Supporting Information for

Rapid Mass Spectrometric Diagnosis of Bladder Cancer via Urine Carbonyl Metabolic Fingerprints

Yuze Li[1,2], Lixia Jiang[3], Zhenpeng Wang[4], Xiaohua Cao[5], Lingwei Meng[1,2], Jinghan Fan[2], Guanghui Wang[4], Caiqiao Xiong[1]*, and Zongxiu Nie[1,2]*

[1]*Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing, 100190 China*
[2] *University of Chinese Academy of Sciences, Beijing, 100049, China*

[3] *Department of Laboratory Medicine, First Affiliated Hospital of Gannan Medical University Ganzhou, Jiangxi Province, 341000, China*

[4] *National Center for Mass Spectrometry in Beijing, Beijing, 100190, China*

[5] *College of Chemical Engineering, Jiujiang University, Jiujiang, Jiangxi, 332005, China*

** E-mail: xiongcq@iccas.ac.cn*; *znie@iccas.ac.cn*

1. Experimental Procedures.

1.1 The atmospheric pressure glow discharge (APGD) ion source. The APGD consisted of a needle electrode in the axis, surrounded by a tube electrode as shown in Figure S1. Between them, there was a glass tube to insulate these two electrodes but free of the last 2 mm near the outlet of the source (Figure S1b). High voltage (CE1500002T, Rainworm Electronics (Shanghai) Co., Ltd. Shanghai, China) was applied to the needle electrode in current-controlled mode (Max voltage= 1500 V, current= 8 mA), while the tube electrode is grounded. To enhance the oxidation efficiency, 10% oxygen-added-argon gas (0.6 L/min) was used as the discharge gas and introduced to the gap between the tube electrode and the glass insulator. Without the insulation of the glass tube at the last 2 mm nearing the outlet of the source, the discharge gas was penetrated to generate glow discharge plasma when the voltage is higher than 400 V and the current is higher than 2 mA. Parameters such as the angle between APGD and the horizontal line, the distance from the MS inlet to the APGD discharge region, the discharge current and gas flow rate were optimized (Figure S2), and the setup worked in the optimal condition hereafter (the angle = 45°, the distance = 4 mm, the discharge current = 6 mA and the discharge gas flow rate = 0.6 L/min).

1.2 The sand core chip. The sand core is used in chemistry lab as filters to separate samples from precipitate and bacteria. Here, we use it to separate complex urine samples. We bought four types of sand core chips (G2, G3, G4 and G5) from the internet (Taobao.com). The average diameters of the micropores of the four types of sand core chips are 30~50 μm, 16~30 μm, 4~7 μm and 2~4 μm, respectively. The separation effects of four types of sand core chips were characterized and "G5" shows better performance (Figure S3). The tortuous and porous inner structure of the sand core chip (see Figure S3 right panel) provides the separation ability according to the volatility of the analytes and the adhesive force between analytes and the tiny cavity wall (van der Waals force).

2. Mechanism of the derivatives fragmentation.

DMED can label carboxyl, aldehyde, and ketone functional groups. All these three type compounds shifted to $\Delta m/z$ +70 after derivatization, but their fragmentation follow a different pathway. The charge site prefers to keep with nitrogen atoms. Thus, all these derivatives have two isomers, i.e., the charge site keeps in the tertiary amine nitrogen or the bridging nitrogen. Figure S5a, b, and c show the possible isomer pairs of the Schiff base, the carboxamide, and the oxidized Schiff base, respectively. The relative energy differences of each pair are also marked. Both isomers fragmentation pathway follows the charge site-initiated cleavage. The bond cleavage always occurs in the C-N bond next to the charge site, that's the origination of the two main type daughter ions, $\Delta m/z$ -45 (Type I Fragmentation) and *m/z* 72 (Type II Fragmentation). For Schiff base (derivated from aldehyde or ketone), the energy of bridging nitrogen-charged isomer is lower than that of the tertiary amine nitrogen-charged one (Figure S5a), but the difference is small. Figure S5d and g show the fragmentation mechanisms of both the isomers. The peak at *m/z* 72 is the major daughter ion of ketonic and aldehydic compounds (Figure S7, S8). As for carboxamide (derivated from carboxyl acid), the energy difference between two isomers is enormous (Figure S5b), and nearly all the molecular adopt the low-energy tertiary amine nitrogen-charged isomer to keep stable. $\Delta m/z$ -45 is almost the only product in the CID of carboxyl derivatives, whereas *m/z* 72 ions do not appear (Figure S6). The oxidized Schiff base (derivated from aldehydes) is similar to carboxamide, and $\Delta m/z$ -45 daughter ions were observed mainly (Figure S8). The well-matched experimental results verify the reliability of the theory presented here.

3. Discrimination of carboxyl, ketone, and aldehyde groups.

3.1 Carboxyl. Carboxyl acids such as fatty acid and small carboxyl acid metabolites are involved in a lot of biological processes.[1] Due to its unique properties, carboxylic acids are usually detected by ESI-MS under negative ion mode, but with low sensitivity.[2] Here, we employ DMED to label carboxyl acids and detected in positive ion mode by LTQ. The dehydration condensation reaction between carboxyl and amino produce an amide compound with an *m/z* shift of 70. The amide compound will neutral loss a dimethylamine during the CID process with an *m/z* shift of -45.[3] This characteristic feature helps us to discern carboxyl groups. As can be seen in Figure S6a, oleic acid

(*MW*=282) was labeled by DMED and formed a protonated amide compound at *m/z* 353. A daughter ion of Δ*m/z* -45 at *m/z* 308 was generated during the CID process. The peak of *m/z* 308 is the only daughter peak in the CID spectrum, which shows its unique characteristic property. And then, terephthalic acid was tested as an example of aromatic acid (Figure S6b). The peak at *m/z* 237 and 307 are the primary and secondary derivatization products, respectively. Similarly, the primary derivatization product generated a daughter ion of Δ*m/z* -45 at *m/z* 192. As predicted, the secondary derivatization product at *m/z* 307 generated two daughter ions of Δ*m/z* -45 and Δ*m/z* -90, corresponding to one and two molecular of dimethylamine neutral loss from the parent ion. The dicarboxylic acid experiment provides evidence of this method's ability to not only identify but also quantify functional groups by the order of derivatization. Glutaric acid, a fatty dicarboxylic acid, also shows a similar feature with other acids (Figure S6c). The difference is that the CID spectrum of glutaric acid secondary derivatization product shows not only Δ*m/z* -45 and Δ*m/z* -90 fragments ions, but also Δ*m/z* -88 ions. The peak of Δ*m/z* -90 and Δ*m/z* -88 chemical structure were shown nearing the corresponding peak.

3.2 Ketones. Ketones occur in various biological samples such as blood, urine, and exhaled breath.[4] Many hormones in the human endocrine system contain ketone groups. DMED was again employed here to label ketones. Amino reacts with carbonyl to form a carbon-nitrogen double bond compound (Schiff base). Figure S7a shows the spectra of acetone labeled by DMED. The peak at *m/z* 129 is the protonated Schiff base. Unlike carboxyl acid, ketones generate a daughter ion at *m/z* 72 during the CID process. This daughter ion also appears in the case of cyclohexanone (Figure S7b), benzophenone (Figure S7c), and other ketones. In the case of benzophenone, a prominent daughter ion appears at *m/z* 167. The structure of the daughter ion is also shown in Figure S7c. Such a case indicates that the method developed here is not universal, and some special chemical environment should be taken into consideration. Fortunately, in most situations, it does work.

