

A metabolomic strategy by integrating headspace gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry to differentiate the five cultivars of *Chrysanthemum* flower

Lin-Ning Zhang,^a Long Wang,^a Zi-Qi Shi^{*b,c}, Ping Li^a and Hui-Jun Li^{*a}

^aState Key Laboratory of Natural Medicines, China Pharmaceutical University, No. 24 Tongjia Lane, Nanjing, China

^bAffiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, China

^cJiangsu Province Academy of Traditional Chinese Medicine, Nanjing, China

(*) Corresponding authors:

Prof. Shi, Key Laboratory of New Drug Delivery Systems of Chinese Materia Medica, Jiangsu Provincial Academy of Chinese Medicine, Nanjing 210028, China

Tel.: +86 25 85608672, fax: +86 25 85608672, E-mail: shiziqi@126.com

and

Prof. Li, State Key Laboratory of Natural Medicines, China Pharmaceutical University, No. 24 Tongjia Lane, Nanjing, China

Tel.: +86 25 8327 1379, fax: +86 25 8327 1379, E-mail: cpuli@163.com

Fig. S1 Flowchart of the general strategy to differentiate and classify the five official *Chrysanthemum* flowers

Fig. S2 Collected regions and the number of samples of *Chrysanthemum* flowers

Fig. S3 The overlapping total ion current (TIC) chromatograms of QC samples from HSGC-MS (A) and UHPLC-QTOF/MS (B) analysis

Fig. S4 The typical TIC chromatograms of BJ, CJ, GJ, HbJ and HJ by HSGC-MS.

Fig. S5 The typical TIC chromatograms of BJ, CJ, GJ, HbJ and HJ by UHPLC-QTOF-MS in negative ion mode.

Fig. S6 PCA scores plots of data of HSGC-MS and UHPLC-QTOF/MS from different normalization methods. Obviously, there were no obvious separation trends for groups of five FCs. A and B are plots of HSGC-MS and UHPLC-QTOF/MS data sets normalized by respective IS, C and D are plots of HSGC-MS and UHPLC-QTOF/MS normalized by total peak area.

Fig. S7 The plot of response permutation testing, $R^2=(0.0, 0.406)$, $Q^2=(0.0, -0.290)$
The criterions of validity have two important characteristics: the blue regression line of the Q^2 (cum) points has a negative intercept, and all the permuted values (Q^2 (cum) and R^2 (cum)) to the right are higher than the original points to the left.

Fig. S8 ROC analysis of the identified marker compounds in similar species of *Chrysanthemum* flowers with BJ-HJ, CJ-HbJ, GJ-HbJ and CJ-GJ.

