping samples, and QC is the quality control sample, in which the two reference lines perpendicular to the X-axis correspond to the CV values of 0.2 and 0.3, and the number of substances corresponding to the reference line parallel to the X-axis accounts for 80% of the total number of substances.

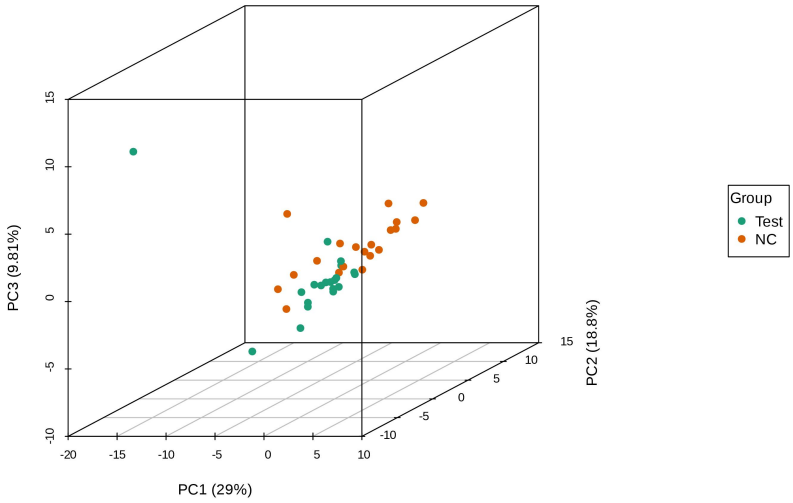

Fig.5 PCA three-dimensional image

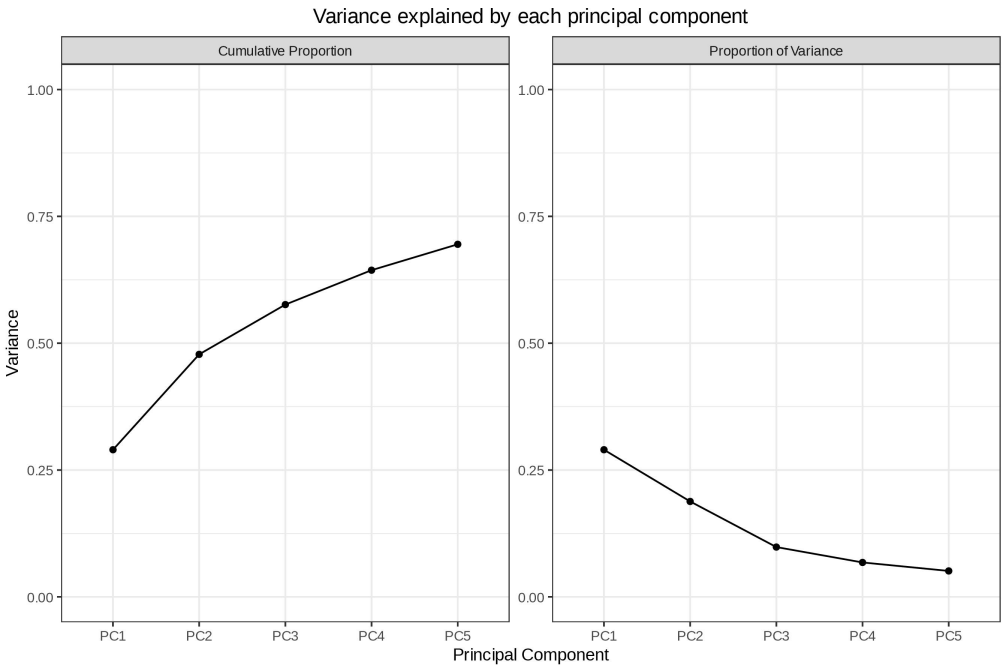

Fig.6 Principal component analysis can explain the variogram. Horizontal coordinates indicate the individual principal components, vertical coordinates indicate the rate of explanation of the principal components to the data set, the left graph shows the cumulative rate of explanation, and the right graph shows the rate of explanation of the individual principal components.

