Supplementary information

Screening of Hepatocellular Carcinoma via Machine Learning Based on Atmosphere Pressure Glow Discharge Mass Spectrometry

Jinghan Fan [a, b], Xiao Wang [a, b], Yile Yu [a, b], Yuze Li [c, *], and Zongxiu Nie [a, b *]

[a] Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

[c] State Key Laboratory of High-efficiency Utilization of Coal and Green Chemical Engineering, College of Chemistry and Chemical Engineering, Ningxia University, Yinchuan 750021, China

* Corresponding authors:
Zongxiu Nie
Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing
100190, China
Email: znie@iccas.ac.cn
Yuze Li
State Key Laboratory of High-efficiency Utilization of Coal and Green Chemical Engineering, College of Chemistry and Chemical Engineering, Ningxia University, Yinchuan
750021, China
Email: liyuze@nxu.edu.cn

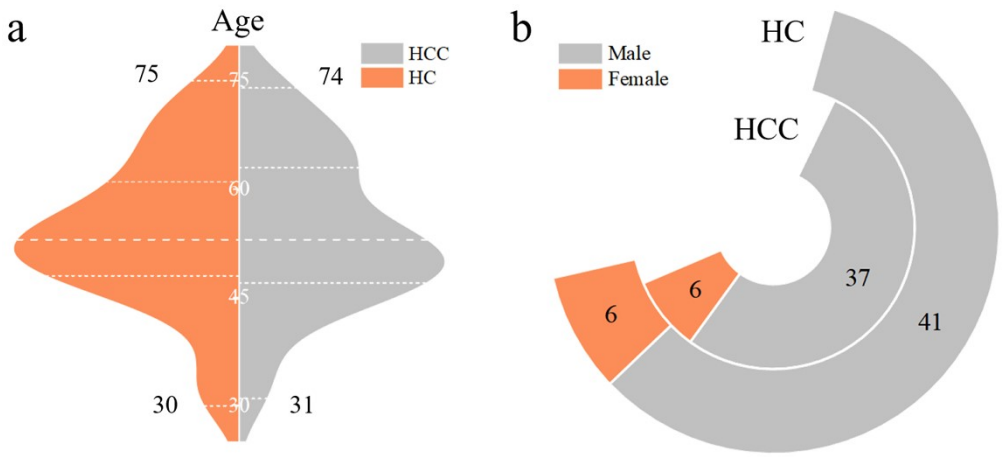


Fig. S1 The characteristics of (a) age and (b) gender in HC and HCC cohorts.

Table S1 Summary of 43 HCC patients and 47 HC participants for APGD-MS analysis of urine

Subject	Age	Sex	Subject	Age	Sex
HC	30	male	HCC	31	female
HC	32	male	HCC	33	female
HC	33	male	HCC	36	male
HC	35	male	HCC	41	male
HC	41	male	HCC	43	female
HC	43	male	HCC	45	male
HC	45	male	HCC	46	male
HC	46	male	HCC	46	male
HC	47	male	HCC	46	female
HC	47	male	HCC	46	male
HC	47	female	HCC	47	male
HC	48	male	HCC	47	male
HC	49	male	HCC	48	male
HC	49	male	HCC	49	male
HC	50	male	HCC	49	male
HC	50	male	HCC	49	male
HC	50	male	HCC	50	male
HC	50	male	HCC	51	male
HC	51	female	HCC	51	male
HC	52	female	HCC	52	male
HC	52	male	HCC	52	male
HC	53	male	HCC	53	male
HC	53	male	HCC	53	male
HC	53	male	HCC	53	male
HC	53	male	HCC	54	male
HC	54	male	HCC	55	male

HC	54	male	HCC	55	female
HC	55	male	HCC	56	male
HC	55	male	HCC	60	female
HC	56	male	HCC	61	male
HC	57	male	HCC	61	male
HC	58	male	HCC	62	male
HC	58	male	HCC	63	male
HC	60	male	HCC	65	male
HC	60	male	HCC	66	male
HC	61	male	HCC	66	male
HC	63	male	HCC	66	male
HC	63	female	HCC	67	male
HC	65	male	HCC	68	male
HC	66	male	HCC	70	male
HC	67	male	HCC	74	male
HC	68	female	HCC	74	male
HC	69	male	HCC	74	male
HC	71	female			
HC	72	male			
HC	73	male			
HC	75	male			

