



Innovative Metabolomics Insights for Better Health

Anthocyanidin Metabolomics Assay Final Report

Metware Biotechnology Inc.

www.metwarebio.com

Contents

1	Abstract	3
2	The experimental process	3
2.1	Sample information	6
2.2	Reagents and instruments	7
2.3	Sample extraction process	7
2.4	Chromatography-mass spectrometry acquisition conditions	8
2.5	Qualitative and quantitative principles of metabolites	8
3	Data evaluation	9
3.1	Data pre-processing	9
3.2	Standard Solution Preparation	10
3.3	Quantification Results	11
3.4	Sample Quality Control Analysis	12
3.5	Sample quantification histogram	13
3.6	Hierarchical Cluster Analysis	14
4	Analysis results	15
4.1	Differential metabolite screening	15
4.2	Functional annotation and enrichment analysis of differential metabolites in KEGG database	17
5	References	20
6	Appendix	21
6.1	Analytical methods	21
6.2	List of software and versions	21

MWXS-23-xxx Anthocyanidin Metabolomics Assay Final Report

1 Abstract

Metabolites are the basis of organism phenotype and can help to understand biological processes and mechanisms more intuitively and effectively. Based on qualitative and quantitative analysis of metabolites, metabolomics can be used to analyze metabolic pathways or metabolic networks, study the metabolic basis of macroscopic phenotypic phenomena of different organisms, response mechanisms of metabolites stimulated by physical, chemical or pathogenic organisms such as diseases and drugs, and safety evaluation of food and drugs. Anthocyanidin metabolomics is a targeted quantitative detection method. It has accurate characteristics. There are 108 metabolites in the anthocyanidin metabolomics database. They include pelargonidins, cyanidins, delphinidins, peonidins, petunidins, malvidins, procyanidins, and so on.

2 The experimental process

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) can detect and quantify compounds with high polarity and poor thermal stability, and accurately quantify them. The overall process is as follows:

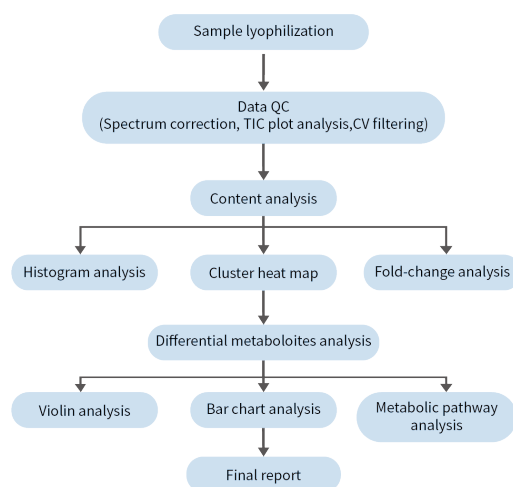


Fig 1: Flow chart of metabolomics analysis

Compounds to be detected:

Table 1: List of compounds in the panel

Number	Compounds	Index
1	Cyanidin-3-(6-O-p-caffeoyl)-glucoside	Anthocyanidin_01
2	Cyanidin-3-(6"-caffeoylsophoroside)-5-glucoside	Anthocyanidin_02
3	Cyanidin-3,5,3'-O-triglucoside	Anthocyanidin_03

Table 1: List of compounds in the panel

Number	Compounds	Index
4	Cyanidin-3,5-O-diglucoside	Anthocyanidin_04
5	Cyanidin-3-O-(6-O-malonyl-beta-D-glucoside)	Anthocyanidin_05
6	Cyanidin-3-O-(6-O-p-coumaroyl)-glucoside	Anthocyanidin_06
7	Cyanidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	Anthocyanidin_07
8	Cyanidin-3-O-arabinoside	Anthocyanidin_08
9	Cyanidin-3-O-(6"-ferulysophoroside)-5-glucoside	Anthocyanidin_09
10	Cyanidin-3-O-galactoside	Anthocyanidin_10
11	Cyanidin-3-O-rutinoside-5-O-glucoside	Anthocyanidin_13
12	Cyanidin-3-O-sambubioside	Anthocyanidin_14
13	Cyanidin-3-O-sophoroside	Anthocyanidin_15
14	Cyanidin-3-O-xyloside	Anthocyanidin_16
15	Cyanidin-3-O-sambubioside-5-O-glucoside	Anthocyanidin_17
16	Delphinidin	Anthocyanidin_18
17	Delphinidin-3,5-O-diglucoside	Anthocyanidin_19
18	Delphinidin-3-O-(6-O-acetyl)-glucoside	Anthocyanidin_20
19	Delphinidin-3-O-(6-O-malonyl-beta-D-glucoside)	Anthocyanidin_21
20	Delphinidin-3-O-(6-O-malonyl)-glucoside-3'-glucoside	Anthocyanidin_22
21	Delphinidin-3-O-(6-O-p-coumaroyl)-glucoside	Anthocyanidin_23
22	Delphinidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	Anthocyanidin_24
23	Delphinidin-3-O-arabinoside	Anthocyanidin_25
24	Delphinidin-3-O-galactoside	Anthocyanidin_26
25	Delphinidin-3-O-glucoside	Anthocyanidin_27
26	Delphinidin-3-O-rhamnoside	Anthocyanidin_28
27	Delphinidin-3-O-rutinoside	Anthocyanidin_29
28	Delphinidin-3-O-sambubioside	Anthocyanidin_30
29	Delphinidin-3-O-sophoroside	Anthocyanidin_31
30	Delphinidin-3-O-rutinoside-5-O-glucoside	Anthocyanidin_32
31	Delphinidin-3-O-sambubioside-5-O-glucoside	Anthocyanidin_33
32	Afzelin	Anthocyanidin_34
33	Chalcone	Anthocyanidin_35
34	Dihydrokaempferol	Anthocyanidin_36
35	Dihydromyricetin	Anthocyanidin_37
36	Kaempferol-3-O-rutinoside	Anthocyanidin_38
37	Naringenin	Anthocyanidin_39
38	Naringenin-7-O-glucoside	Anthocyanidin_40
39	Quercetin-3-O-glucoside	Anthocyanidin_41
40	Rutin	Anthocyanidin_42
41	Malvidin	Anthocyanidin_43
42	Malvidin-3,5-O-diglucoside	Anthocyanidin_44
43	Malvidin-3-O-(6"-acetylglucoside)-5-glucoside	Anthocyanidin_45
44	Malvidin-3-O-(6-O-malonyl-beta-D-glucoside)	Anthocyanidin_46
45	Malvidin-3-O-(6-O-p-coumaroyl)-glucoside	Anthocyanidin_47
46	Malvidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	Anthocyanidin_48
47	Malvidin-3-O-arabinoside	Anthocyanidin_49
48	Malvidin-3-O-galactoside	Anthocyanidin_50

