Online Supporting Material

Supplemental text. Condition of UPLC tandem MS.

Agilent 1290-UHPLC

The mobile phase was composed of 0.1% formic acid in water (solvent A) and ACN (solvent B). The gradient profile was the following: 0–1.5 min: 2% B; 1.5–9 min: linear gradient from 2 to 50% B; 9–14 min: linear gradient from 50 to 95% B; 14–15 min: 95% B. The flow rate was 0.3 mL min^{-1}. The injection volume was 2μL.

Agilent 6540 Quadrupole-Time-of-Flight mass system

A Jet Stream electrospray ionization source was used for sample ionization. The following parameters were used throughout the study: curtain gas: Gas temperature (325 °C), gas flow (5 L/min), Nebulizer (40 psi), sheath gas temperature (325°C), sheath gas flow (10 L/min), capillary voltage (40 kV for positive and 35 kV for negative), and fragmentor (120 V).The mass scan range was m/z = 50-1700

Data transformation, normalization and standardization

Before sPLS-DA, a generalized log2 transformation was carried out, and several normalization methods were performed and evaluated by using SMART software[4]. The methods consisted of (1) no normalization; (2) quantile normalization (by sample); (3) sample-based standardization; (4) metabolite-based standardization. Performance of those methods were further evaluated according to the number of poor-quality replicate samples, which should be removed, in post-normalization. The more replicates were removed, the poorer the method was. In summary, 2, 2, 2 and 7 replicate samples should be removed after applying no normalization (Figure S1(A)), quantile normalization by sample (Figure S1(B)), sample-based standardization (Figure S1(C)), and metabolites-based standardization (Figure S1(D)), respectively. In other words, the first three methods performed equally well in terms of the number of poor-quality replicate samples. Next, we compared the results of cluster patterns before and after standardization. The results showed that a cluster pattern was not improved or even worse after a standardization was done (e.g., Figure S2(A) and S2(B) for the category of common plant oil). Therefore, we suggested a generalized log2 transformation and no standardization in our analysis.

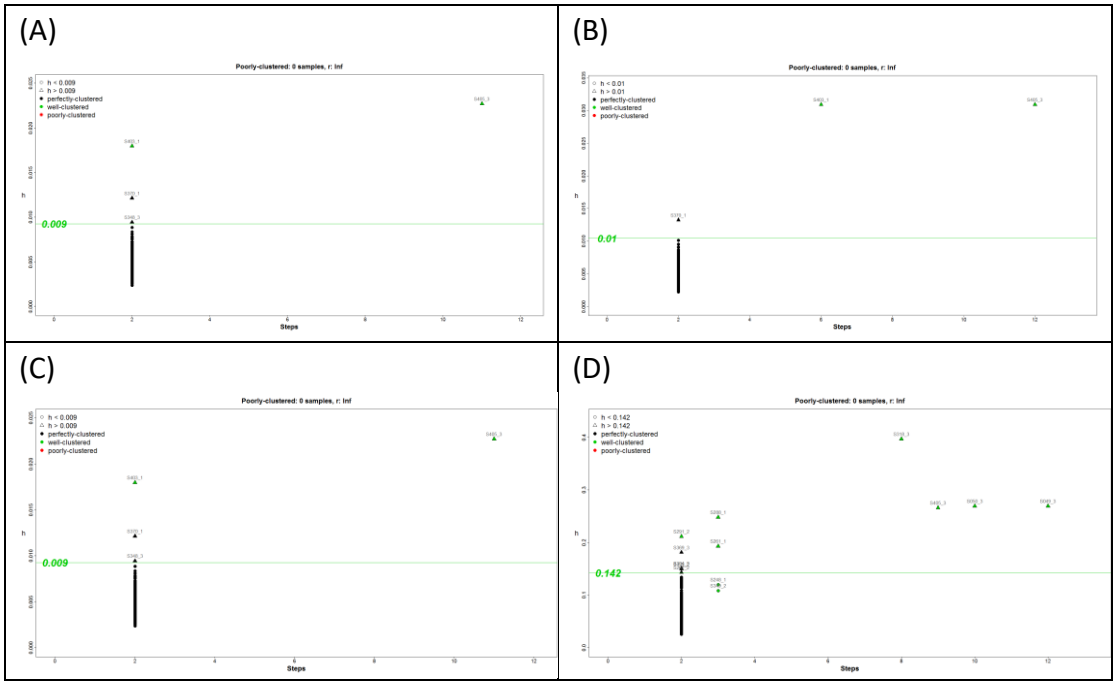

Figure S1. Representation of quality control. (A) No normalization; (B) Quantile normalization; (C) Sample-based standardization; (D) Metabolite-based standardization. Poorly-clustered replicate samples (marked by red if any) and well-clustered replicate samples (marked by green) and above the green horizontal reference line, should be removed after a normalization.

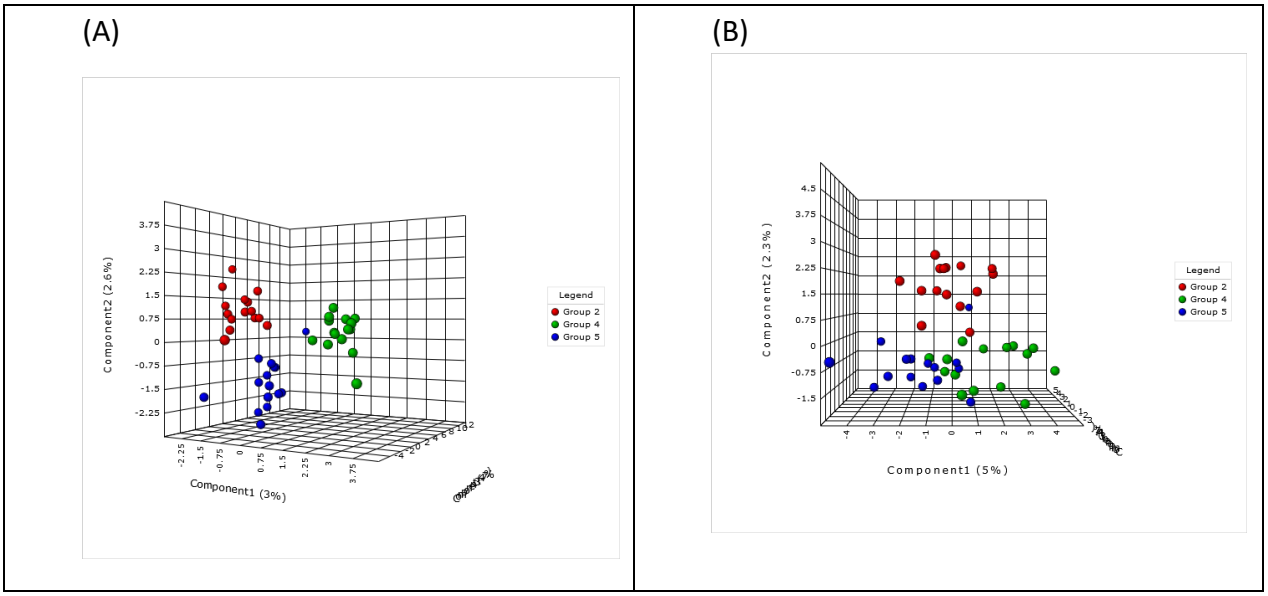

Figure S2. (A) Cluster pattern before a normalization; (B) Cluster pattern after a normalization.