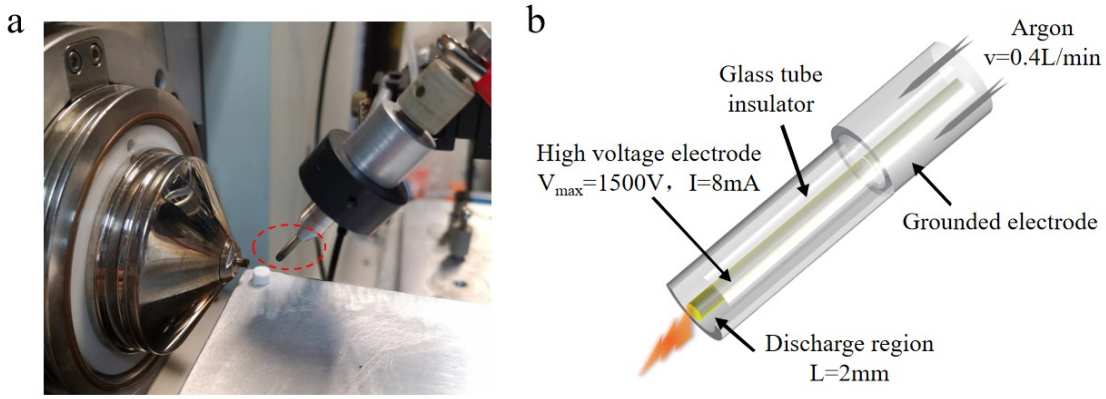
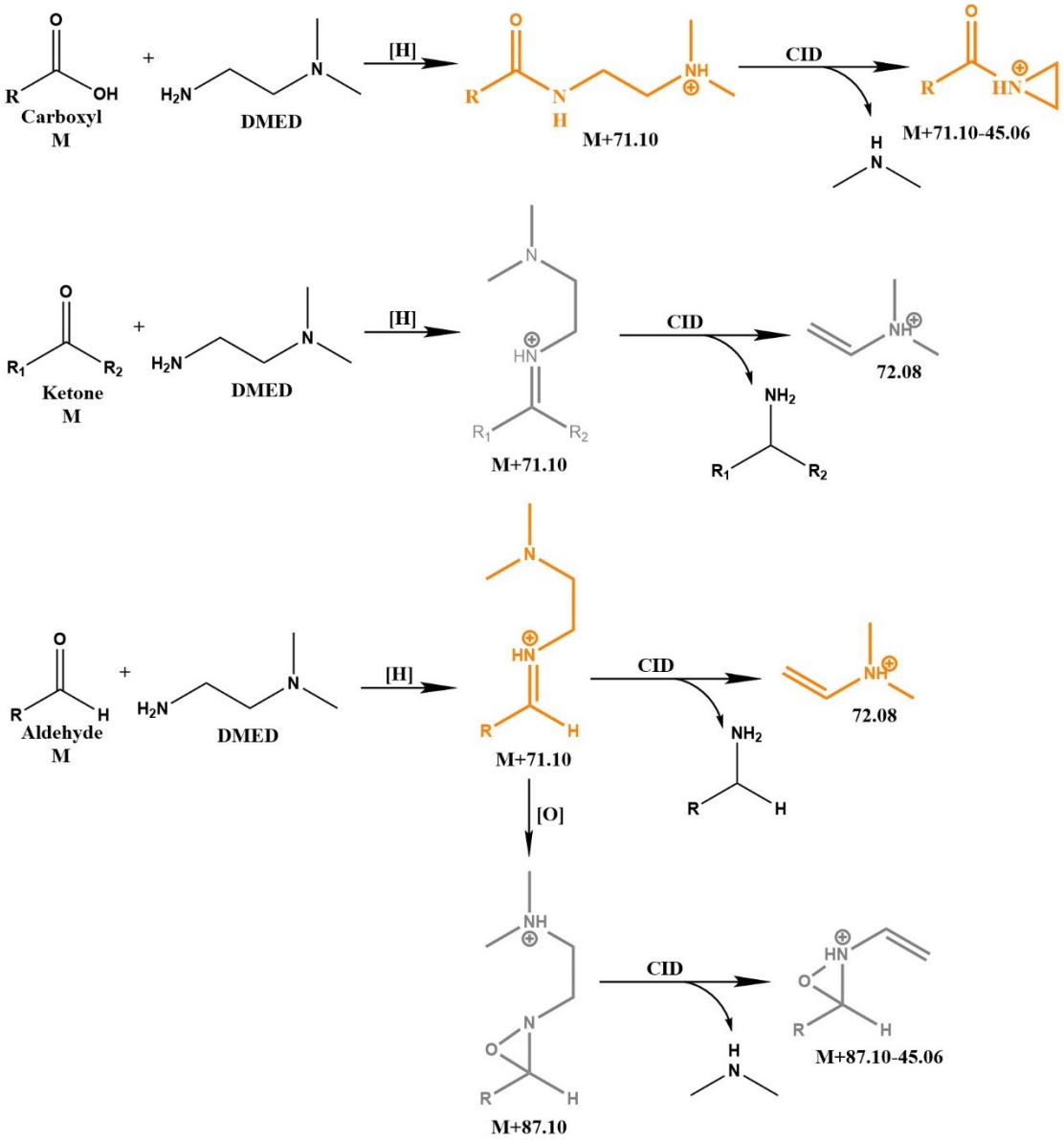