3.3 Aldehydes. Aldehydes play important roles in the biological system as they are the products of lipid peroxidation, which has been linked to the etiology of several diseases, including Alzheimer's disease.[5] Accurate determination of aldehydes is of great significance for early disease

screening and treatment. Several aldehyde determination methods have been developed, mainly based on LC-MS,[5, 6] GC-MS,[7-9] and proton-transfer-reaction mass spectrometry (PTR-MS).[10] DMED can react with aldehydes to form a Schiff base in the same way with ketones. The difference is that the aldehydes-derivated Schiff base can be oxidized further due to a lower steric hindrance. There are plenty of reactive oxygen species such as ozone and hydroxyl radicals that can be the oxidants in the glow discharge plasma, especially in the case of 10% oxygen-added-argon discharge gas. The peak at *m/z* 177 in Figure S8a is the protonated benzaldehyde-derivated Schiff base, and the red-marked *m/z* 193 is its oxidized peak, which was not observed in ketones. The Schiff base of benzaldehyde –derivated also generates *m/z* 72 daughter ion in the CID process like ketones. A dialdehyde can generate primary and secondary derivatives as shown in Figure S8b. Both derivatives were further oxidized and the peaks of $\Delta m/z$ +16 appeared as red-marked. During the CID process, the primary product generated an *m/z* 72 daughter ion, while the oxidized derivative generated a $\Delta m/z$ -45 daughter ion.

4. Simultaneous determination of carboxyl and hydroxyl groups.

This method had also been proven to be an efficient tool for the simultaneous determination of carboxyl and hydroxyl groups. A piece of obvious evidence is the derivatization reaction of DL-malic acid, a natural product with two carboxyls and a hydroxyl group. As can be seen in Figure S9a and d, the derivatization reaction of DL-malic acid labeled by DMED generated a primary product at *m/z* 205 and a secondary product at *m/z* 275. The carboxyl acid derivatives, both primary product and secondary product generate $\Delta m/z$ -45 daughter ions (Figure S9e and f). Interestingly, peaks at $\Delta m/z$ -18 appear near the protonated derivatives, which indicate the presence of a hydroxyl group in the molecule. Figure S9b and c show the proposed dehydration processes of the primary and secondary products. The hydroxyl takes off followed by the oxygen of the carbonyl takes off to form a carbon-nitrogen double bond. The CID spectra in Figure S9e, f, g and h verified the assumption proposed in Figure S9b and c. In other carboxyl acids without hydroxyl, such dehydration products peak not appear.

5. Identification of the metabolites in urine samples.

The identification of those unknown metabolites in the urine samples starts from the determination of the functional groups, followed by searching the HMDB database. Specifically, in an MS scan of a urine sample, all the peaks with a signal-to-noise ratio (S/N) higher than 3 were isolated for further collision induced dissociation (CID). The collision energy was 25 and the mass window was 1 Da. And then, those peaks with a characteristic CID pattern (type I or type II fragmentation, Figure S5) were selected as the potential biomarkers. Thereby, the functional groups of these potential biomarkers can be deduced. Then searching the molecular weight and the corresponding functional groups in HMDB will help us determine the precise structure of the metabolites. The assignments were further confirmed by high-resolution MS and listed in Table S1. High-resolution MS was performed on an FT-ICR-MS and a Q-TOF-MS. The former one has a very high resolution but moderate sensitivity, while the latter one is the opposite. That's why both two instruments were used.

6. Quantification details

To explore the limit of the detection (LOD), a series of oleic acid standards solutions with different concentrations were prepared. Figure S10a shows the derivatization reaction of oleic acid labeled by DMED, and Figure S10b shows the mass spectrum of the reaction system in the concentration of 3.5×10^{-13} mol/L. The signal to noise ratio S/N is higher than 10 in this concentration, which indicates a better LOD than 0.35 pM. Typically, carboxylic compounds can be detected in the negative ion mode with moderate sensitivity. To show the sensitivity advantage of the labeling strategy, the LOD of underived oleic acid was measured as 30μM as shown in Figure S10c.

Nine-Phenylacridine was used as an internal standard (IS) to achieve the relative quantification of the potential biomarkers in urine samples. Figure S11a shows the chemical structure of the IS. Creatinine is one of the major components in the urine sample, and we employed it as an example to show the relative quantification method. Figure S11b shows the mass spectra of the mixtures of a constant concentration of IS and increased concentration of creatinine. Figure S11c shows the ratio of creatinine intensity to IS intensity (R(A/I)) increases with the increase of creatinine

concentration in good linearity. We believe that the fluctuation of the ratio R reflects the concentration changes of biomarkers.

7. Machine learning.

7.1 Matrices processing. Totally 42 carboxylic, aldehydic and ketonic metabolites were identified and relatively quantified. The relative concentrations of these metabolites were set as the features, and the BC/HC status was set as the target. Samples from 28 BC patients and 38 controls were tested and repeated three times for each. One hundred and ninety-eight data points formed the matrix for machine learning (Figure S12). Every column in the matrix represents a single repeat of a sample, and every row represents a potential biomarker. And the colors indicate the relative concentration. The matrix shows visible differences between BC and HC**.** And then, in a typical procedure, the matrix was split into two parts (3:1) randomly to training and testing the machine learning models.

7.2 Model Selection. Several commonly used models were applied to classify the data, including Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). To evaluate these models, 5 indicators were calculated for each model and listed in Table S2. Area Under the Curve (AUC) was calculated from the average Receiver Operating Characteristic (ROC) curve of each model. In a confusion matrix, the data can be separated into four classes, True Positive (TP), False positive (FP), True negative (TN), and False negative (FN).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

Besides, Principal Component Analysis (PCA) was also applied to classify the data in the first three principal components. However, the first three components only explained 55% variance as shown in Figure S13.

In general, Logistic Regression performs best, and to be the selected one.

7.3 Details about the Logistic Regression (LR) model.

LR can be regarded as a linear regression model in a broad sense**.** Consider a linear regression model:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Where z is the BC/HC status and x_1, x_2 ... x_n are the features.

Logarithm of both sides:

$$g(z) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}} = \frac{1}{1 + e^{-z}}$$

This transforms the predicted value of the linear regression model into a probability value between [0,1] (nonlinear logistic regression).

The equation was solved by a home-written python script based on the sklearn package, and the major parameters are set as follows: penalty='l2', tol=0.00001, fit_intercept=True, class_weight=None, max_iter=100, solver='lbfgs'.

7.4 Five-fold cross-validation. Five-fold cross-validation was performed to evaluate the LR model more comprehensively. As shown in Figure S14, the dataset was split into five parts and in each fold, take one of the five as testing data, and the others as training data. After five-folds, all the data has been used for training and testing.