Table S1 Nine extracted ions in ten QC samples from HSGC-MS analysis

Table S2 Nine extracted ions in eleven QC samples from UHPLC-QTOF/MS analysis

Table S3 IS-1 in ten QC samples from HSGC-MS analysis

Table S4 IS-2 in eleven QC samples from UHPLC-QTOF/MS analysis

Table S5 ROC curve analysis of the identified marker compounds from CJ and GJ

Table S6 ROC curve analysis of the identified marker compounds from CJ and HbJ

Table S7 ROC curve analysis of the identified marker compounds from GJ and HbJ

Table S8 ROC curve analysis of the identified marker compounds from HJ and BJ

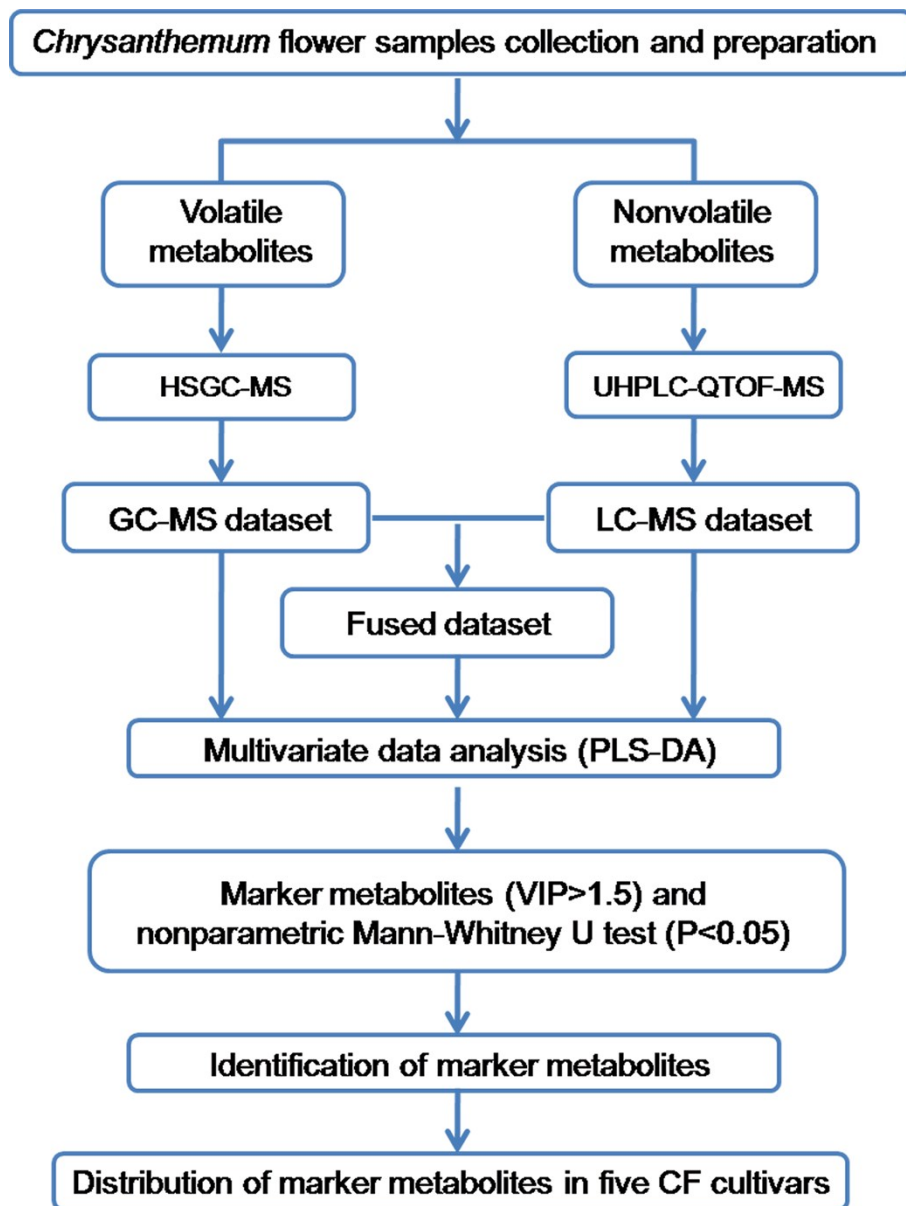
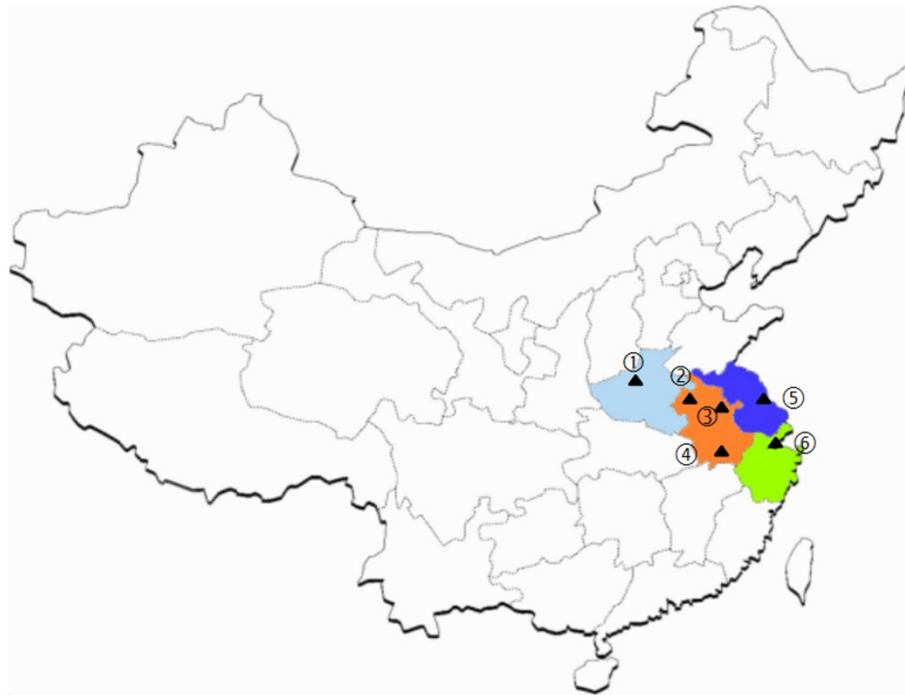


Fig. S1 Flowchart of the general strategy to differentiate and classify the five official *Chrysanthemum* flowers



① Jiaozuo, Henan—"HJ" ② Bozhou, Anhui—"BJ" ③ Chuzhou, Anhui—"CJ" ④ Huangshan, Anhui—"GJ"
 ⑤ Yancheng, Jiangsu—"HbJ" ⑥ Tongxiang, Zhejiang—"HbJ"

Materials	Sample number	Source
HuaiJu (HJ)	11	Jiaozuo, Henan
BoJu (BJ)	10	Bozhou, Anhui
ChuJu (CJ)	11	Chuzhou, Anhui
GongJu (GJ)	10	Huangshan, Anhui
HangJu (HbJ)	11	Yancheng Jiangsu and Tongxiang, Zhejiang