Fig. S2 (a) A photograph of the DSI platform. (b) The internal structure of APGD ion source.



Scheme S1 The derivatization reaction of carbonyl compounds with DMED and their possible fragmentation pathways.

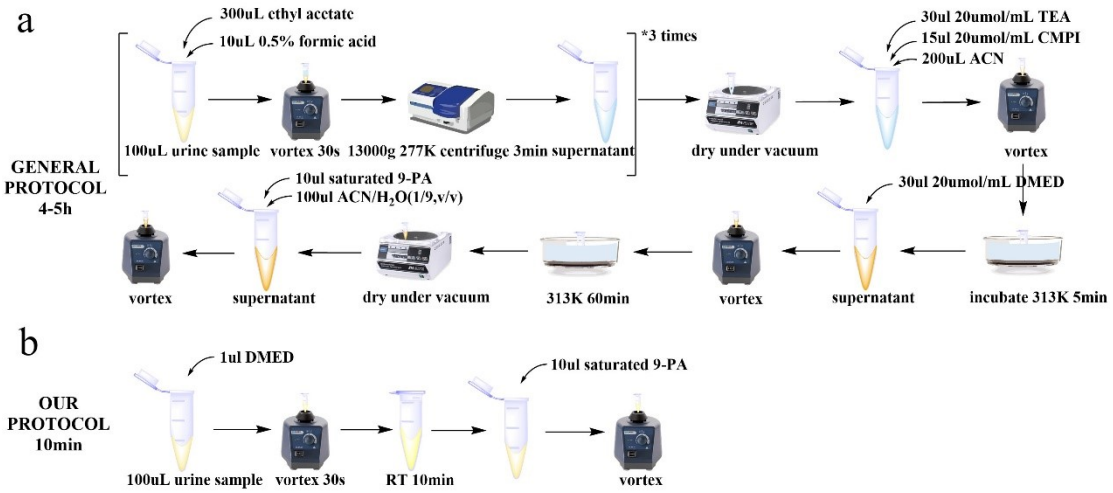


Fig. S3 Operational processes of (a) general and (b) our derivatization protocols. The processing time were 4-5 hours and 10 minutes respectively.