7.5 FreeViz. FreeViz was performed by an open-source software Orange 3.0 as shown in Figure S15. Firstly, assuming all the features contribute equally to the classification and formed a cycle as Figure S15a. After optimized by FreeViz, the most important features were shown in the longer arrows. The length of the arrow represents the contribution of the feature to the classification as shown in Figure S15b. It can be seen that *m/z* 250, 203, 237,187, 193, 224, 158 are more important than other ones in the classification.

SI References

(1) Kloos, D.; Lingeman, H.; Mayboroda, O. A.; Deelder, A. M.; Niessen, W. M. A.; Giera, M., Analysis of biologically-active, endogenous carboxylic acids based on chromatography-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2014,** *61*, 17-28.

(2) Zhang, T.-Y.; Li, S.; Zhu, Q.-F.; Wang, Q.; Hussain, D.; Feng, Y.-Q., Derivatization for liquid chromatography-electrospray ionization-mass spectrometry analysis of small-molecular weight compounds. *TrAC Trends in Analytical Chemistry* **2019,** *119*.

(3) Zhu, Q. F.; Zhang, Z.; Liu, P.; Zheng, S. J.; Peng, K.; Deng, Q. Y.; Zheng, F.; Yuan, B. F.; Feng, Y. Q., Analysis of liposoluble carboxylic acids metabolome in human serum by stable isotope labeling coupled with liquid chromatography-mass spectrometry. *J Chromatogr A* **2016,** *1460*, 100-9.

(4) Diskin, A. M.; ˇSpanˇel, P.; Smith1, D., Time variation of ammonia, acetone, isoprene and ethanol in breath: a quantitative SIFT-MS study over 30 days. *Physiological Measurement* **2003,** *24*, 107-119.

(5) Williams, T. I.; Lovell, M. A.; Lynn, B. C., Analysis of derivatized biogenic aldehydes by LC tandem mass spectrometry. *Analytical chemistry* **2005,** *77* (10), 3383-9.

(6) Guo, N.; Peng, C. Y.; Zhu, Q. F.; Yuan, B. F.; Feng, Y. Q., Profiling of carbonyl compounds in serum by stable isotope labeling - Double precursor ion scan - Mass spectrometry analysis. *Anal Chim Acta* **2017,** *967*, 42-51.

(7) Serrano, M.; Gallego, M.; Silva, M., Quantitative analysis of aldehydes in canned vegetables using static headspace-gas chromatography-mass spectrometry. *J Chromatogr A* **2017,** *1524*, 21-28.

(8) Anton, A. P.; Ferreira, A. M.; Pinto, C. G.; Cordero, B. M.; Pavon, J. L., Headspace generation coupled to gas chromatography-mass spectrometry for the automated determination and quantification of endogenous compounds in urine. Aldehydes as possible markers of oxidative stress. *J Chromatogr A* **2014,** *1367*, 9-15.

(9) Serrano, M.; Gallego, M.; Silva, M., Analysis of endogenous aldehydes in human urine by static headspace gas chromatography-mass spectrometry. *J Chromatogr A* **2016,** *1437*, 241-246.

(10) Blake, R. S.; Patel, M.; Monks, P. S.; Ellis, A. M.; Inomata, S.; Tanimoto, H., Aldehyde and ketone discrimination and quantification using two-stage proton transfer reaction mass spectrometry. *International Journal of Mass Spectrometry* **2008,** *278* (1), 15-19.

Figures and Tables

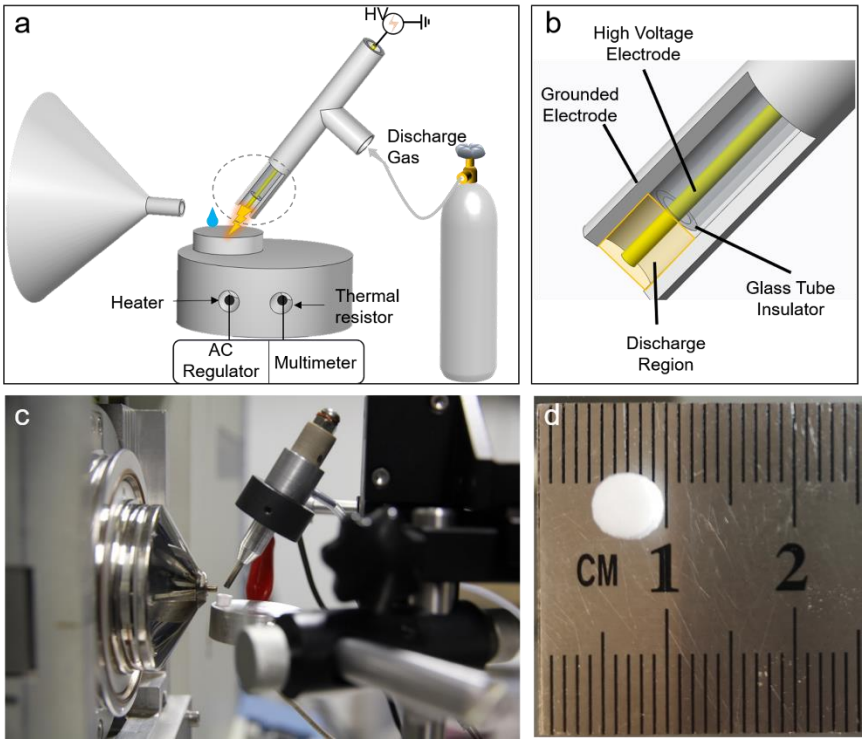


Figure S1. Details of the DSI platform. a) Schematic diagram and (b) partial section view of the DSI-MS platform. A photograph of (c) the DSI-MS platform and (d) the sand core chip.

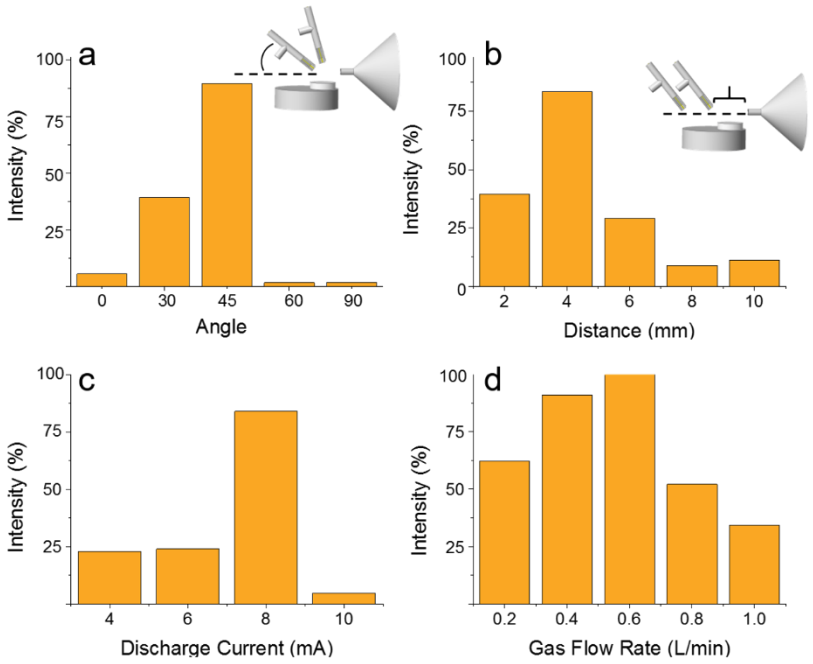


Figure S2. The angle between APGD and the horizontal line (a), the distance from the MS inlet to the APGD discharge region (b), the discharge current, (c) and gas flow rate (d) were optimized.