Fig. S2 Collected regions and the number of samples of *Chrysanthemum* flowers

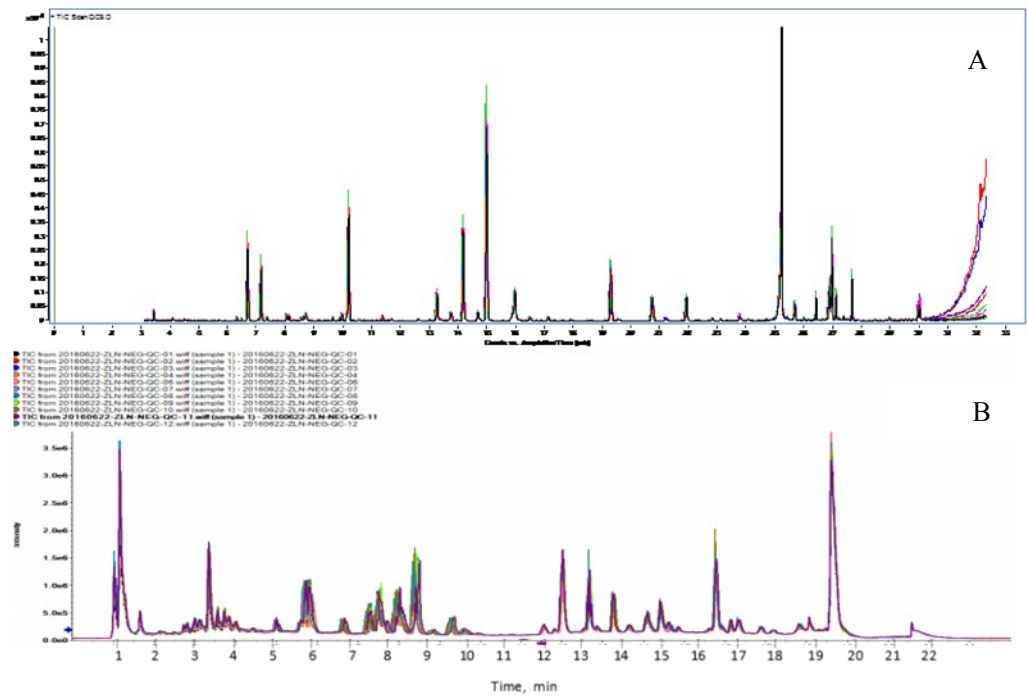


Fig S3 The overlapping total ion current (TIC) chromatograms of QC samples from HSGC-MS (A) and UHPLC-QTOF/MS (B) analysis