Table 1: List of compounds in the panel

Number	Compounds	Index
49	Malvidin-3-O-glucoside	Anthocyanidin_51
50	Malvidin-3-O-rutinoside	Anthocyanidin_52
51	Malvidin-3-O-sambubioside	Anthocyanidin_53
52	Malvidin-3-O-sophoroside	Anthocyanidin_54
53	Malvidin-3-O-sambubioside-5-O-glucoside	Anthocyanidin_55
54	Pelargonidin	Anthocyanidin_56
55	Pelargonidin-3-(6''-caffeylsophoroside)-5-glucoside	Anthocyanidin_57
56	Pelargonidin-3-sophoroside-5-glucoside	Anthocyanidin_58
57	Pelargonidin-3,5-O-diglucoside	Anthocyanidin_59
58	Pelargonidin-3-O-(6''-ferulylsambubioside)-5-O-(malonyl)-glucoside	Anthocyanidin_60
59	Pelargonidin-3-O-(6-O-malonyl-beta-D-glucoside)	Anthocyanidin_61
60	Pelargonidin-3-O-(6-O-p-coumaroyl)-glucoside	Anthocyanidin_62
61	Pelargonidin-3-O-[2-O-glucosyl-6-O-p-coumaroyl-glucoside]-5-O-glucoside	Anthocyanidin_63
62	Pelargonidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	Anthocyanidin_64
63	Pelargonidin-3-O-[6-O-feruloyl-2-O-glucosyl-glucoside]-5-O-glucoside	Anthocyanidin_65
64	Pelargonidin-3-O-arabinoside	Anthocyanidin_66
65	Pelargonidin-3-O-galactoside	Anthocyanidin_67
66	Pelargonidin-3-O-glucoside	Anthocyanidin_68
67	Pelargonidin-3-O-rutinoside	Anthocyanidin_69
68	Pelargonidin-3-O-rutinoside-5-O-glucoside	Anthocyanidin_70
69	Pelargonidin-3-O-sambubioside	Anthocyanidin_71
70	Pelargonidin-3-O-sophoroside	Anthocyanidin_72
71	Pelargonidin-3-O-sophoroside-5-O-(malonyl)-glucoside	Anthocyanidin_73
72	Pelargonidin-3-O-sambubioside-5-O-glucoside	Anthocyanidin_74
73	Peonidin	Anthocyanidin_75
74	Peonidin-3-(caffeoyl-glucosyl-glucoside)-5-glucoside	Anthocyanidin_76
75	Peonidin-3-sophoroside-5-glucoside	Anthocyanidin_77
76	Peonidin-3,5-O-diglucoside	Anthocyanidin_78
77	Peonidin-3-O-(6-O-malonyl-beta-D-glucoside)	Anthocyanidin_79
78	Peonidin-3-O-(6-O-p-coumaroyl)-glucoside	Anthocyanidin_80
79	Peonidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	Anthocyanidin_81
80	Peonidin-3-O-arabinoside	Anthocyanidin_82
81	Peonidin-3-O-caffeoyl-feruloyl-sophoroside-5-glucoside	Anthocyanidin_83
82	Peonidin-3-O-(6''-ferulylsophoroside)-5-glucoside	Anthocyanidin_84
83	Peonidin-3-O-galactoside	Anthocyanidin_85
84	Peonidin-3-O-glucoside	Anthocyanidin_86
85	Peonidin-3-O-P-hydroxybenzoylsophoroside-5-glucoside	Anthocyanidin_87
86	Peonidin-3-O-rutinoside	Anthocyanidin_88

Table 1: List of compounds in the panel

Number	Compounds	Index
87	Peonidin-3-O-sambubioside	Anthocyanidin_89
88	Peonidin-3-O-sophorose	Anthocyanidin_90
89	Peonidin-3-O-sambubioside-5-O-glucoside	Anthocyanidin_91
90	Petunidin-3,5-O-diglucoside	Anthocyanidin_92
91	Petunidin-3-O-(6-O-malonyl-beta-D-glucoside)	Anthocyanidin_93
92	Petunidin-3-O-(6-O-p-coumaroyl)-glucoside	Anthocyanidin_94
93	Petunidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	Anthocyanidin_95
94	Petunidin-3-O-arabinoside	Anthocyanidin_96
95	Petunidin-3-O-galactoside	Anthocyanidin_97
96	Petunidin-3-O-glucoside	Anthocyanidin_98
97	Petunidin-3-O-rutinoside	Anthocyanidin_99
98	Petunidin-3-O-sambubioside	Anthocyanidin_100
99	Petunidin-3-O-sophorose	Anthocyanidin_101
100	Petunidin-3-O-sambubioside-5-O-glucoside	Anthocyanidin_102
101	Procyanidin A1	Anthocyanidin_103
102	Procyanidin A2	Anthocyanidin_104
103	Procyanidin B1	Anthocyanidin_105
104	Procyanidin B2	Anthocyanidin_106
105	Procyanidin B3	Anthocyanidin_107
106	Procyanidin C1	Anthocyanidin_108

Original file path: Final report/data/metware_component.xlsx

2.1 Sample information

This project has 24 samples divided into 4 groups. Sample information is shown in the following table:

Table 2: Sample information table

Species	Tissues	MW_ID	Sample_ID
-	-	A1	A1
-	-	A2	A2
-	-	A3	A3
-	-	A4	A4
-	-	A5	A5
-	-	A6	A6
-	-	B1	B1
-	-	B2	B2
-	-	B3	B3
-	-	B4	B4
-	-	B5	B5
-	-	B6	B6
-	-	C1	C1
-	-	C2	C2
-	-	C3	C3
-	-	C4	C4