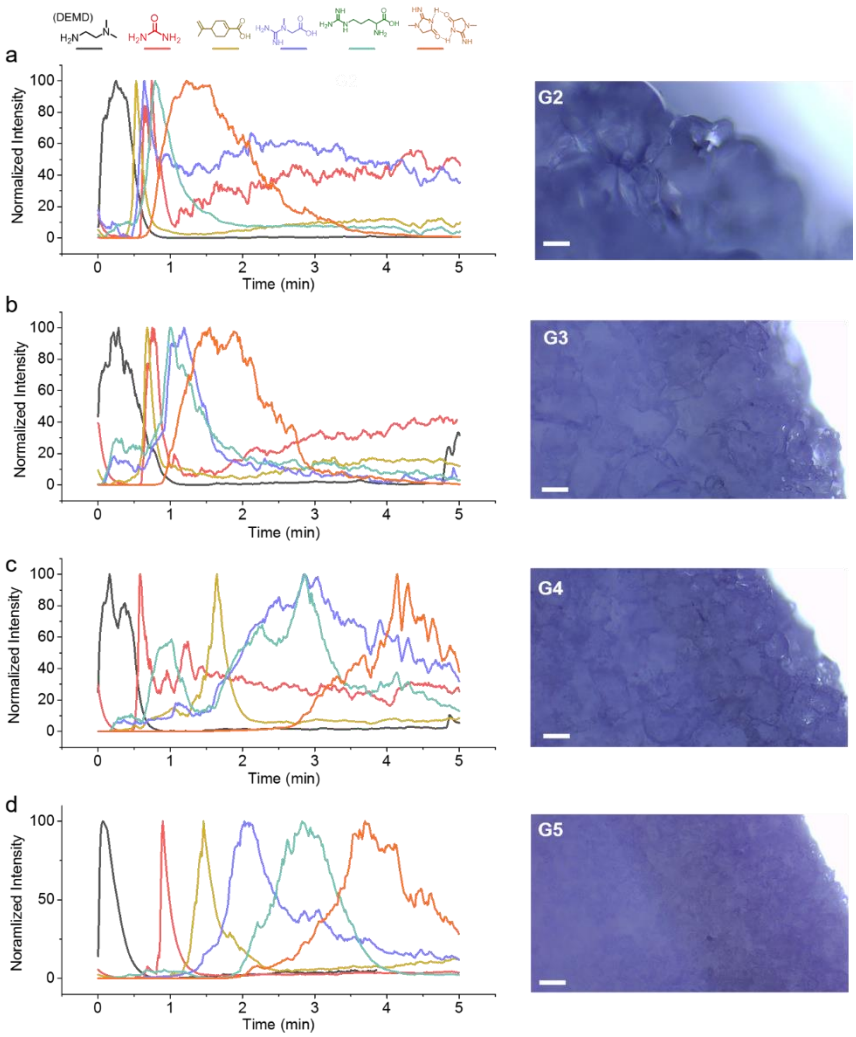


Figure S3. The separation effects of G2 (a), G3 (b), G4 (c), and G5 (d) type of sand core chips characterized by the extracted ion chromatogram (EIC) of several representative compounds in the urine. The right sides are the corresponding optical microscope photos of the chips, and the bar represents 50μm.

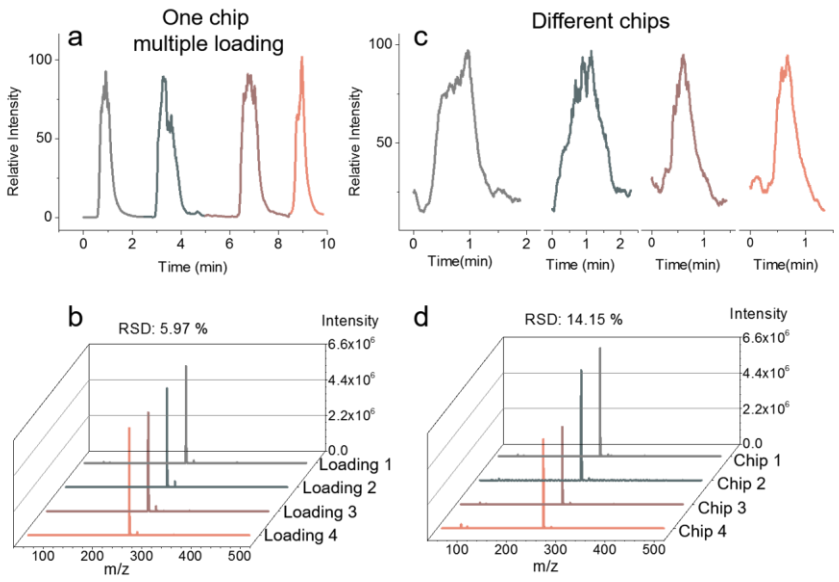


Figure S4. The reproducibility of the DSI-MS platform. During multiple loadings of the 9-Phenylacridine solution to the same sand core chip, (a) the extracted ion chromatography and (b) the mass spectrum is relatively stable and the relative standard deviation (RSD) is 5.97%. And the RSD is 14.15% between different chips (c and d) for the same sample.