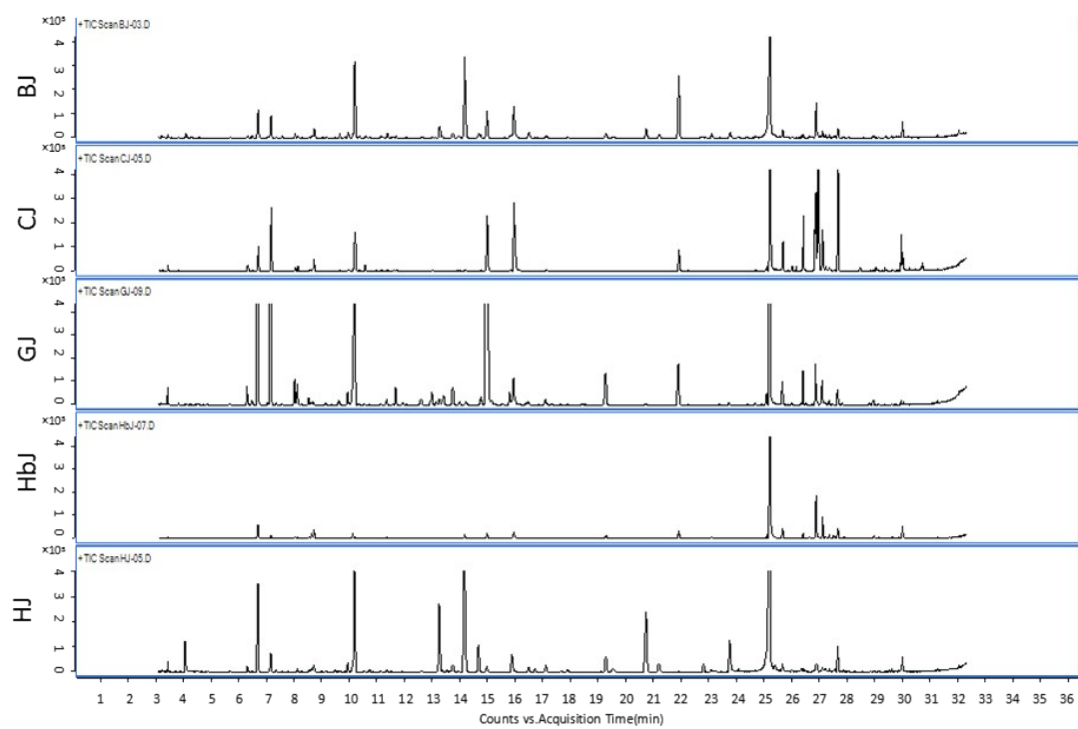


Fig. S4 The typical TIC chromatograms of BJ, CJ, GJ, HbJ and HJ by HSGC-MS

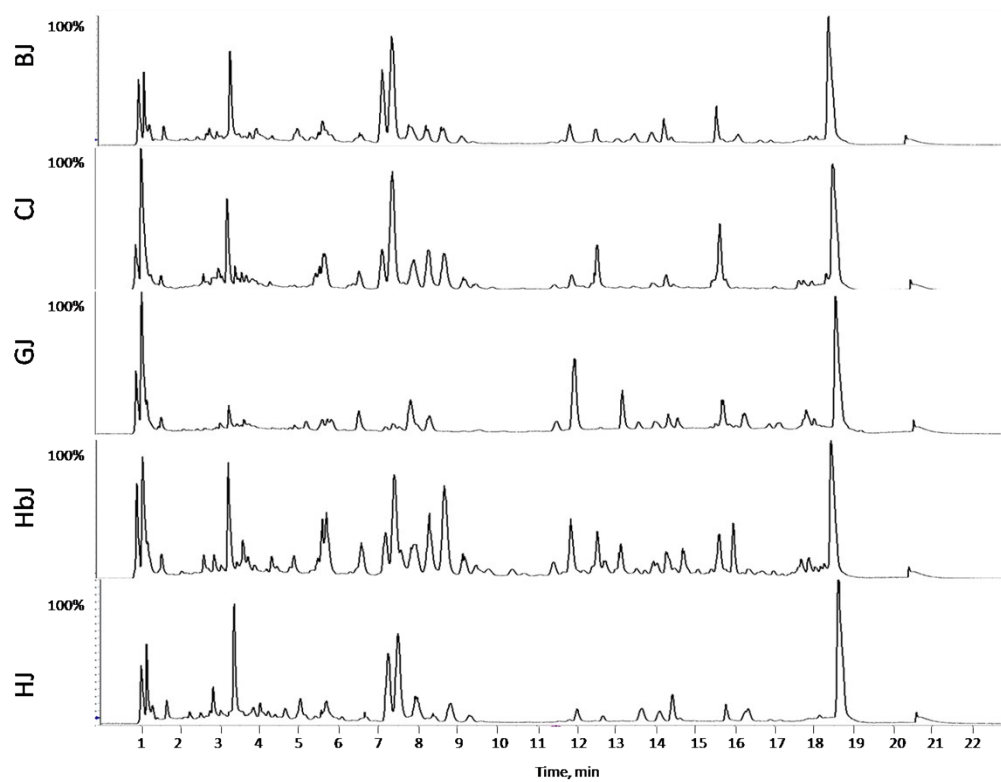


Fig. S5 The typical TIC chromatograms of BJ, CJ, GJ, HbJ and HJ by UHPLC-QTOF-MS in negative ion mode

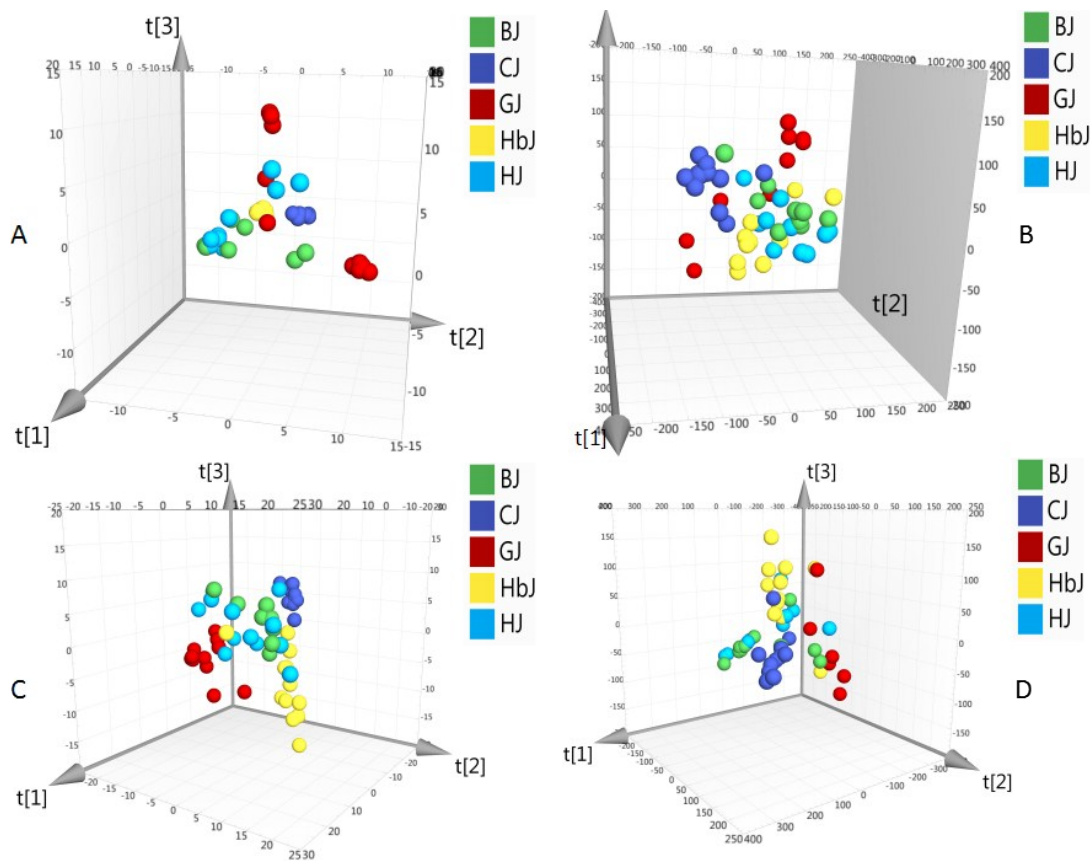


Fig. S6 PCA scores plots of data of HSGC-MS and UHPLC-QTOF/MS from different normalization methods. Obviously, there were no obvious separation trends for groups of five FCs. A and B are plots of HSGC-MS and UHPLC-QTOF/MS data sets normalized by respective IS, C and D are plots of HSGC-MS and UHPLC-QTOF/MS normalized by total peak area.