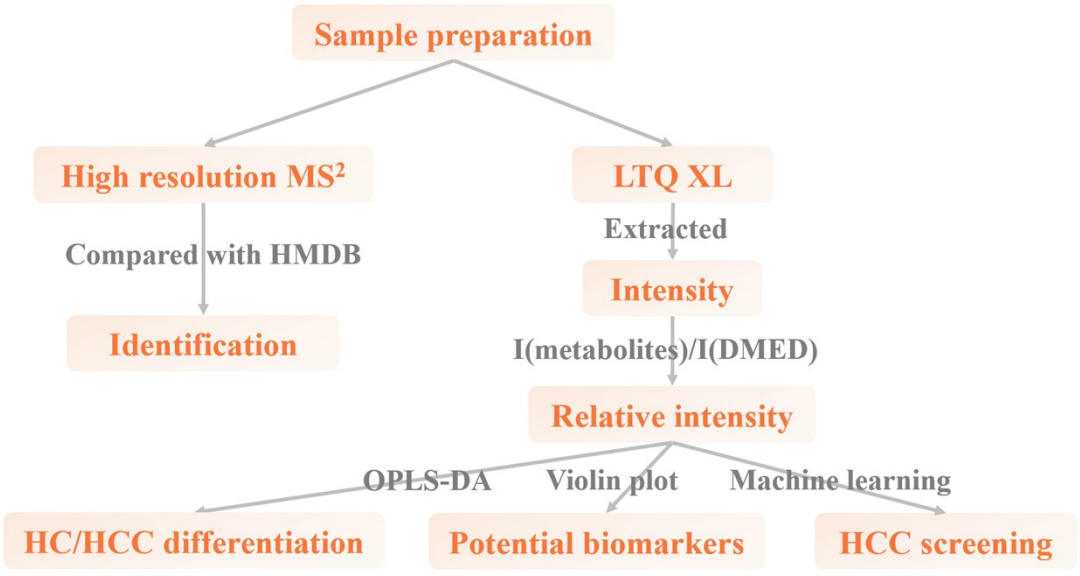


Fig. S4 The flow chart of data processing procedure.

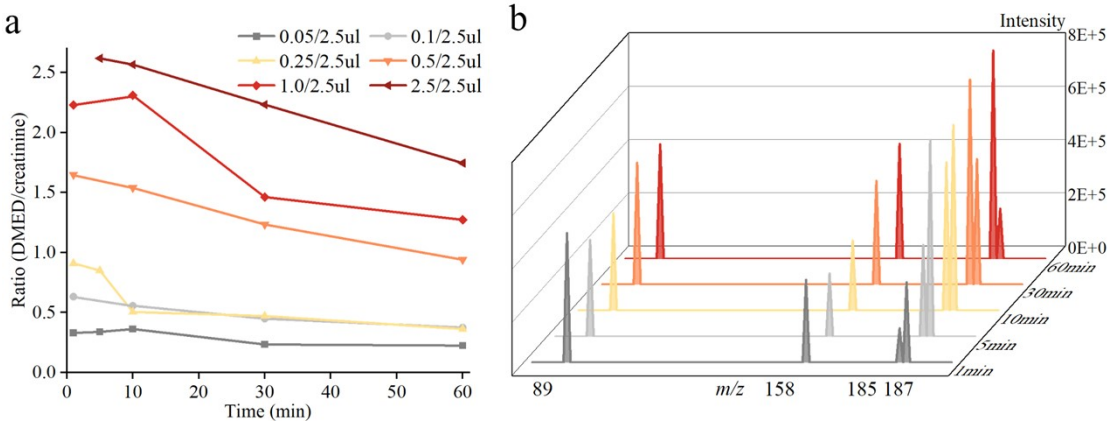


Fig. S5 Trends of intensity of DMED over time. (a) The trend of relative peak intensity of DMED under different adding volumes and derivatization times. (b) The trend of peak intensity of DMED (*m/z* 89) and its products (*m/z* 158, 185, 187) over time under the condition of adding 2.5 μL DMED to 100 μL water.