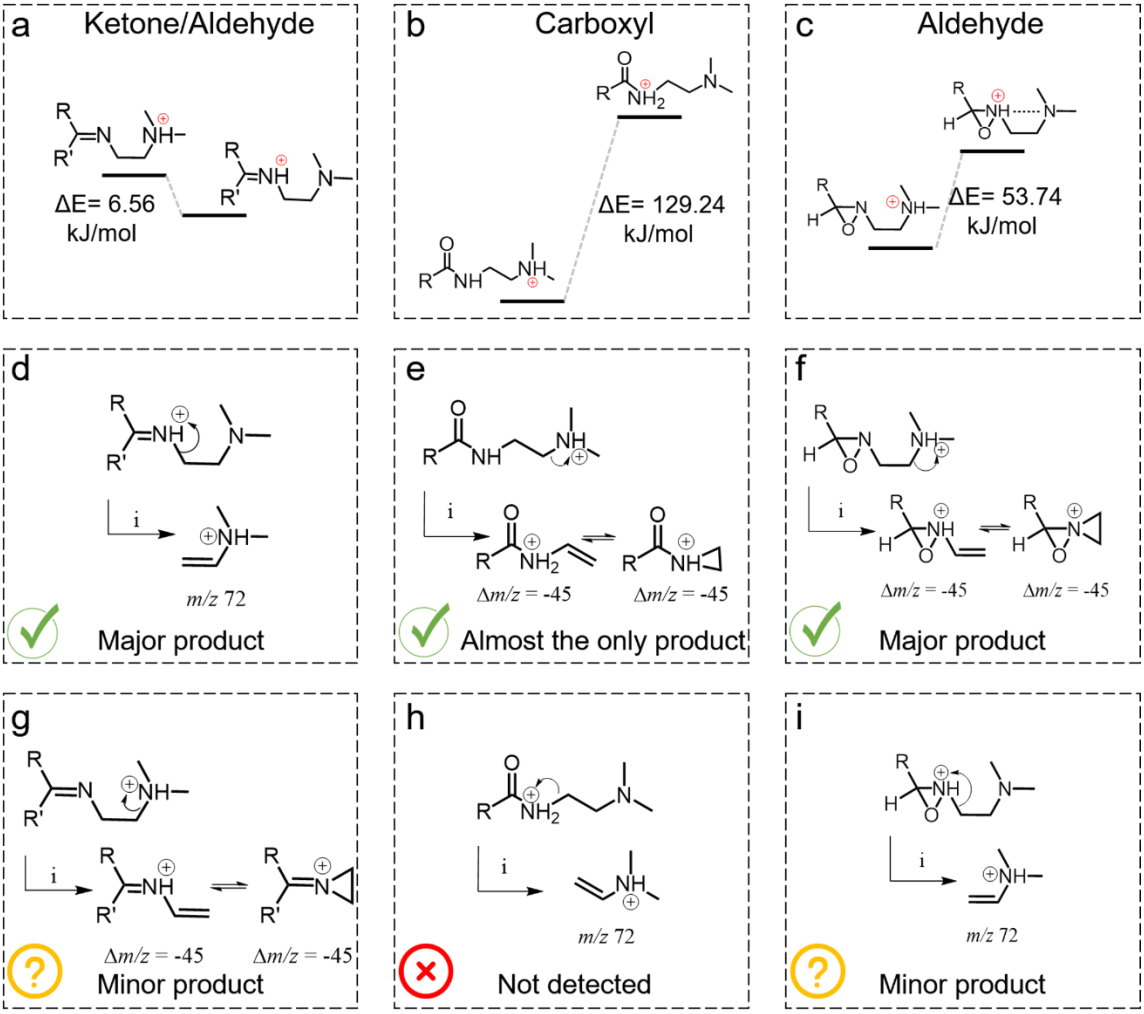


Figure S5. The possible isomer pairs and energy differences of (a) Schiff base, (b) carboxamide, and (c) oxidized Schiff base. The corresponding possible fragmentation pathway of each pair are shown in (d) and (g), (e) and (h), (f) and (i), respectively.

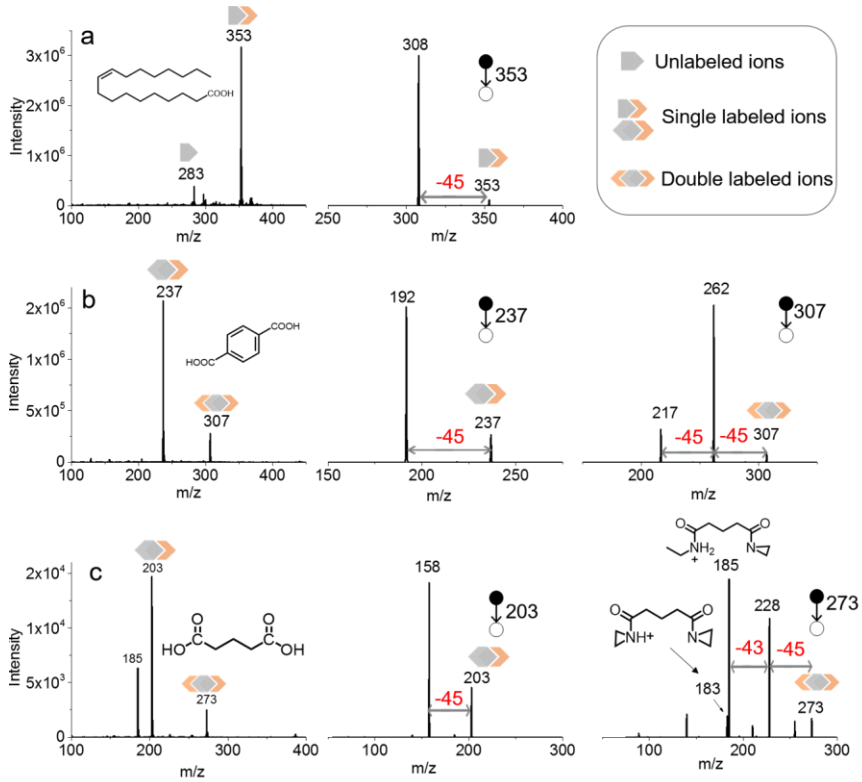


Figure S6. Mass spectra of (a) oleic acid, (b) terephthalic acid, and (c) glutaric acid labeled by DMED. The CID spectra of the corresponding derivatives are shown on the right side.

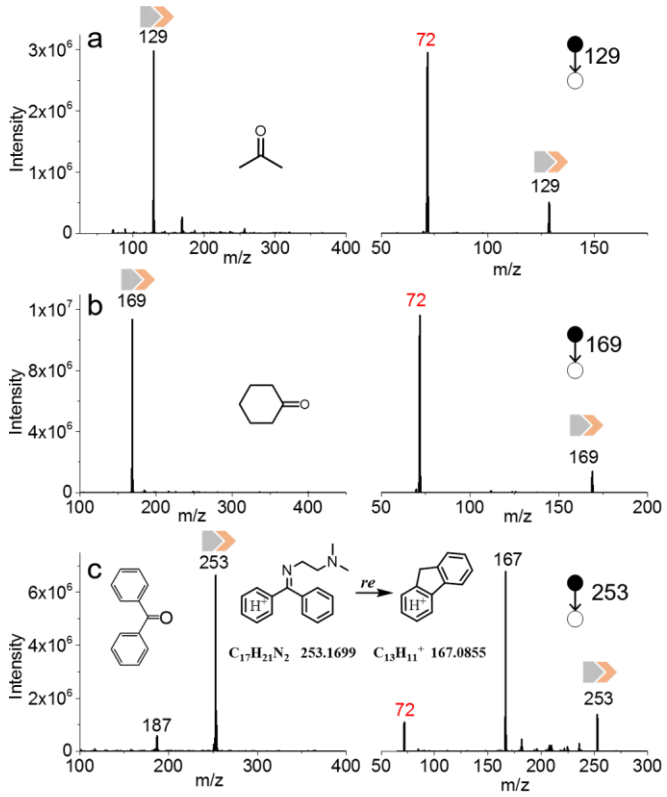


Figure S7. Mass spectra of (a) acetone, (b) cyclohexanone, and (c) benzophenone labeled by DMED. The CID spectra of the corresponding derivatives are shown on the right side.