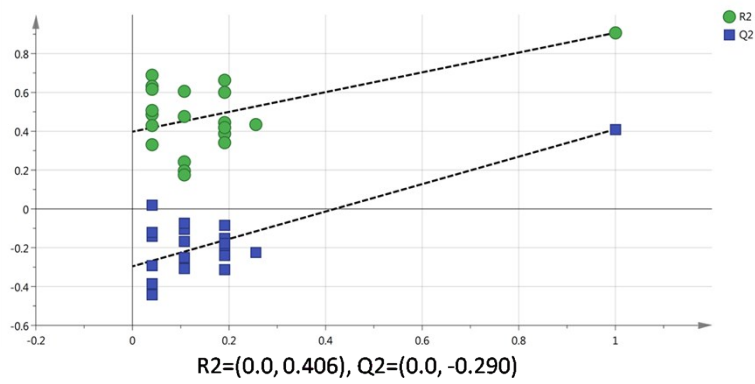


Fig. S7 The plot of response permutation testing, $R2=(0.0, 0.406)$, $Q2=(0.0, -0.290)$
 The criteria of validity have two important characteristics: the blue regression line of the Q2 (cum) points has a negative intercept, and all the permuted values (Q2 (cum) and R2 (cum)) to the right are higher than the original points to the left

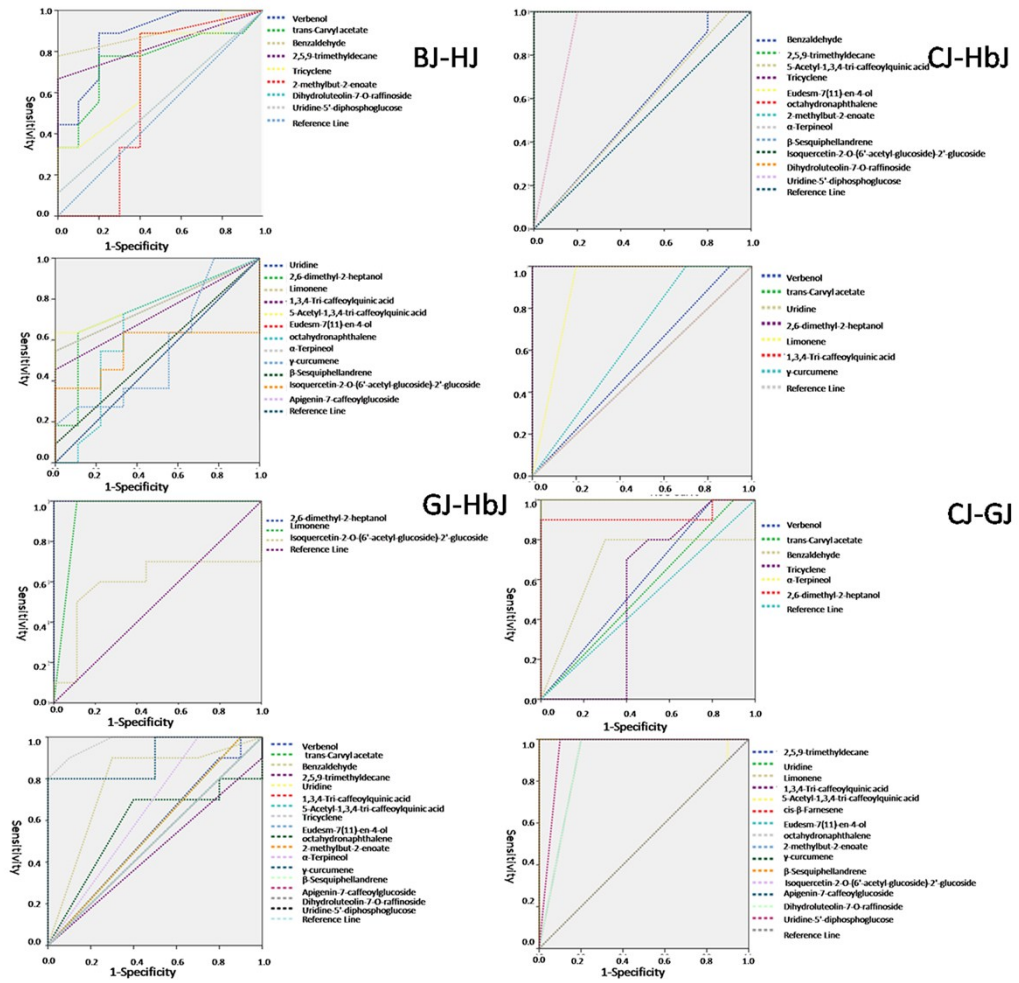


Fig. S8 ROC analysis of 21 biomarkers in similar species of *Chrysanthemum* flowers with BJ-HJ, CJ-HbJ, GJ-HbJ and CJ-GJ