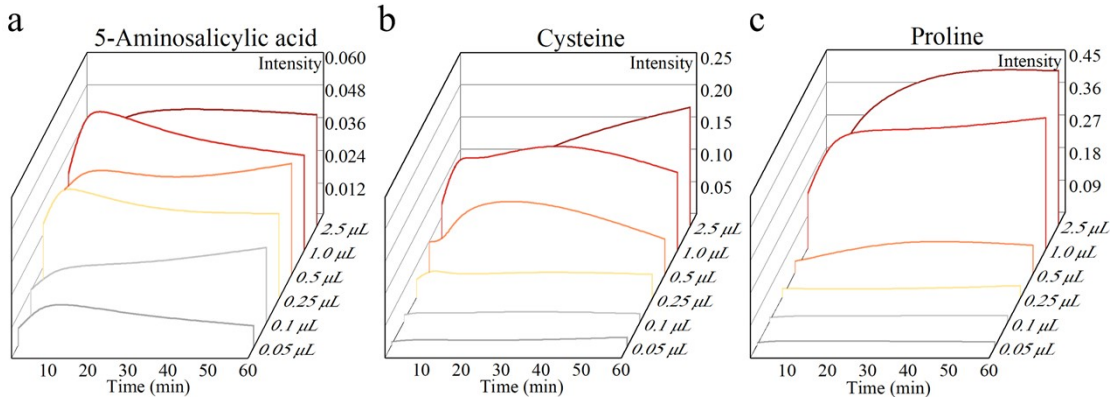


Fig. S6 The trend of relative peak intensity of (a) 5-aminosalicylic acid, (b) Cysteine and (c) proline under different adding volumes of DMED and derivatization times. The derivatization could be finished within 10 minutes and the sufficient adding volumes of DMED are 0.25, 0.5 and 1.0 μL respectively.

Table S2 Metabolites identified using different protocols by high-resolution MS.

Metabolites Identified Through Our Protocol						Metabolites Identified Through General Protocol					
No.	***m/z***	**Ion Type**	**CID**	**Classification**	**Assignment**	**No.**	***m/z***	**Ion Type**	**CID**	**Classification**	**Assignment**
	89	$[M+H]^+$		Label reagent	DMED		89	$[M+H]^+$		Label reagent	DMED
1	114	$[M+H]^+$			Creatinine†	**1**	114	$[M+H]^+$			Creatinine†
2	115	$[M\_DMED+H]^+$	m/z=72	Ketone	Acetaldehyde†	**2**	115	$[M\_DMED+H]^+$	m/z=72	Ketone	Acetaldehyde†
3	117	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Formic acid†	**3**	117	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Formic acid†
4	121	$[2M+H]^+$			Urea (Dimer)†	**4**	121	$[2M+H]^+$			Urea (Dimer)†
5	129	$[M\_DMED+H]^+$	m/z=72	Ketone	Acetone	**5**	131	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Acetic acid†
6	131	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Acetic acid†	**6**	132	$[M+H]^+$			Creatine†
7	132	$[M+H]^+$			Creatine†	**7**	157	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Crotonic acid†
8	145	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Propionic acid	**8**	161	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Lactic acid†
9	146	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Glycine	**9**	189	$[M\_DMED+H]^+$	Δm/z=-45	Acid	β-Hydroxyisovaleric acid
10	147	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Glycolic acid	**10**	193	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Benzoin acid†
11	156	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Cyanoacetic acid	**11**	200	$[M\_DMED+H]^+$	Δm/z=-45	Acid	L-Pyroglutamic acid†
12	157	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Crotonic acid†	**12**	201	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Heptanoic acid†
13	159	$[M\_DMED+H]^+$	m/z=72	Ketone	Pyranic acid	**13**	203	$[M\_DMED+H]^+$	Δm/z=-45	Acid	L-Ornithine†
14	160	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Alanine	**14**	216	$[M\_DMED+H]^+$	Δm/z=-45	Acid	N-Isobutyrylglycine
15	161	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Lactic acid†	**15**	223	$[M\_DMED+H]^+$	Δm/z=-45	Acid	4-Hydroxyphenylacetic acid
16	169	$[M\_DMED+H]^+$	Δm/z=-45	Acid	2,4-Pentadienoic acid	**16**	227	$[2M+H]^+$			Creatinine (dimer)†
17	171	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Allylacetic acid	**17**	229	$[M\_DMED+H]^+$	Δm/z=-45	Acid	4-hydroxycyclohexylacetic acid
18	173	$[M\_DMED+H]^+$	Δm/z=-45	Acid	N-Valeric Acid	**18**	231	$[M\_DMED+H]^+$	Δm/z=-45	Acid	3-Methyladipic acid
19	175	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Hydroxypyruvic acid	**19**	237	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Perillic acid†