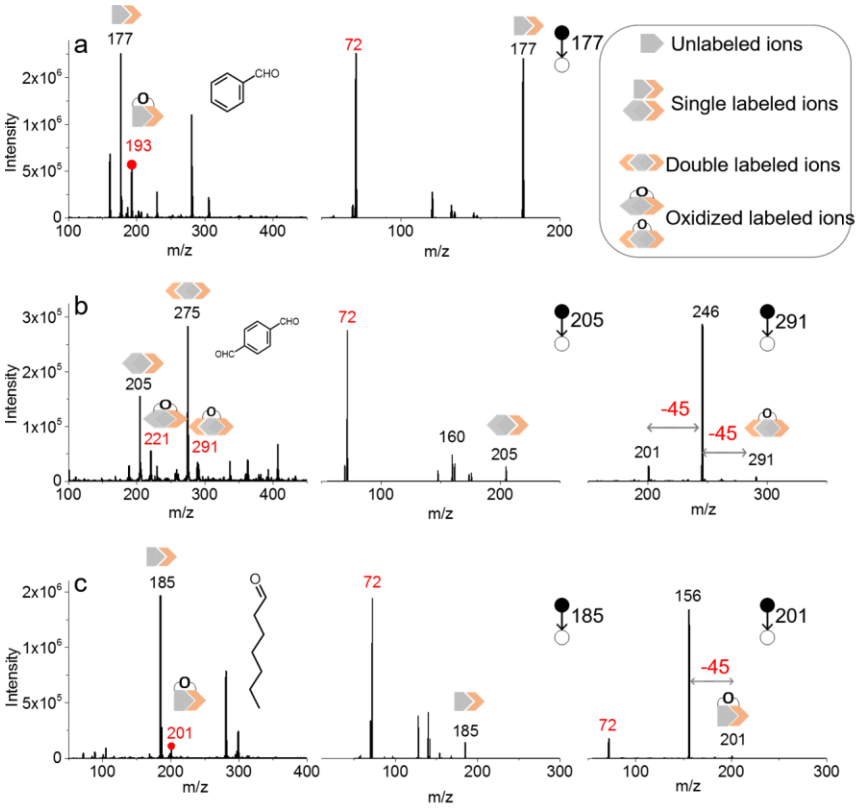


Figure S8. Mass spectra of (a) benzaldehyde, (b) terephthalaldehyde, and (c) heptaldehyde labeled by DMED. The CID spectra of the corresponding derivatives are shown on the right side.

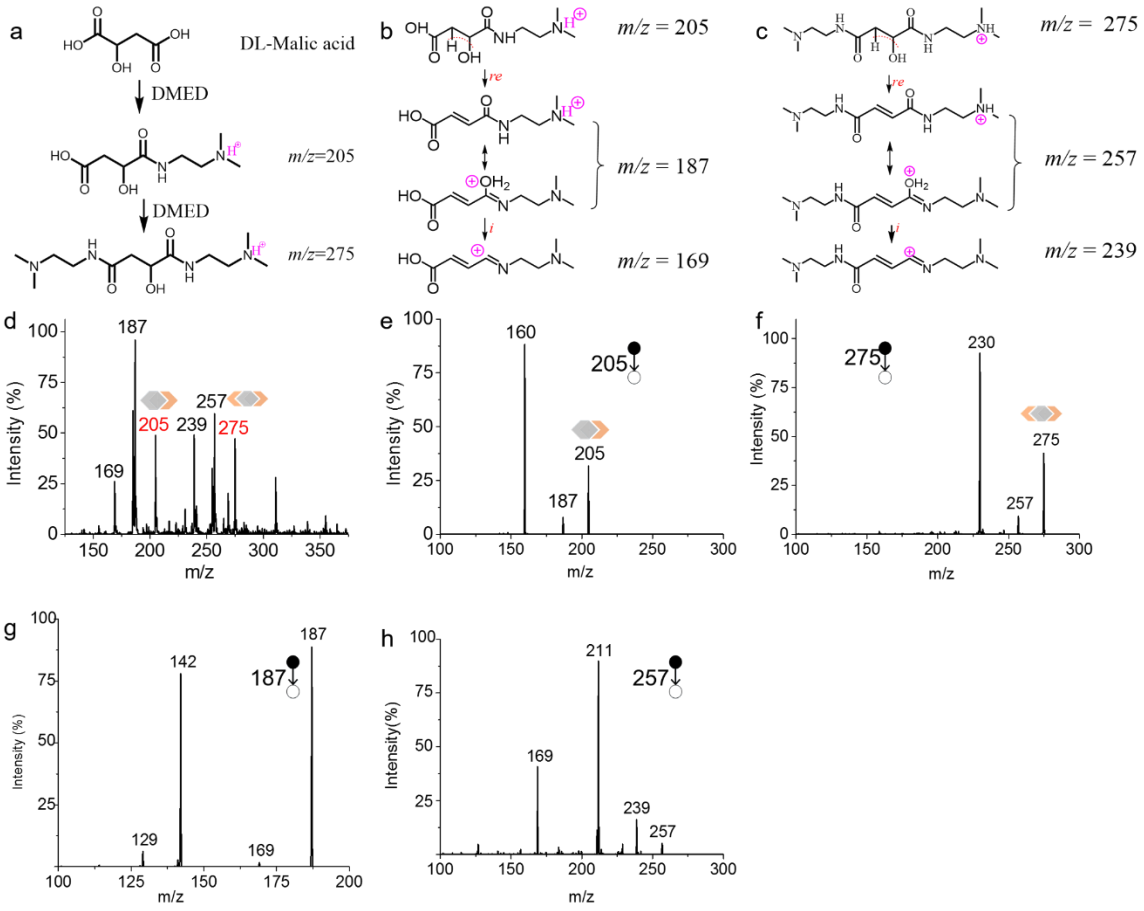


Figure S9. Simultaneous determination of carboxyl and hydroxyl groups. (a) The reaction of DL-malic acid labeled by DMED; The proposed dehydration processes of primary (b) and secondary (c) derivatization products; Mass spectra of the reaction system (d) and the corresponding CID spectra of *m/z* 205 (e), *m/z* 275 (f), *m/z* 187 (g) and *m/z* 257 (h).

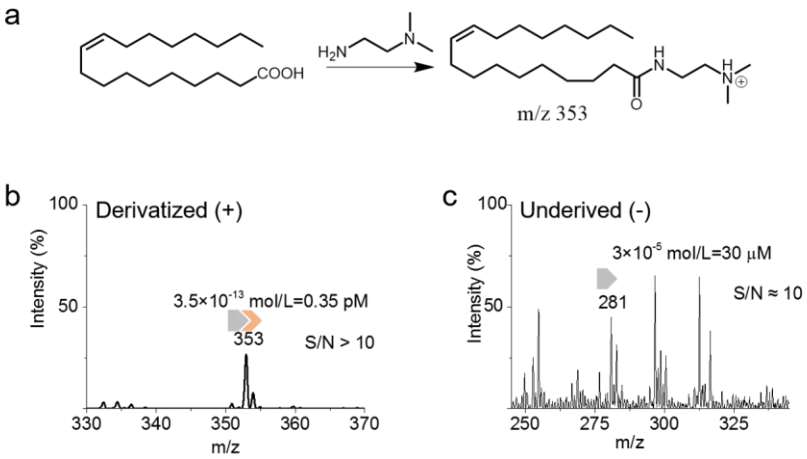


Figure S10. (a) The reaction of oleic acid labeled by DMED and (b) the corresponding mass spectrum in the concentration of 0.35 pM detected in positive ion mode. For comparison, (c) show the mass spectrum of underived oleic acid detected in negative ion mode.