Table S1 Nine extracted ions in ten QC samples from HSGC-MS analysis

Extracted ion (<i>m/z</i> @ <i>t_R</i>)	No.	<i>t_R</i> (min)	Peak area
93.0 @ 6.310	1	6.298	43964.16
	2	6.292	43824.39
	3	6.293	44756.61
	4	6.298	44869.55
	5	6.298	47211.60
	6	6.304	39326.86
	7	6.31	37705.17
	8	6.328	38925.68
	9	6.316	35655.54
	10	6.322	31911.78
	RSD (%)	0.1989	11.92
170.0 @ 8.630	1	8.624	60746.08
	2	8.618	58350.04
	3	8.618	55715.41
	4	8.618	60489.43
	5	8.624	63608.81
	6	8.636	52188.82
	7	8.636	52047.18
	8	8.653	54544.30
	9	8.647	47228.54
	10	8.647	47260.30
	RSD (%)	0.1563	10.20
184.0 @ 11.591	1	11.672	26202.92
	2	11.667	28933.25
	3	11.667	26839.84
	4	11.655	25924.77
	5	11.667	29324.15
	6	11.678	24665.22
	7	11.684	19952.83
	8	11.696	18652.77
	9	11.696	21167.37
	10	11.696	16260.51
	RSD (%)	0.1256	18.91
152.0 @ 15.875	1	15.933	587763.30
	2	15.927	596975.58

	3	15.927	636672.73
	4	15.933	629611.11
	5	15.927	645130.14
	6	15.939	591700.71
	7	15.945	533237.40
	8	15.968	561636.09
	9	15.957	577224.54
	10	15.951	465806.16
	RSD (%)	0.08921	9.209
<hr/>			
59.0 @ 17.105	1	17.087	62179.45
	2	17.081	62263.02
	3	17.081	66336.31
	4	17.087	64096.98
	5	17.093	71012.48
	6	17.099	59542.14
	7	17.105	48947.34
	8	17.122	52998.31
	9	17.122	55521.11
	10	17.105	43813.22
	RSD (%)	0.08924	14.19
<hr/>			
109.0 @ 22.811	1	22.805	32248.08
	2	22.794	37089.30
	3	22.794	35080.15
	4	22.794	36304.52
	5	22.794	37013.48
	6	22.8	28666.52
	7	22.805	24769.84
	8	22.823	27433.84
	9	22.811	28985.96
	10	22.805	24652.08
	RSD (%)	0.04151	15.85
<hr/>			
204.0 @ 26.827	1	26.856	470109.48
	2	26.856	489787.83
	3	26.856	518002.05
	4	26.856	470226.78
	5	26.862	502496.38
	6	26.862	429707.94
	7	26.868	403009.85
	8	26.874	423725.64
	9	26.874	416462.44

	10	26.874	339497.73
	RSD (%)	0.02987	12.11
204.0 @ 26.903	1	26.949	762982.70
	2	26.949	779225.64
	3	26.949	807477.85
	4	26.949	768627.68
	5	26.95	835865.37
	6	26.956	722203.99
	7	26.955	676027.72
	8	26.967	690816.30
	9	26.961	699411.46
	10	26.961	596450.59
	RSD (%)	0.02425	0.9671
204.0 @ 27.666	1	27.655	377574.43
	2	27.655	397247.97
	3	27.655	413888.28
	4	27.655	376448.09
	5	27.655	411370.23
	6	27.661	354501.63
	7	27.666	323594.54
	8	27.672	343883.30
	9	27.672	344385.39
	10	27.666	275931.11
	RSD (%)	0.02617	11.79

Table S2 Nine extracted ions in eleven QC samples from UHPLC-QTOF/MS

analysis

Extracted ion	No.	t_R	Peak area
277.0331	1	1.1558	15162605.83
	2	1.1619	16415434.34
	3	1.1543	17315510.79
	4	1.1546	16942635.35
	5	1.1580	16883041.17
	6	1.1554	17767263.02
	7	1.1558	20361786.25
	8	1.1562	18851267.17
	9	1.1573	19575234.07
	10	1.1553	19316328.54
	11	1.1611	18950580.34
	RSD (%)	0.2174	8.780
243.0604	1	1.6478	1809560.45
	2	1.6350	2086357.71
	3	1.6479	2033042.99
	4	1.6330	2007974.29
	5	1.6340	2089584.81
	6	1.6300	2070715.15
	7	1.6380	2832767.23
	8	1.6420	2577345.57
	9	1.6430	2667555.61
	10	1.6380	2412917.27
	11	1.6390	2076676.72
	RSD (%)	0.3563	14.53
667.1477	1	3.6270	1330756.44
	2	3.6570	1440613.31
	3	3.6470	1201413.10
	4	3.6570	1302689.58
	5	3.6410	1262582.24
	6	3.6380	1365451.25
	7	3.6460	1263921.57
	8	3.6450	1306795.21
	9	3.6680	1219223.86
	10	3.6320	1447656.33