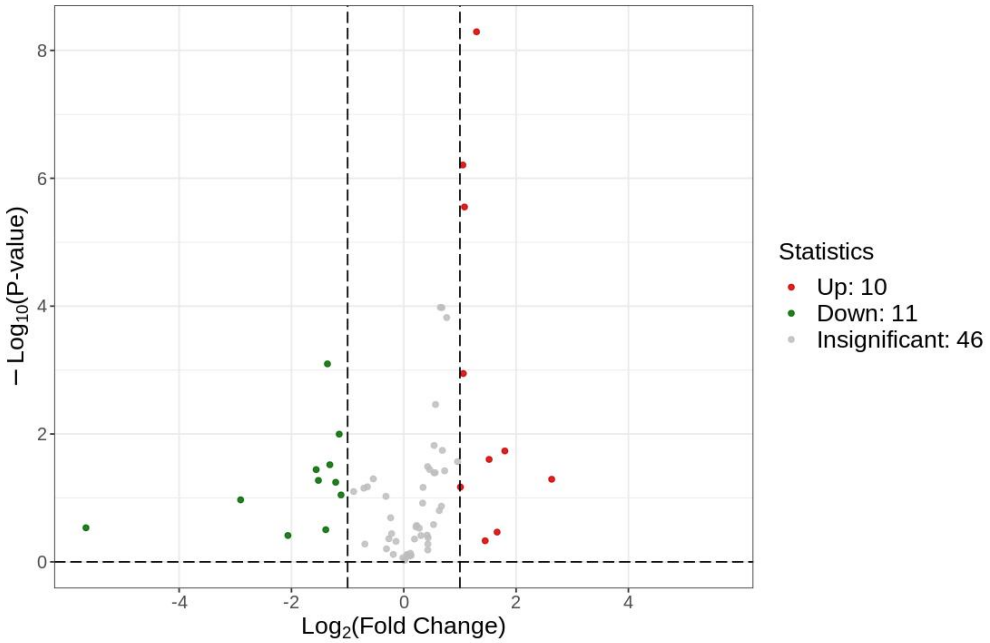

Fig.7 Differential metabolite volcano map. Under the FC + *P*-value double screening condition, each point in the volcano plot represents a metabolite, the horizontal coordinate indicates the fold change of metabolite differences in different subgroups (log2FoldChange), and the vertical coordinate indicates the level of significance of the differences (-log10p-value), and the greater the absolute value of the horizontal coordinate, the greater the fold difference of expression between the two samples of the expressions The larger the absolute value of the horizontal coordinate, the greater the fold difference in expression between the two samples; Larger vertical coordinate values indicate more significant differential expression The green points in the graph represent down-regulated differentially expressed metabolites, the red points represent up-regulated differentially expressed metabolites, and the gray points represent metabolites that were detected but not differentially significant.

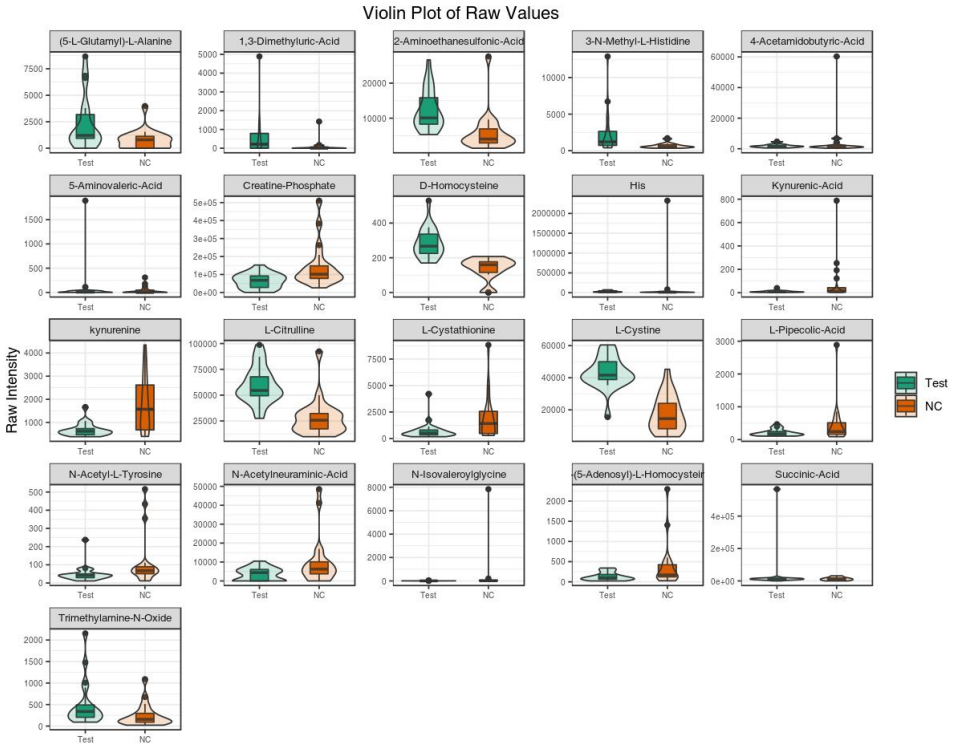

Fig. 8 Differential metabolite violin map. Horizontal coordinates are groups, vertical coordinates are expressions. The box shape in the center indicates the interquartile range, the thin black line extending from it represents the 95% confidence interval, the black horizontal line right in the middle is the median, and the outer shape indicates the density of the distribution of the data.