Table 2: Sample information table

Species	Tissues	MW_ID	Sample_ID
-	-	C5	C5
-	-	C6	C6
-	-	D1	D1
-	-	D2	D2
-	-	D3	D3
-	-	D4	D4
-	-	D5	D5
-	-	D6	D6

Original file path: Final report/0.data/sample_info.xlsx

2.2 Reagents and instruments

Table 3: Instrument information

Instrument	Model	Manufacturer
LC-MS/MS	QTRAP 6500+	SCIEX
Centrifuge	5424R	Eppendorf
Electronic balance	AS 60/220.R2	RADWAG
Ball mill instrument	MM400	Retsch
Multitube vortex oscillator	MIX-200	ShangHaiJingXin
Ultrasonic cleaning apparatus	KQ5200E	KunShanShuMei

Table 4: Information of standards and reagents

Reagent	level	brand
Methanol	Chromatographically pure	Merck
Acetonitrile	Chromatographically pure	Merck
Formic acid	Chromatographically pure	Sigma-Aldrich
Hydrochloric acid	Premium pure	Xinyang Chemical Reagent Factory
Standard	Chromatographically pure	ExtraSynthese

2.3 Sample extraction process

The sample was freeze-dried, ground into powder (30 Hz, 1.5 min), and stored at -80°C until needed. 50 mg powder was weighted and extracted with 0.5 mL methanol/water/hydrochloric acid (500:500:1, V/V/V). Then the extract was vortexed for 5 min and ultrasound for 5 min and centrifuged at 12, 000 g under 4 °C for 3 min. The residue was re-extracted by repeating the above steps again under the same conditions. The supernatants were collected, and filtrated through a membrane filter (0.22 µm, Anpel) before LC-MS/MS analysis.

2.4 Chromatography-mass spectrometry acquisition conditions

The sample extracts were analyzed using an UPLC-ESI-MS/MS system (UPLC, ExionLC™ AD, <https://sciex.com.cn/> ; MS, Applied Biosystems 6500 Triple Quadrupole, <https://sciex.com.cn/>).

UPLC Conditions were as follows:

The analytical conditions were as follows, UPLC: column, Waters ACQUITY BEH C18 (1.7 μ m, 2.1 mm*100 mm); solvent system, water (0.1% formic acid): methanol (0.1% formic acid); gradient program, 95:5 V/V at 0 min, 50:50 V/V at 6 min, 5:95 V/V at 12 min, hold for 2 min, 95:5 V/V at 14 min; hold for 2 min; flow rate, 0.35 mL/min; temperature, 40°C; injection volume, 2 μ L.

The mass spectrum conditions were as follows:

Linear ion trap (LIT) and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (QTRAP), QTRAP® 6500+ LC-MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in positive ion mode and controlled by Analyst 1.6.3 software (Sciex). The ESI source operation parameters were as follows: ion source, ESI+; source temperature 550 °C; ion spray voltage (IS) 5500 V; curtain gas (CUR) was set at 35 psi, respectively. Anthocyanins were analyzed using scheduled multiple reaction monitoring (MRM). Data acquisitions were performed using Analyst 1.6.3 software (Sciex). MultiQuant 3.0.3 software (Sciex) was used to quantify all metabolites. Mass spectrometer parameters including the declustering potentials (DP) and collision energies (CE) for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period.

2.5 Qualitative and quantitative principles of metabolites

Metabolites were quantified by multiple reaction monitoring (MRM) using triple quadrupole mass spectrometry. In MRM mode, the first quadrupole screened the precursor ions for the target substance and excluded ions of other molecular weights. After ionization induced by the impact chamber, the precursor ions were fragmented, and a characteristic fragment ion was selected through the third quadrupole to exclude the interference of non-target ions. After obtaining the metabolite spectrum data from different samples, the peak area was calculated on the mass spectrum peaks of all substances and analyzed by standard curves.

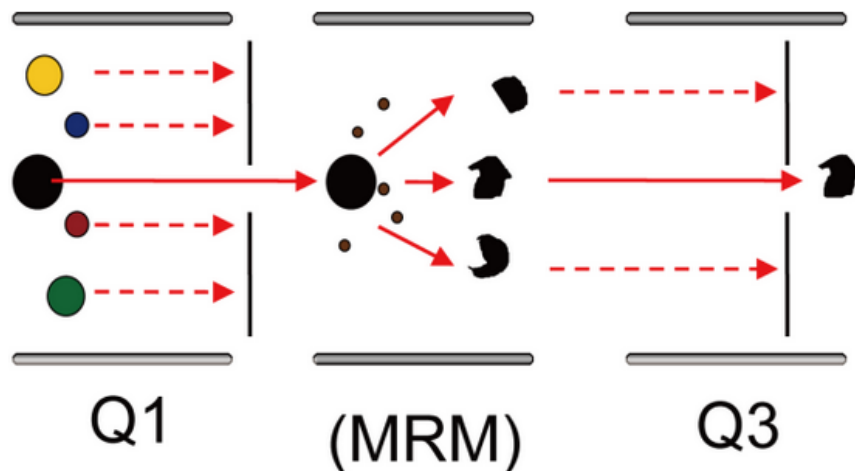


Fig 2:
Schematic diagram of multiple reaction monitoring mode by mass spectrometry

3 Data evaluation

3.1 Data pre-processing

Analyst 1.6.3 was used to process mass spectrum data. The following figure shows the total ions current (TIC) and MRM metabolite detection multi-peak diagram (XIC) of the mixed QC samples. The X-axis shows the Retention time (RT) from metabolite detection, and the Y-axis shows the ion flow intensity from ion detection (intensity unit: CPS, count per second).