	11	3.6290	1444435.75
	RSD (%)	0.3504	6.710
639.1177	1	3.8092	1396419.41
	2	3.8190	1251343.27
	3	3.8890	1078519.96
	4	3.8850	1165813.75
	5	3.8380	1230658.76
	6	3.8510	1597973.13
	7	3.8590	1710070.52
	8	3.8880	1689719.66
	9	3.8850	1750538.57
	10	3.8900	1784064.86
	11	3.8570	1666478.07
	RSD (%)	0.7596	17.73
338.0776	1	4.4089	297359.68
	2	4.4459	330481.21
	3	4.4463	255370.07
	4	4.4240	319886.52
	5	4.4515	271766.76
	6	4.4330	267939.73
	7	4.4347	303094.66
	8	4.4371	344778.95
	9	4.4445	383009.72
	10	4.4254	335967.50
	11	4.4100	378356.68
	RSD (%)	0.3259	13.48
521.1998	1	5.2090	98735.12
	2	5.2230	63931.28
	3	5.2360	53656.56
	4	5.1990	70115.63
	5	5.2290	59519.67
	6	5.2100	76021.79
	7	5.2000	77674.37
	8	5.2120	77674.37
	9	5.2160	58253.28
	10	5.2150	57502.18
	11	5.2060	81518.88
	RSD (%)	0.2208	19.15
359.0839	1	6.0680	827987.89
	2	6.0650	810085.26

	3	6.1180	747019.97
	4	6.1240	837427.90
	5	6.1200	788153.09
	6	6.1290	888531.55
	7	6.1100	927707.82
	8	6.0910	949082.06
	9	6.1110	974361.76
	10	6.1300	1132204.22
	11	6.1420	1028432.78
	RSD (%)	0.4107	12.77
627.2366	1	18.3000	583684.54
	2	18.2990	445501.54
	3	18.2970	626878.24
	4	18.2928	472987.72
	5	18.2800	569508.48
	6	18.2830	480817.22
	7	18.2920	574336.74
	8	18.2900	520702.34
	9	18.2860	726789.53
	10	18.2930	671134.93
	11	18.2880	639975.57
	RSD (%)	0.0350	15.42
552.2392	1	18.4621	583684.54
	2	18.4680	881950.92
	3	18.4660	685300.48
	4	18.4538	648280.42
	5	18.4611	569508.48
	6	18.4624	512681.87
	7	18.4582	574336.74
	8	18.4607	513266.78
	9	18.4662	561686.55
	10	18.4627	491729.73
	11	18.4570	657266.54
	RSD (%)	0.02280	18.19

Table S3 IS-1 in ten QC samples from HSGC-MS analysis

No.	Peak area	Area sum %
1	5707375.20	29.02
2	4823403.13	25.35
3	5652064.94	27.72
4	5206285.93	26.44
5	6141511.16	28.77
6	4887036.11	27.39
7	4802649.01	27.66
8	4629018.10	26.32
9	3514539.57	22.47
10	3918542.40	26.71
RSD (%)	16.31	7.040

Table S4 IS-2 in eleven QC samples from UHPLC-QTOF/MS analysis

No.	Area	Area sum %
1	24405550.87	21.54
2	25679208.51	22.21
3	26305779.54	22.05
4	25253467.97	21.91
5	26017727.80	22.32
6	25854050.58	21.80
7	25735910.16	18.19
8	25735910.16	18.19
9	25999923.91	19.19
10	26217085.79	18.79
11	26743970.60	19.38
RSD (%)	2.350	8.450

Table S5 ROC curve analysis of the identified biomarkers from CJ and GJ

Test Result Variable(s)	Area Under the Curve	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Verbenol	0.600	0.130	0.450	0.346	0.854
trans-Carvyl acetate	0.550	0.132	0.705	0.292	0.808
Benzaldehyde	0.680	0.133	0.174	0.420	0.940
2,5,9-trimethyldecane	1.000	0.000	0.000	1.000	1.000
Uridine	0.500	0.132	1.000	0.241	0.759
2,6-dimethyl-2-heptanol	0.920	0.078	0.001	1.000	1.000
Limonene	0.500	0.132	1.000	0.241	0.759
1,3,4-Tri-caffeoylquinic acid	0.500	0.132	1.000	0.241	0.759
5-Acetyl-1,3,4-tri-caffeoylquinic acid	0.505	0.132	0.970	0.236	0.754
Tricyclene	0.535	0.147	0.791	0.247	0.823
cis- β -Farnesene	1.000	0.000	0.000	1.000	1.000
Eudesm-7(11)-en-4-ol	1.000	0.000	0.000	1.000	1.000
octahydronaphthalene	1.000	0.000	0.000	1.000	1.000
2-methylbut-2-enoate	1.000	0.000	0.000	1.000	1.000
α -Terpineol	1.000	0.000	0.000	1.000	1.000
γ -curcumene	1.000	0.000	0.000	1.000	1.000
β -Sesquiphellandrene	1.000	0.000	0.000	1.000	1.000
Isoquercetin-2-O-(6'-acetyl-glucoside)-2'-glucoside	0.950	0.058	0.001	0.837	1.000
Apigenin-7-caffeoylglucoside	0.900	0.079	0.002	0.744	1.000
Dihydroluteolin-7-O-raffinose	0.900	0.079	0.002	0.744	1.000
Uridine-5'-diphosphoglucose	0.900	0.058	0.001	0.837	1.000

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Table S6 ROC curve analysis of the identified biomarkers from CJ and HbJ

