Electronic Supplementary Material (ESI) for Analyst.

Compound identification in GC-MS by simultaneously evaluating mass spectrum and retention index

Xiaoli Wei[a] Imhoi Koo[a] Seongho Kim [b] and Xiang Zhang,[a]

[a] Department of Chemistry, University of Louisville, Louisville, KY 40292, United States

[b] Biostatistics Core, Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI 48201, United States

Table S1. Results of the five times of 5-fold cross validation for analysis of the query dataset containing retention indices acquired on standard non-polar column. Each datum is the average of the results of a 5-fold cross validation with its standard deviation.

	CV 1	CV 2	CV 3	CV 4	CV 5
a	0.05±0.02	0.05±0.03	0.06±0.03	0.04±0.03	0.04±0.03
b	23±6	25±8	23±7	21±8	24±6
w	0.66±0.02	0.71±0.08	0.64±0.08	0.69±0.04	0.67±0.03
Training error (%)	11.5±0.1	11.5±0.2	11.4±0.3	11.4±0.2	11.5±0.4
Testing error (%)	11.8±0.4	11.9±0.8	11.9±0.9	11.9±0.7	12.0±1.8

Table S2. Results of the five times of 5-fold cross validation for analysis of the query dataset containing retention indices acquired on standard polar column. Each datum is the average of the results of a 5-fold cross validation with its standard deviation.

	CV 1	CV 2	CV 3	CV 4	CV 5
a	0.02±0.01	0.02±0.01	0.03±0.02	0.02±0.01	0.01±0.01
b	26±7	32±3	26±6	29±1	26±7
w	0.72±0.07	0.67±0.07	0.73±0.13	0.73±0.09	0.75±0.05
Training error (%)	13.7±0.2	13.7±0.5	13.7±0.5	13.7±0.6	13.7±0.2
Testing error (%)	14.3±1.0	14.1±2.0	14.3±1.8	14.3±2.3	14.4±1.4