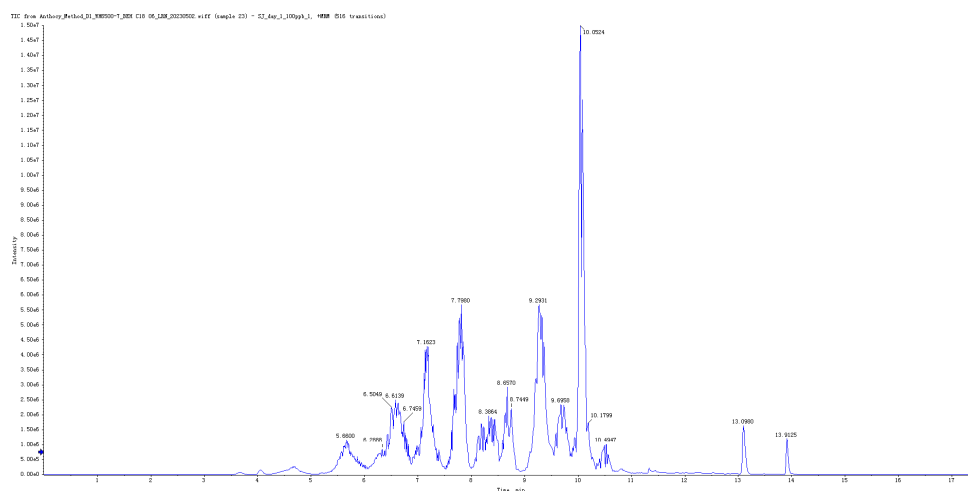


Fig 3: Total ion current diagram of mixed phase mass spectrum analysis

Original file path: Final report/0.data/QC/*QC_MS_TIC.png

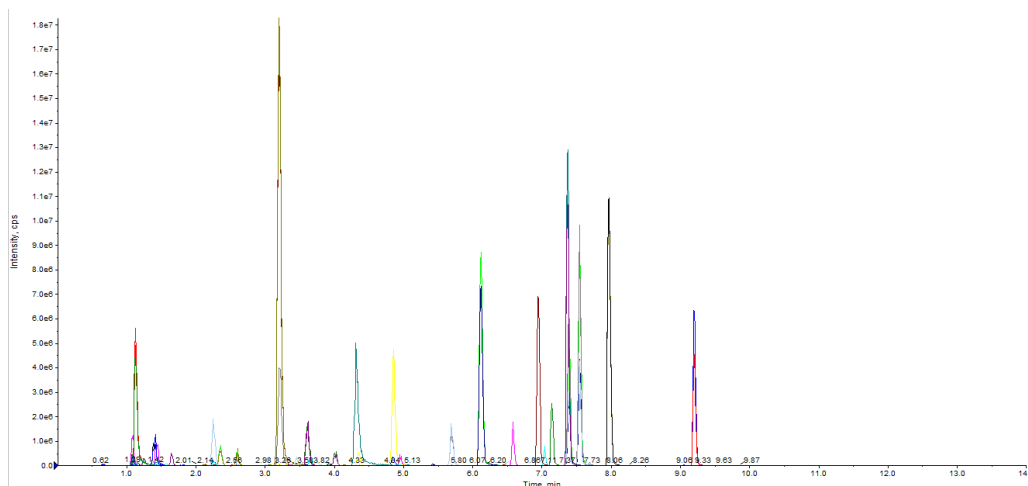


Fig 4: Extraction ion flow chromatogram

Original file path: Final report/0.data/QC/*MRM_detection_of_multimodal_maps*

The mass spectrometry data was analyzed using MultiQuant 3.0.3 software. The mass spectrum peaks detected in different samples were scored and corrected based on retention time and peak shape of the standard. The figure below shows the correction results of quantitative analysis of a substance randomly selected from different samples.

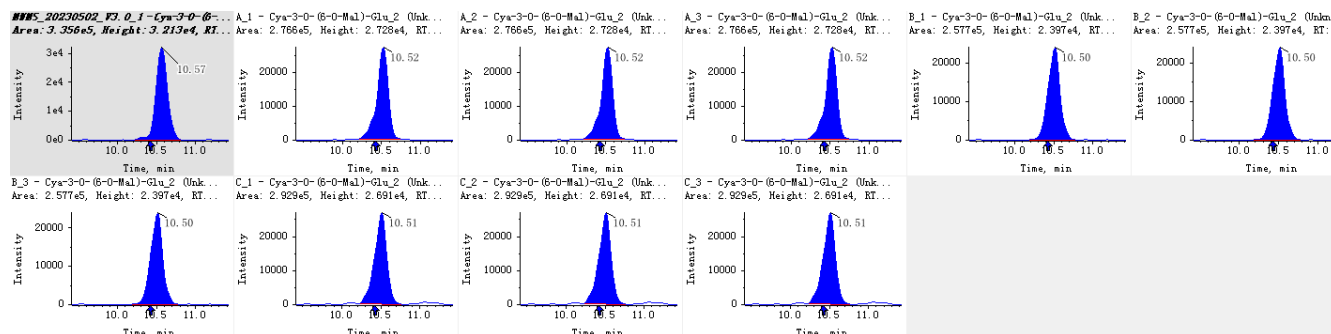


Fig 5: Scoring correction diagram for quantitative analysis of metabolites

Note: The figure shows the quantitative analysis integral correction results of randomly selected metabolites in different samples. The x-axis is the retention time (min) of metabolite detection, the y-axis is the ion flow intensity (CPS) of a certain metabolite ion detection, and the peak area represents the relative content of the substance in the sample.

Original file path: Final report/0.data/QC/*Integral_correction.png

3.2 Standard Solution Preparation

Prepare 0.01 ng/mL, 0.05 ng/mL, 0.1 ng/mL, 0.5 ng/mL, 1 ng/mL, 5 ng/mL, 10 ng/mL, 50 ng/mL, 100 ng/mL, 500 ng/mL, 1000 ng/mL, 2000 ng/mL, 5000 ng/mL a series of different concentrations standard so-

lutions, and obtain the mass spectral peak intensity data and calibration curve. The x-axis is the standard concentration, the y-axis is the standard area. The equation of calibration curve are shown in the following table.