20	183	$[M\_DMED+H]^+$	Δm/z=-45	Acid	2-Furoic acid	**20**	243	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Octendioic acid
21	184	$[M\_DMED+H]^+$	Δm/z=-45	Acid	1-pyrroline-5-carboxylic acid	**21**	244	$[M\_DMED+H]^+$	Δm/z=-45	Acid	N-Acetylglycine
22	186	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Proline	**22**	245	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Arginine†
23	188	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Guanidoacetic acid	**23**	246	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Indole-3-acetic acid
24	192	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Cysteine	**24**	250	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Hippuric acid†
25	193	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Benzoin acid†	**25**	270	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Octenoylglycine
26	194	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Nicotinic acid	**26**	271	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Decenedioic acid
27	197	$[M\_DMED+H]^+$	Δm/z=-45	Acid	5-Methylfuran-2-carboxylic acid	**27**	272	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Caprylylglycine
28	199	$[M\_DMED+H]^+$	Δm/z=-45	Acid	3-Heptenioc acid	**28**	290	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Pantothenic acid
29	200	$[M\_DMED+H]^+$	Δm/z=-45	Acid	L-Pyroglutamic acid†						
30	201	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Heptanoic acid†						
31	202	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Creatine						
32	203	$[M\_DMED+H]^+$	Δm/z=-45	Acid	L-Ornithine†						
33	209	$[M\_DMED+H]^+$	Δm/z=-45	Acid	4-Hydroxybenzioc acid						
34	217	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Glutamine						
35	224	$[M\_DMED+H]^+$	Δm/z=-45	Acid	5-Aminosalicylic acid						
36	225	$[M\_DMED+H]^+$	Δm/z=-45	Acid	2,5-Dihydroxybenzioc acid						
37	227	$[2M+H]^+$			Creatinine (dimer)†						
38	237	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Perillic acid†						
39	238	$[M\_DMED+H]^+$	m/z=72	Ketone	N-Acetyltaurine						
40	245	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Arginine†						
41	250	$[M\_DMED+H]^+$	Δm/z=-45	Acid	Hippuric acid†						

†The metabolites were identified in both protocols.

Table S3 Detailed information on metabolites only identified in general protocol.

Compound	Formula	Theoretical *m/z*	Experimental *m/z*	Adduct	Delta (ppm)
β-Hydroxyisovaleric acid	$C_5H_{10}O_3$	189.1603	189.1608	$[M\_DMED+H]^+$	2.6739
N-Isobutyrylglycine	$C_6H_{11}NO_3$	216.1712	216.1713	$[M\_DMED+H]^+$	0.4626
4-Hydroxyphenylacetic acid	$C_8H_8O_3$	223.1446	223.1448	$[M\_DMED+H]^+$	0.8963
4-hydroxycyclohexylacetic acid	$C_8H_{14}O_3$	229.1916	229.1917	$[M\_DMED+H]^+$	0.4612
3-Methyladipic acid	$C_7H_{12}O_4$	231.1709	231.1709	$[M\_DMED+H]^+$	0
Octendioic acid	$C_8H_{12}O_4$	243.1709	243.1708	$[M\_DMED+H]^+$	0.2422
N-Acetylglycine	$C_4H_7NO_3$	244.2025	244.2025	$[M\_DMED+H]^+$	0
Indole-3-acetic acid	$C_{10}H_9NO_2$	246.1606	246.1605	$[M\_DMED+H]^+$	0.4062

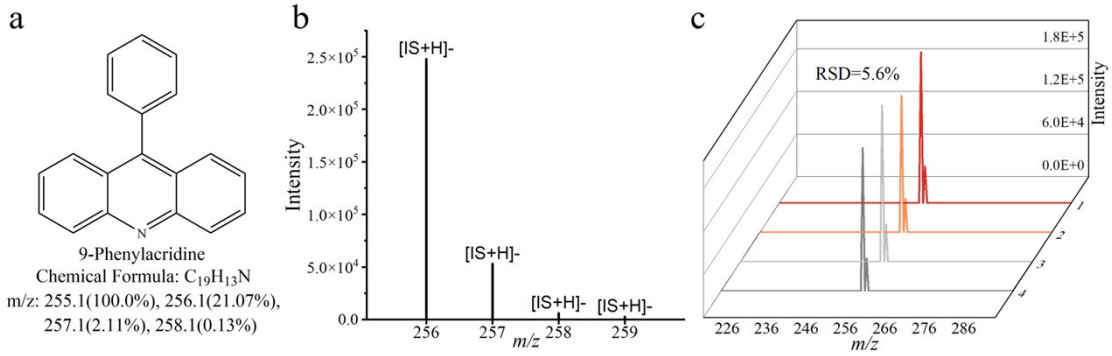


Fig. S7 (a) The structure of 9-phenylacridine. (b) The mass spectrum of 9-phenylacridine (saturated aqueous solution) in positive ion mode. (c) The repeatability of 9-PA. 100 μL water was mixed with 10 μL saturated aqueous solution of 9-PA.