Test Result Variable(s)	Area Under the Curve	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
2,6-dimethyl-2-heptanol	1.000	0.000	0.000	1.000	1.000
Limonene	0.944	0.064	0.001	0.819	1.000
Isoquercetin-2-O-(6'-acetylglucoside)-2'-glucoside	1.000	0.000	0.000	0.000	1.000
Verbenol	0.556	0.136	0.683	0.289	0.822
trans-Carvyl acetate	0.500	0.136	1.000	0.233	0.767
Benzaldehyde	0.560	0.132	0.462	0.302	0.818
2,5,9-trimethyldecane	1.000	0.000	0.000	1.000	1.000
Uridine	0.500	0.136	1.000	0.233	0.767
1,3,4-Tri-caffeoylquinic acid	0.500	0.136	1.000	0.233	0.767
5-Acetyl-1,3,4-tri-caffeoylquinic acid	0.550	0.135	0.713	0.292	0.808
Tricyclene	1.000	0.000	0.000	1.000	1.000
Eudesm-7(11)-en-4-ol	1.000	0.000	0.000	1.000	1.000
octahydronaphthalene	1.000	0.000	0.000	1.000	1.000
2-methylbut-2-enoate	1.000	0.000	0.000	1.000	1.000
α -Terpineol	1.000	0.000	0.000	1.000	1.000
γ -curcumene	0.611	0.134	0.414	0.349	0.874
β -Sesquiphellandrene	1.000	0.000	0.000	1.000	1.000
Apigenin-7-caffeoylglucoside	0.950	0.058	0.001	0.837	1.000
Dihydroluteolin-7-O-raffinose	0.900	0.080	0.003	0.744	1.000
Uridine-5'-diphosphoglucose	0.900	0.080	0.003	0.744	1.000

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Table S7 ROC curve analysis of the identified biomarkers from GJ and HbJ

Test Result Variable(s)	Area Under the Curve	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Verbenol	0.544	0.135	0.744	0.280	0.809
trans-Carvyl acetate	0.550	0.135	0.713	0.286	0.814
Benzaldehyde	0.850	0.095	0.010	0.663	1.000
2,5,9-trimethyldecane	0.500	0.136	1.000	0.233	0.767
Uridine	0.500	0.136	1.000	0.233	0.767
2,6-dimethyl-2-heptanol	1.000	0.000	0.000	1.000	1.000
Limonene	0.944	0.064	0.001	0.819	1.000
1,3,4-Tri-caffeoylquinic acid	0.500	0.136	1.000	0.233	0.767
5-Acetyl-1,3,4-tri-caffeoylquinic acid	0.550	0.135	0.713	0.286	0.814
Tricyclene	0.972	0.032	0.001	0.910	1.000
cis- β -Farnesene	0.500	0.136	1.000	0.233	0.767
Eudesm-7(11)-en-4-ol	0.556	0.142	0.683	0.277	0.835
octahydronaphthalene	0.550	0.135	0.713	0.286	0.814
2-methylbut-2-enoate	0.650	0.128	0.270	0.399	0.901
α -Terpineol	0.889	0.081	0.004	0.731	1.000
γ -curcumene	0.500	0.136	1.000	0.233	0.767
β -Sesquiphellandrene	0.594	0.143	0.488	0.314	0.875
Isoquercetin-2-O-(6'-acetyl-glucoside)-2'-glucoside	0.500	0.136	1.000	0.233	0.767
Apigenin-7-caffeoylglucoside	0.500	0.136	1.000	0.233	0.767
Dihydroluteolin-7-O-raffinose	0.500	0.136	1.000	0.233	0.767

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Table S8 ROC curve analysis of the identified biomarkers from HJ and BJ

Test Result Variable(s)	Area Under the Curve	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Isoquercetin-2-O-(6'-acetyl-glucoside)-2'-glucoside	0.456	0.148	0.744	0.165	0.746
trans-Carvyl acetate	0.761	0.119	0.055	0.529	0.994
Benzaldehyde	0.889	0.087	0.004	0.718	1.000
2,5,9-trimethyldecane	0.833	0.103	0.014	0.631	1.000
Uridine	0.773	0.107	0.027	0.562	0.984
1,3,4-Tri-caffeoylquinic acid	0.727	0.123	0.142	0.502	0.952
5-Acetyl-1,3,4-tri-caffeoylquinic acid	0.798	0.111	0.041	0.502	0.952
Tricyclene	0.722	0.120	0.102	0.487	0.957
Eudesm-7(11)-en-4-ol	0.545	0.135	0.713	0.288	0.803
octahydronaphthalene	0.652	0.138	0.348	0.390	0.913
2-methylbut-2-enoate	0.594	0.145	0.488	0.311	0.878
α -Terpineol	0.545	0.135	0.713	0.288	0.803
β -Sesquiphellandrene	0.545	0.135	0.713	0.288	0.803
Apigenin-7-caffeoylglucoside	0.500	0.136	1.000	0.233	0.767
Dihydroluteolin-7-O-raffinose	0.556	0.136	0.683	0.289	0.822
Uridine-5'-diphosphoglucose	0.556	0.136	0.683	0.289	0.822
Verbenol	0.878	0.079	0.006	0.722	1.000
2,6-dimethyl-2-heptanol	0.747	0.120	0.086	0.562	0.984
Limonene	0.773	0.116	0.066	0.562	0.984
γ -curcumene	0.556	0.138	0.624	0.239	0.761

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5