Table 5: Equation of calibration curve

Index	Class	RT	Equation
Anthocyanidin_15	Cyanidin	-	$y = 7.20269e4 x - 389.25100$
Anthocyanidin_14	Cyanidin	-	$y = 9.86888e4 x + 2506.45717$
Anthocyanidin_08	Cyanidin	-	$y = 1.48554e5 x + 24914.02769$
Anthocyanidin_05	Cyanidin	-	$y = 7.07968e4 x + 5734.36764$
Anthocyanidin_17	Cyanidin	-	$y = 26702.89565 x - 1503.99152$
Anthocyanidin_04	Cyanidin	-	$y = 6.77465e4 x - 11707.89877$
Anthocyanidin_10	Cyanidin	-	$y = 7.51123e4 x + 13501.61290$
Anthocyanidin_16	Cyanidin	-	$y = 9.23008e4 x - 919.23969$
Anthocyanidin_01	Cyanidin	-	$y = 3.25688e4 x - 12731.31621$
Anthocyanidin_02	Cyanidin	-	$y = 3.25688e4 x - 12731.31621$

Original file path: Final report/0.data/equation.xlsx

3.3 Quantification Results

Substitute the integrated peak area of all the detected samples into the equation of calibration curve for calculation, and further substitute into the calculation formula for calculation, and finally obtain concentration of the sample.

Concentration of the sample ($\mu\text{g/g}$) = $c \cdot V / 1000000 / m$

Meaning of each letter in the formula:

c (ng/mL) is the concentration obtained by substituting the sample peak area into the equation of calibration curve.

V (μL) is the volume of extraction solution.

m (g) is the quality of the sample.

The metabolite ID, concentration and corresponding metabolite names of some metabolites detected in this experiment are shown in the following table:

Table 6: Statistical Table of metabolite quantity

Index	A1	A2
Anthocyanidin_01	0.0338851	0.0337765
Anthocyanidin_17	0.0177332	0.0148256
Anthocyanidin_16	0.0872831	0.0719825
Anthocyanidin_14	0.0416203	0.0413638
Anthocyanidin_13	0.0758341	0.07174
Anthocyanidin_10	0.0697135	0.0689639
Anthocyanidin_09	0.0198515	0.0198012
Anthocyanidin_15	0.0202636	0.0205767
Anthocyanidin_07	0.0467777	0.0360037
Anthocyanidin_06	0.02598	0.0203927

Original file path: Final report/0.data/*level.xlsx

3.4 Sample Quality Control Analysis

3.4.1 Total Ion Chromatogram Analysis

Using the mixed solution as the QC sample, one QC sample was inserted every 10 detection samples for analysis during the detection by the system. The stability of the device during the detection of the project can be assessed by analyzing the overlapped total ion flow chromatograms (TICs) obtained from the mass spectrometry detection and analysis of the same QC samples. The high stability of the testing device is a vital safeguard for the reproducibility and reliability of the data.

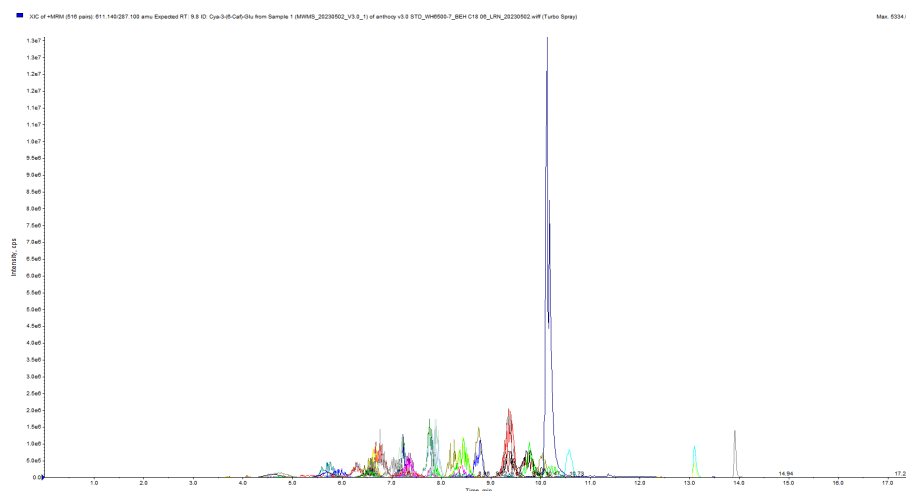


Fig 6: TIC overlap diagram detected by QC sample essence spectrum

Note: Superimposed spectrum from different QC samples. The results showed that the spectrum of total ion flow were highly consistent indicating that the signal stability was good when the same sample was detected at different times by mass spectrometry.

Original file path: Final report/0.data/picture/*QC_MS_tic_overlap*

3.4.2 CV value distribution of all samples

The Coefficient of Variation (CV) value is the ratio between the standard deviation of the original data and the mean of the original data, which can reflect the degree of data dispersion. The Empirical Cumulative Distribution Function (ECDF) can be used to analyze the frequency of CV of substances that is smaller than the reference value. The higher the proportion of substances with low CV value in QC samples is, the more stable the experimental data is. The proportion of substances with CV value less than 0.3 in QC samples was higher than 80% , indicating that the experimental data were relatively stable. The proportion of substances with CV value less than 0.2 in QC samples was higher than 80%, indicating that the experimental data were very stable.

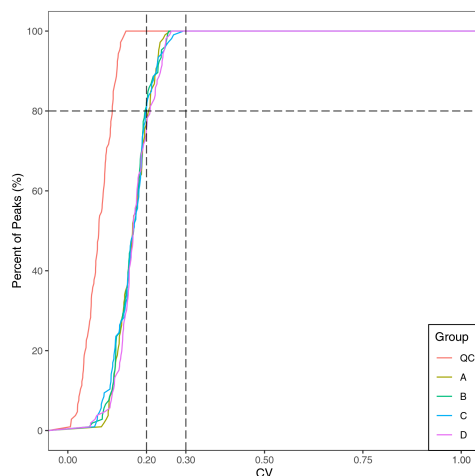


Fig 7: CV distribution of each group

Note: The X-axis represents the CV value, the Y-axis represents the proportion of metabolites with CV value less than a corresponding reference value. Different colors represent different sample groups. QC indicates quality control samples. The two dash lines on X-axis correspond to 0.2 and 0.3; the two dash lines on Y-axis correspond to 80% .

Original file path: Final report/1.Data_Assess/CV/*ECDF*

3.5 Sample quantification histogram

The results of sample content are grouped by statistics, and the statistical results are shown in the following table.