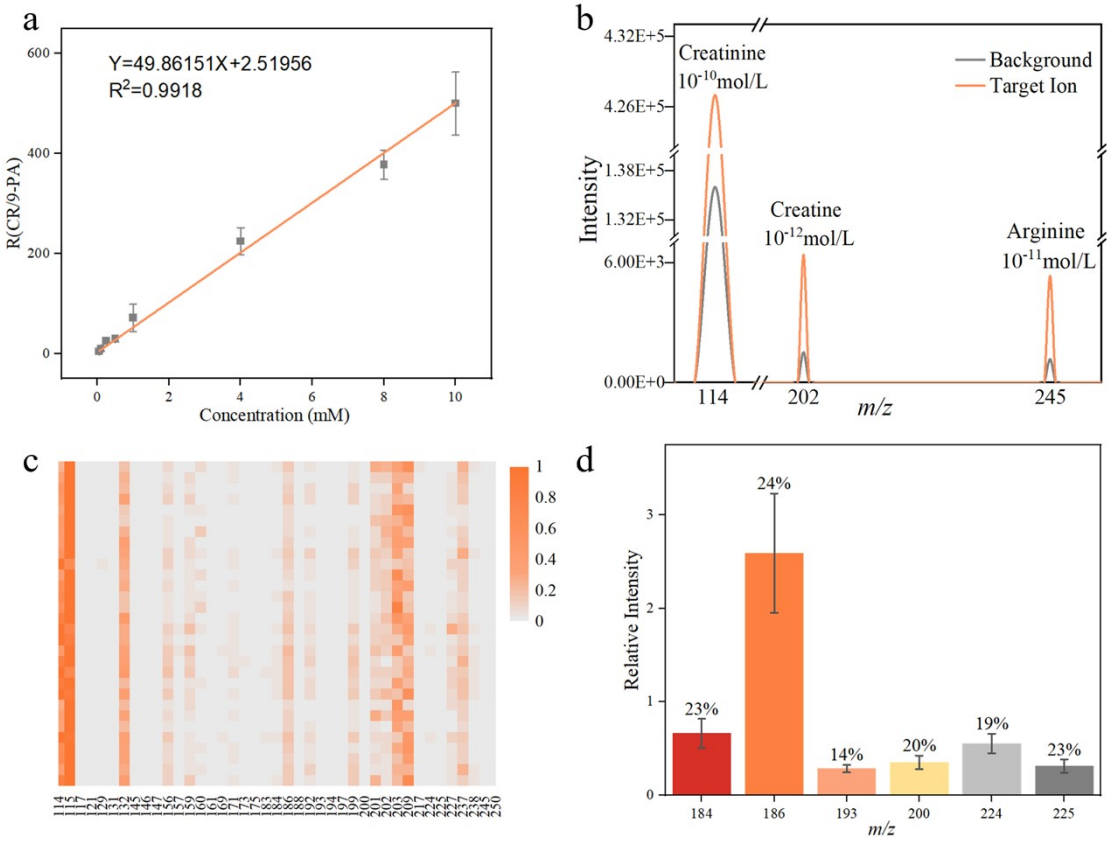


Fig. S8 Quantification study of the screening method. (a) The internal calibration curve for relative quantification of creatinine according to the ion intensity ratio of $[creatinine+H]^+$ and $[9\text{-}PA+H]^+$. 100 μL standard solution of creatinine was mixed with 10 μL saturated aqueous solution of 9-PA. (b) The LOD of creatinine, creatine and arginine. (c) The repeatability heatmap of metabolites relative peak intensities. The similarity can be found from the plot. (n=30) (d) The RSD of some typical metabolites. There were all within the acceptable range.