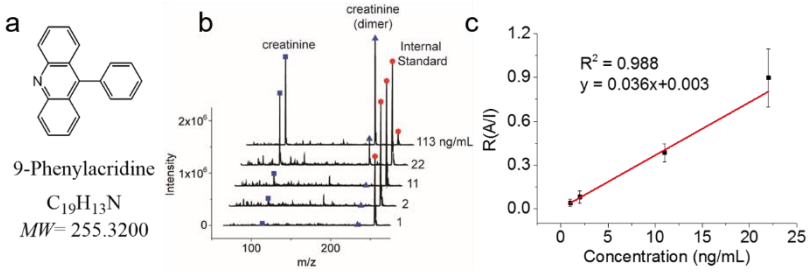


Figure S11. (a) The molecule structure of the internal standard (IS) 9-phenylacridine. (b) Mass spectra of the mixtures of a constant concentration of IS (1.5 μmol/mL) and increased concentration of creatinine. (c) The ratio of creatinine intensity to IS intensity (ratio R(A/I)) increases with the increase of creatinine concentration in good linearity.

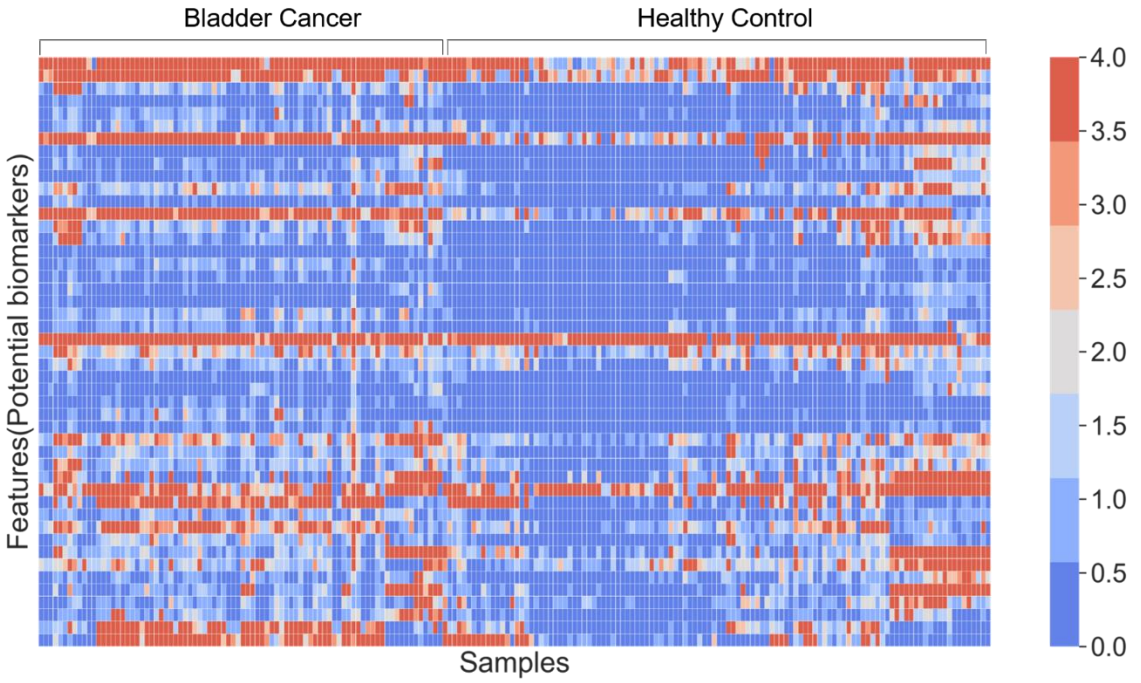


Figure S12. The acquired relative concentrations of potential biomarkers data were reorganized to a data matrix for machine learning.

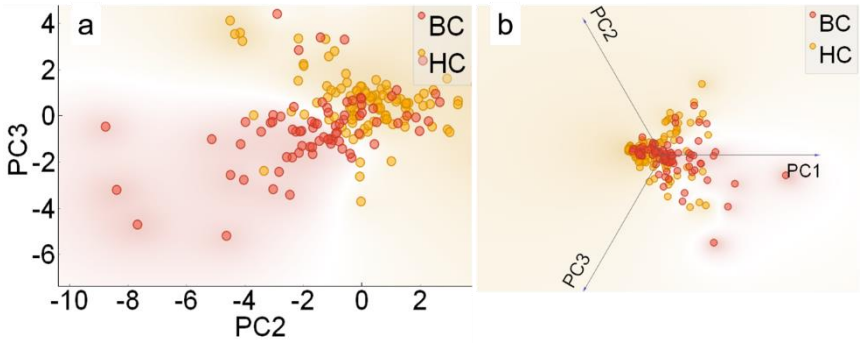
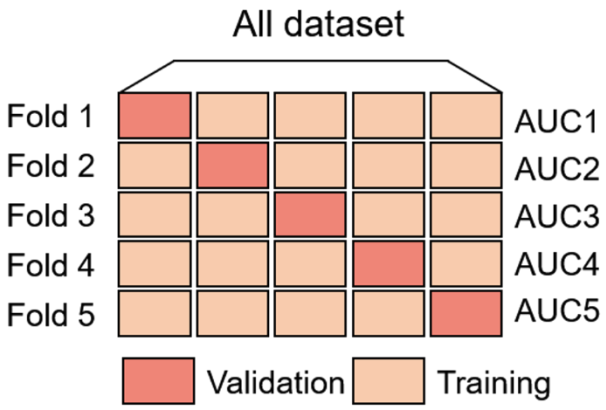


Figure S13. (a) 2D and (b) 3D plot of principal component analysis (PCA) of the dataset. The first three principal components explained 55% of the variance.



$$Mean\ AUC = \frac{1}{5}\sum_{i=1}^{5} AUC_i$$

Figure S14. Five-fold cross-validation of the dataset. The dataset was split into 5 parts and in each fold, one of the five parts was used for model testing while the others for training. After 5 folds, all these five parts have been tested.

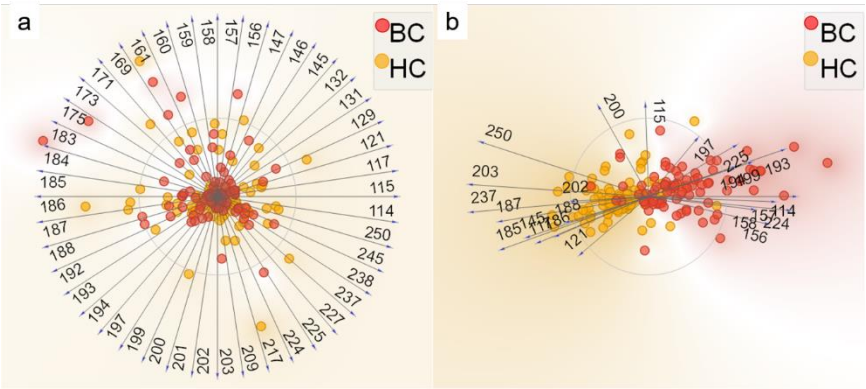


Figure S15. Features (a) before and (b) after FreeViz optimization. Those features with a longer arrow in (b) contribute more to the classification.