Table 7: Statistical results table

Index	Group	N	Mean
Anthocyanidin_01	D	6	0.094095
Anthocyanidin_01	C	6	0.071224
Anthocyanidin_01	B	6	0.026301
Anthocyanidin_01	A	6	0.037508
Anthocyanidin_02	D	6	0.068214
Anthocyanidin_02	C	6	0.071139
Anthocyanidin_02	B	6	0.058028
Anthocyanidin_02	A	6	0.051672
Anthocyanidin_03	D	6	0.04595
Anthocyanidin_03	C	6	0.037382

Original file path: Final report/1.Data_Assess/histogram/groups*.xlsx

The bar chart below shows the content difference of each substance in different groups.

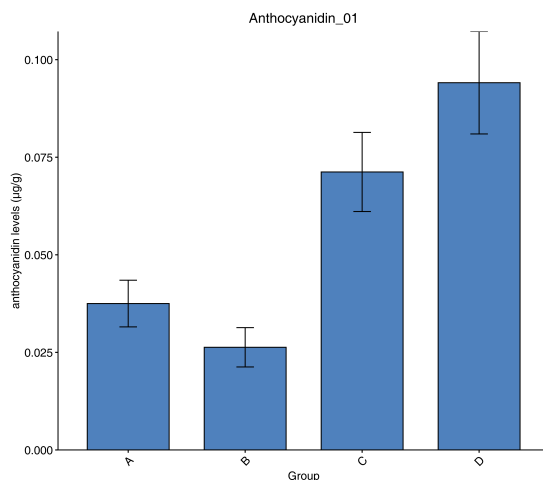


Fig 8: Sample content histogram

Note: The x-axis is the groups, the y-axis is the content, error bars are standard deviations.

Original file path: Final report/1.Data_Assess/histogram/histogram_compounds/*.png

3.6 Hierarchical Cluster Analysis

3.6.1 Principles of cluster analysis

Hierarchical Cluster Analysis (HCA) is a type of multivariate statistical analysis method. The samples are classified according to their features such that highest homogeneity is achieved between sample from the same group and highest heterogeneity is achieved between samples from different groups. In this report, the compound quantification data was normalized (Unit Variance Scaling, UV Scaling) and heatmaps were drawn by R software Pheatmap package. Hierarchical Cluster Analysis (HCA) was used to cluster the samples.

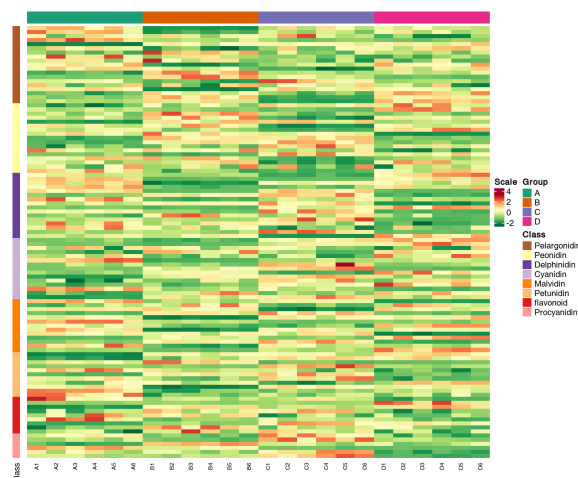


Fig 9: Sample clustering diagram

Note: X-axis indicates the sample name and the Y-axis are the metabolites. Group indicates sample groups. Z-Score indicates the relative quantification of each metabolite with red representing higher content and green representing lower content. Cluster analysis was performed on both metabolites (vertical cluster tree) and samples (horizontal cluster tree). “all_heatmap_class”: Heat map based on metabolite classification; “all_heatmap_no_cluster”: Showing only heatmap.

Original file path: Final report /1.Data_Assess/*all_heatmap*

4 Analysis results

4.1 Differential metabolite screening

It is often necessary to combine univariate statistical analysis and multivariate statistical analysis for large high dimensional datasets such as metabolomics datasets to accurately identify differential metabolites. Univariate statistical analysis methods include parametric test and nonparametric test. Multivariate statistical analysis methods include principal component analysis and partial least square discriminant analysis. Based on the results of OPLS-DA (biological repetition ≥ 2), multivariate analysis of Variable Importance in Projection (VIP) from OPLS-DA modeling was used to preliminarily select differential metabolites from different samples. The fold-change and statistical significance (p-value) from univariate analysis can be used in conjunction to further identify differential metabolites. If biological replicates were < 3 , differential metabolites are screened based on Fold Change value. If there were ≥ 3 biological replicates, VIP and P-values were used in combination to screen for differential metabolites. The detailed screening criteria is as follows:

For two sets of comparisons:

1. Metabolites with Fold Change ≥ 2 and Fold Change ≤ 0.5 were considered as significant and selected.

A partial result from the screening criteria is seen below:

Table 8: Screening results of differential metabolites

Index	Compounds	Type
Anthocyanidin_14	Cyanidin-3-O-sambubioside	down
Anthocyanidin_10	Cyanidin-3-O-galactoside	up
Anthocyanidin_15	Cyanidin-3-O-sophoroside	down
Anthocyanidin_06	Cyanidin-3-O-(6-O-p-coumaroyl)-glucoside	down
Anthocyanidin_05	Cyanidin-3-O-(6-O-malonyl-beta-D-glucoside)	down
Anthocyanidin_08	Cyanidin-3-O-arabinoside	down
Anthocyanidin_24	Delphinidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	up
Anthocyanidin_25	Delphinidin-3-O-arabinoside	up
Anthocyanidin_26	Delphinidin-3-O-galactoside	up
Anthocyanidin_28	Delphinidin-3-O-rhamnoside	up

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/group-ID*_vs_group-ID*filter.xlsx.

4.1.1 Bar chart of differential metabolites

The following figure shows the result of top differentially expressed metabolites in each comparison with fold-change value shown as $\log_2$ values .