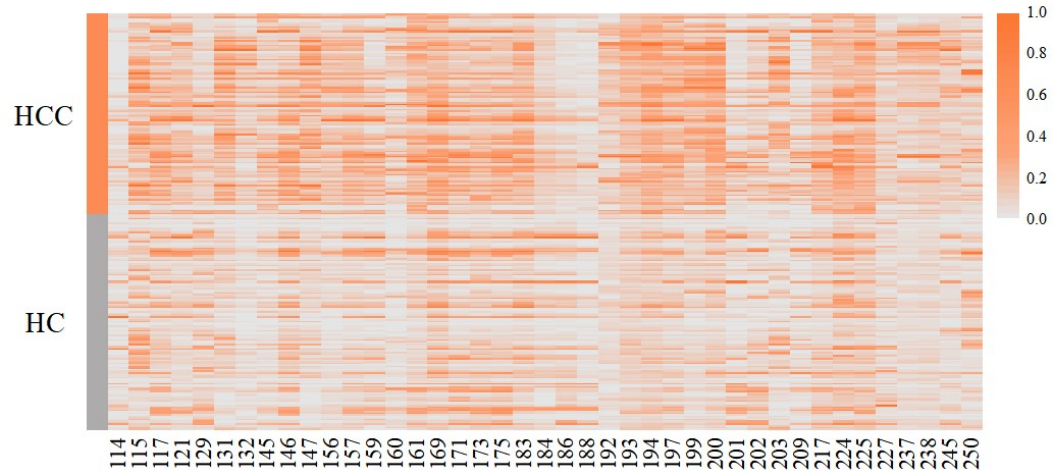


Fig. S9 Heatmap of 90 urine samples with 270 mass spectra.

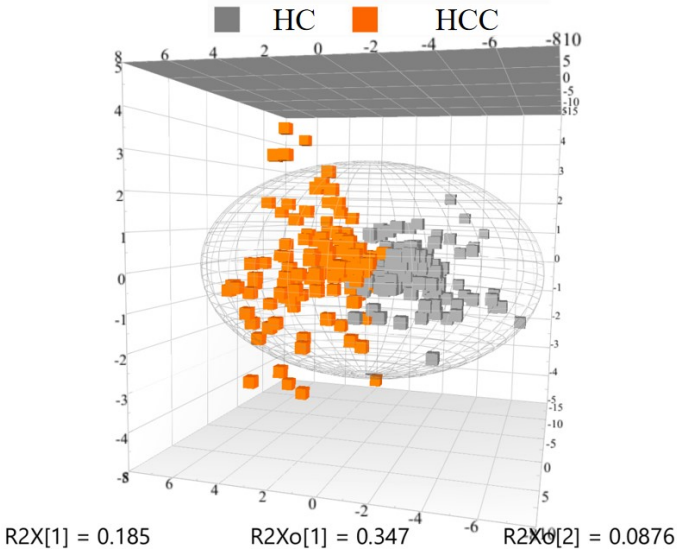


Fig. S10 The OPLS-DA plot of HCC and HC cohorts.

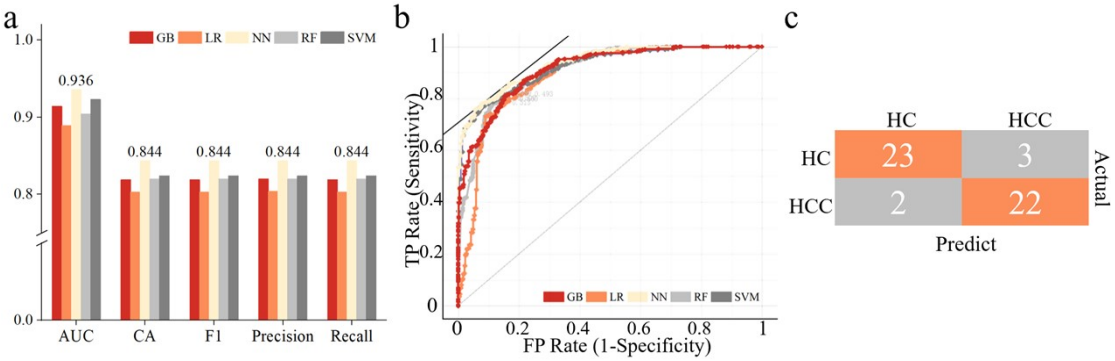


Fig. S11 Screening results of machine learning models using 11 potential urinary biomarkers of HCC as features. (a) Evaluation results of 5 different models. (b) ROC curves of 5 different models. (c) Confusion matric of the test cohort.