Table S1. Information about the identified metabolites in the urine samples. [a][b].

no.	Retention Time (min)	m/z (Calculated)	m/z (observed)	Δm/z (ppm)	Ion type	Major Fragment	Classification	Assignment	Formula
1	0.33	89.1073	89.1068	5.61	$[M+H]^+$	-	Lable Reagent	DMED	$C_4H_{12}N_2$
2	0.4	158.1652	158.1656	2.53	$[M\_DMED+H]^+$	m/z= 72	Ketone	Amino Butanone	C_4H_9NO
3	0.44	201.1961	201.1890	35.29	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Heptanoic Acid	$C_7H_{14}O_2$
4	0.45	209.1284	209.1275	4.30	$[M\_DMED+H]^+$	Δm/z= -45	Acid	4-Hydroxybenzoic acid	$C_7H_6O_3$
5	0.57	238.1220	238.1129	38.22	$[M\_DMED+H]^+$	m/z= 72	Ketone	N-Acetyltaurine	$C_4H_9NO_4S$
6	0.58	117.1022	117.1079	48.68	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Formic acid	CH_2O_2
7	0.58	256.1121	256.1057	24.99	$[M+H]^+$	-	Internal Standard	9-Phenylacridine	$C_{19}H_{13}N$
8	0.59	157.1335	157.1350	9.55	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Isocrotonic acid	$C_4H_6O_2$
9	0.63	131.1179	131.1181	1.53	$[M\_DMED+H]^+$	Δm/z= -45	Acid	AceticAcid	$C_2H_4O_2$
10	0.64	156.1131	156.1135	2.56	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Cyanoacetic acid	$C_3H_3NO_2$
11	0.64	160.1444	160.1449	3.12	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Alanine	$C_3H_7NO_2$
12	0.64	199.1805	199.1760	22.59	$[M\_DMED+H]^+$	Δm/z= -45	Acid	3-Heptenoic acid	$C_7H_{12}O_2$
13	0.65	146.1288	146.1212	52.01	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Glycine	$C_2H_5NO_2$
14	0.68	183.1128	183.1136	4.37	$[M\_DMED+H]^+$	Δm/z= -45	Acid	2-Furoic acid	$C_5H_4O_3$
15	0.68	184.1444	184.1409	19.01	$[M\_DMED+H]^+$	Δm/z= -45	Acid	1-pyrroline-5-carboxylic acid	$C_5H_7NO_2$
16	0.69	132.1131	132.1133	1.51	$[M+H]^+$	-	Acid	Creatine	$C_4H_9N_3O_2$
17	0.69	217.1659	217.1686	12.43	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Glutamine	$C_5H_{10}N_2O_3$
18	0.69	224.1393	224.1517	55.32	$[M\_DMED+H]^+$	Δm/z= -45	Acid	5-Aminosalicylic acid	$C_7H_7NO_3$
19	0.78	121.0720	121.0772	42.95	$[2M+H]^+$	-	-	Urea (Dimer)	$(CH_4N_2O)_2$
20	0.81	185.1285	185.1293	4.32	$[M\_DMED+H]^+$	Δm/z= -45	Acid/acetone	Acetylacrylic acid	$C_5H_6O_3$
21	0.86	203.1866	203.1876	4.92	$[M\_DMED+H]^+$	Δm/z= -45	Acid	L-Ornithine	$C_5H_{12}N_2O_2$
22	0.87	186.1601	186.1528	39.21	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Proline	$C_5H_9NO_2$
23	0.95	175.1077	175.1036	23.41	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Hydroxypyruvic acid	$C_3H_4O_4$
24	1.11	171.1492	171.1447	26.29	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Allylacetic acid	$C_5H_8O_2$
25	1.11	188.1393	188.1458	34.55	$[M\_DMED+H]^+$	Δm/z= -45	Acid	N-Acetylglycine	$C_4H_7NO_3$
26	1.11	197.1397	197.1454	28.91	$[M\_DMED+H]^+$	Δm/z= -45	Acid	2-(1-Imidazolyl) acetic Acid	$C_5H_6N_2O_2$
27	1.11	200.1393	200.1403	5.00	$[M\_DMED+H]^+$	Δm/z= -45	Acid	L-Pyroglutamic acid	$C_5H_7NO_3$

28	1.25	159.1123	159.1029	59.08	$[M\_DMED+H]^+$	m/z= 72	Ketone\Acid	Pyruvic acid	$C_3H_4O_3$
29	1.25	187.1805	187.1880	40.07	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Hexanoic acid	$C_6H_{12}O_2$
30	1.25	225.1234	225.1107	56.41	$[M\_DMED+H]^+$	Δm/z= -45	Acid	2,5-Dihydroxybenzoic acid	$C_7H_6O_4$
31	1.52	237.1961	237.2087	53.12	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Perillic acid	$C_{10}H_{14}O_2$
32	1.91	173.1648	173.1557	52.55	$[M\_DMED+H]^+$	Δm/z= -45	Acid	N-Valeric Acid	$C_5H_{10}O_2$
33	1.92	161.1284	161.1271	8.07	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Lactic acid	$C_3H_6O_3$
34	1.97	250.1550	250.1565	6.00	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Hippuric acid	$C_9H_9NO_3$
35	2.05	202.1662	202.1666	1.98	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Creatine	$C_4H_9N_3O_2$
36	2.4	115.1230	115.1219	9.55	$[M\_DMED+H]^+$	m/z= 72	Aldehyde	Acetaldehyde	C_2H_4O
37	2.59	114.0662	114.0661	0.88	$[M+H]^+$	-	-	Creatinine	$C_4H_7N_3O$
38	2.59	145.1335	145.1396	42.03	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Propionic acid	$C_3H_6O_2$
39	2.59	147.1228	147.1255	18.35	$[M+H]^+$	Δm/z= -45	Acid	Glycolic acid	$C_2H_4O_3$
40	2.59	192.1165	192.1212	24.46	$[M\_DMED+H]^+$	Δm/z= -45	Acid	L-Cysteine	$C_3H_7NO_2S$
41	2.61	169.1335	169.1325	5.91	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Vynilacrylic acid	$C_5H_6O_2$
42	2.61	193.1335	193.1335	0.00	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Benzoic acid	$C_7H_6O_2$
43	2.61	194.1288	194.1273	7.73	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Nicotinic acid	$C_6H_5NO_2$
44	2.62	129.1386	129.1457	54.98	$[M\_DMED+H]^+$	m/z= 72	Ketone	Acetone	C_3H_6O
45	2.97	245.2084	245.2012	29.36	$[M\_DMED+H]^+$	Δm/z= -45	Acid	Arginine	$C_6H_{14}N_4O_2$
46	3.75	227.1251	227.1264	5.72	$[2M+H]^+$	-	-	Creatinine (dimer)	$(C_4H_7N_3O)_2$

[a]. Labeling reagent, internal standard, urea, and creatinine are also shown in the table as refers.
[b]. The mass tolerances are relatively high because of the lack of performance of the instruments.

Table S2. The comparison of classification performance for the dataset between models.

Model	AUC	Accuracy	F1	Precision	Recall
Logistic Regression	0.84	0.87	0.87	0.87	0.87
Random Forest	0.89	0.82	0.82	0.82	0.82
SVM	090	0.80	0.80	0.80	0.80
kNN	0.86	0.78	0.79	0.79	0.78
Naïve Bayes	0.81	0.69	0.69	0.71	0.69