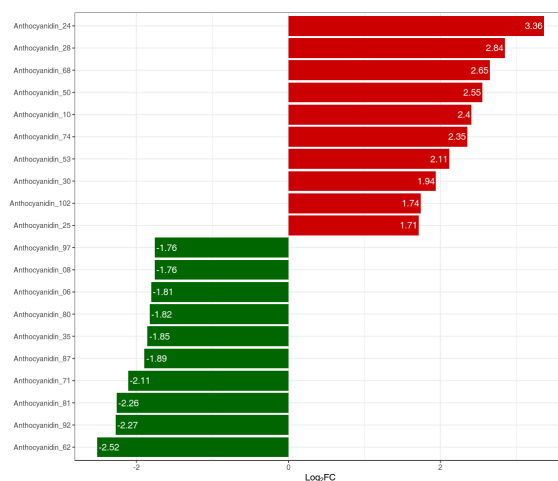


Fig 10: Bar chart of differential metabolites

Note: X-axis refers to $\log_2$ FC values of top differential metabolites,the Y-axis refers to metabolites. Red bars represent up-regulated differential metabolites and green bars represent down-regulated differential metabolites.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/TopFcMetabolites/group-ID*_vs_group-ID*_TopFcMetabolites.*

4.1.2 Violin plot of differential metabolites

Violin plot is used to display data distribution and its probability density. The box in the middle represents the interquartile range, and the middle box represents the 95% confidence interval. The black horizontal line

is the median, and the outer shape represents the distribution density of the data. The following figure shows the result of top 50 differentially compounds with the largest Log₂FC value.

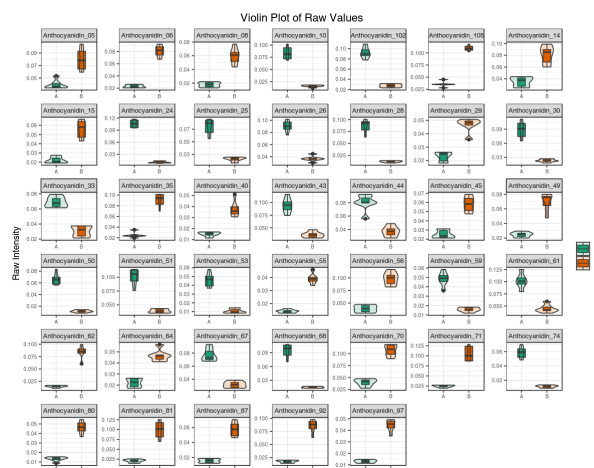


Fig 11: Violin plot of differential metabolites

Note: X-axis refers to sample,the Y-axis refers to content.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/fullViolin/group-ID*_vs_group-ID*_fullViolin_Raw.*;

4.1.3 Differential metabolite statistics

The number of different metabolites in each group is shown in the table below:

Table 9: Statistical table of differential metabolites

group name	All sig diff	down regulated	up regulated
A_vs_B	40	22	18

Statistical table of differential metabolites:Final report/2.Basic_analysis/Difference_analysis/sigMetabolitesCount.xls

4.2 Functional annotation and enrichment analysis of differential metabolites in KEGG database

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database that integrates compounds and genes into metabolic pathways. The KEGG database enabled researchers to study genes with their expression information and compounds with their abundances as a complete network.

4.2.1 Functional annotation of differential metabolites

Metabolites are annotated using the KEGG database, and only metabolic pathways containing differential metabolites are shown. Detailed results are found in the attached results. A portion of the results is shown

below:

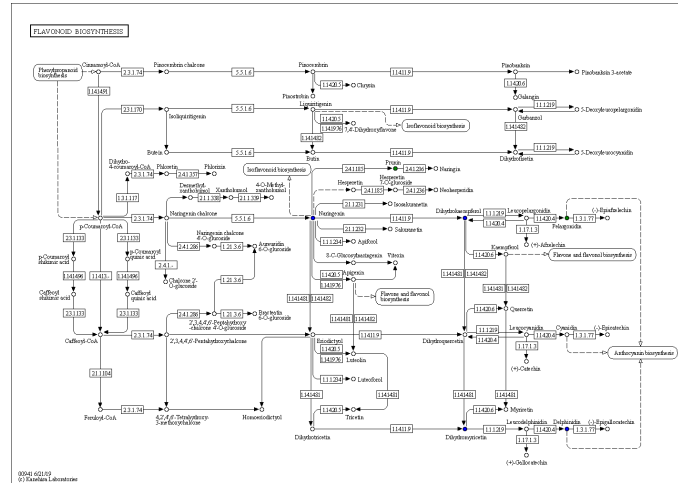


Fig 12: KEGG pathway of metabolites

Note: Red circles indicate that the metabolite content was significantly up-regulated in the experimental group; the blue circles indicate that the metabolite content was detected but did not change significantly; Green circles indicate that the metabolite content was significantly down-regulated in the experimental group. The orange circles indicate a mixture of both up-regulated and down-regulated metabolites. This allows searching for metabolites that may contribute to the phenotypic differences.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/Graph/ko*.

Statistical analysis of KEGG database annotation of screened metabolites with significant differences.

Some of the results are as follows:

Table 10: KEGG annotations for differential metabolites

Index	Compounds	Type	cpd_ID
Anthocyanidin_14	Cyanidin-3-O-sambubioside	down	C20490
Anthocyanidin_10	Cyanidin-3-O-galactoside	up	C08647
Anthocyanidin_15	Cyanidin-3-O-sophoroside	down	C16306
Anthocyanidin_06	Cyanidin-3-O-(6-O-p-coumaroyl)-glucoside	down	C12095
Anthocyanidin_05	Cyanidin-3-O-(6-O-malonyl-beta-D-glucoside)	down	C12643
Anthocyanidin_08	Cyanidin-3-O-arabinoside	down	-
Anthocyanidin_24	Delphinidin-3-O-5-O-(6-O-coumaroyl)-diglucoside	up	C16351
Anthocyanidin_25	Delphinidin-3-O-arabinoside	up	-
Anthocyanidin_26	Delphinidin-3-O-galactoside	up	-
Anthocyanidin_28	Delphinidin-3-O-rhamnoside	up	-

Table 11: Enrichment Statistics of KEGG annotations for differential metabolites

ko_ID	Sig_compound	compound	Sig_compound_all	compound_all
ko00942	18	34	19	42
ko01110	3	12	19	42
ko01100	3	7	19	42
ko00941	2	6	19	42

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_filter_kegg.xlsx.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_KEGG.xlsx.

4.2.2 KEGG classification of differential metabolites

The significant differential metabolites were classified based on pathway annotation . The results are as follows:

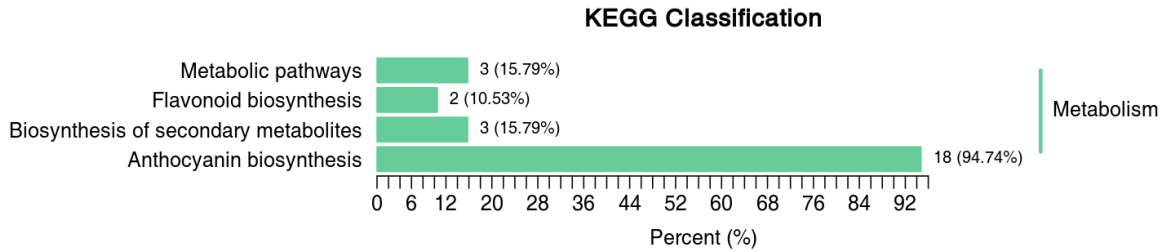


Fig 13: KEGG classification of differential metabolites

Note: the Y-axis shows the name of the KEGG pathway. The number of metabolites and the proportion of the total metabolites are shown next to the bar plot.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_KEGG_barplot.*.

4.2.3 KEGG enrichment analysis of differential metabolites

KEGG pathway enrichment analysis was conducted based on the annotation results. We calculated the Rich Factor for each pathway, which is the ratio of the number of differential metabolites in the corresponding pathway to the total number of metabolites annotated in the same pathway. The greater the Rich Factor, the greater the degree of enrichment. P-value is the calculated using hypergeometric test as shown below:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

N represents the total number metabolites with KEGG annotation, n represents the number of differential

metabolites in N, M represents the number of metabolites in a KEGG pathway in N, and m represents the number of differential metabolites in a KEGG pathway in M. The closer the p-value to 0, the more significant the enrichment. The size of the dots in the figure represents the number of significantly different metabolites enriched in the corresponding pathway. The results are shown below:

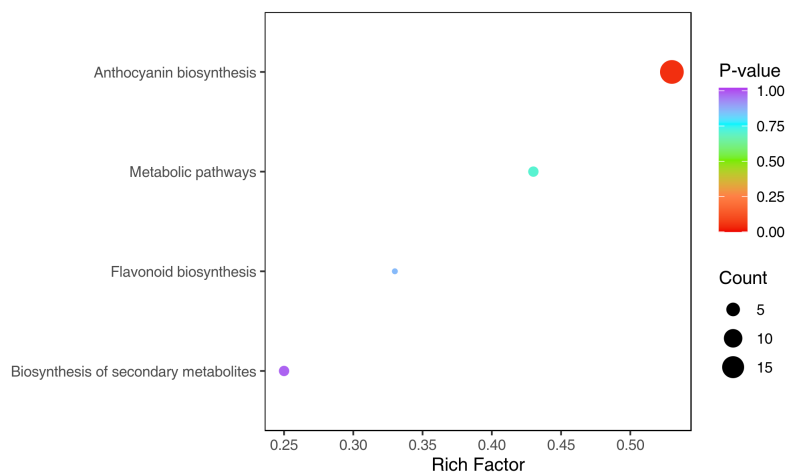


Fig 14: KEGG enrichment diagram of differential metabolites

Note: The X-axis represents the Rich Factor and the Y-axis represents the pathway. The color of points reflects the p-value. The darker the red, the more significant the enrichment. The size of the dot represents the number of enriched differential metabolites.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_KEGG_Enrichment.*.

5 References

1. Paola-Naranjo R D D , José Sánchez-Sánchez, Ana María González-Paramás, et al. Liquid chromatographic-mass spectrometric analysis of anthocyanin composition of dark blue bee pollen from *Echium plantagineum*. *Journal of Chromatography A*, 2004, 1054(1-2):205-210.
2. Alexander Acevedo De la Cruz, Ghislaine Hilbert, Céline Rivière, et al. Anthocyanin identification and composition of wild *Vitis* spp. accessions by using LC-MS and LC-NMR. *Analytica Chimica Acta*, 2012, 732:145-152.
3. De Ferrars R M , Czank C , Saha S , et al. Methods for Isolating, Identifying, and Quantifying Anthocyanin Metabolites in Clinical Samples. *Analytical Chemistry*, 2014, 86(20):10052-8.
4. Qinlong Z, Suize Y, Dongchang Z , et al. Development of “Purple Endosperm Rice” by Engineering Anthocyanin Biosynthesis in the Endosperm with a High-Efficiency Transgene Stacking System. *Molecular Plant*, 2017.

6 Appendix

6.1 Analytical methods

1. Hierarchical Cluster Analysis

The HCA (hierarchical cluster analysis) results of samples and metabolites were presented as heatmaps with dendrograms. HCA was carried out by R package pheatmap. For HCA, normalized signal intensities of metabolites (unit variance scaling) are visualized as a color spectrum.

2. Differential metabolites selected

Significantly regulated metabolites between groups were determined by absolute Log₂FC (fold change).

3. KEGG annotation and enrichment analysis

Identified metabolites were annotated using KEGG compound database (<http://www.kegg.jp/kegg/compound/>), annotated metabolites were then mapped to KEGG Pathway database (<http://www.kegg.jp/kegg/pathway.html>). Pathways with significantly regulated metabolites mapped to were then fed into MSEA (metabolite sets enrichment analysis), their significance was determined by hypergeometric test's p-values.

6.2 List of software and versions

Data processing methods were mainly adopted in the analysis process in two ways:

(1) unit variance scaling (UV)

Unit variance Scaling (UV) is also called Z-Score standardization, i.e., auto scaling. This method standardizes data according to mean and standard deviation of original data. The processed data conform to the standard normal distribution, that is, the mean value is 0 and the standard deviation is 1.

Calculation method: Divide the original data center by standard deviation.

The formula is as follows:

$$x' = \frac{x - \mu}{\sigma}$$

Where μ is the mean value and σ is the standard deviation.

(2) Centralization/zero-mean-centered (Ctr)

Calculation method: subtract the mean of the variables from the original data.

The formula is as follows:

$$x' = x - \mu$$