Supplementary Information

Rapid screening for genitourinary cancer: mass spectrometry-based metabolic fingerprinting of urine

Xiao Wang[1, 2], Yuze Li[1, 2], Jinghan Fan[2], Liuying He[1, 2], Junyu Chen[1, 2], Huihui Liu[1,2*], 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.

*Corresponding author. Email: znie@iccas.ac.cn (Z. Nie), hhliu@iccas.ac.cn (H. Liu).

This PDF file includes:

Materials and Methods

Chemical and materials

All materials and reagents were commercially available. The internal standard 3-Chloro-L-phenylalanine (L-Phe(3-Cl)-OH) (PubChem CID: 2761492) was purchased from Aladdin (Beijing, China). Matrix: 1-Naphthylhydrazine hydrochloride (NHHC, PubChem CID: 519949, from Alfa Aesar), N-(1-naphthyl) ethylenediamine dihydrochloride (NEDC, from Aladdin), 2,5-dihydroxybenzoic acid (DHB, from Alfa Aesar), α-cyano-4-hydroxycinnamic acid (CCA, from Aladdin), 9-aminoacridine (9AA, from fluka), sinapinic acid (SA, from Sigma-Aldrich). Methanol (HPLC grade) and acetonitrile (HPLC grade) were purchased from Thermo Fisher. The deionized water used in all experiments was prepared using a Milli-Q Integral water purification system from Millipore (Milford, MA, U.S.A.). All other solvents and chemicals were of analytical grade.

Urine sample collection

This study was carried out in accordance with relevant ethical regulations and guidelines of the Hospital Scientific Research Ethics Committee of First Affiliated Hospital of Gannan Medical University. All the BC and PC patients were diagnosed by histopathological examination in hospital. The fasting morning urine samples of BC patients and HC volunteers were obtained from the First Affiliated Hospital of Gannan Medical University and stored at -80℃ until analysis. The urine samples of PC patients were obtained from Beijing Hospital in the same procedure. In this study, 38 BC, 39 PC, and 40 HC urine samples were enrolled. The detailed information was summarized in Table S3 and Table S6. No more pretreatment except unfreezing the samples and mixed with matrix and internal standard was used during the experiment.

To explore the metabolic differences between cancer patients and healthy control volunteers, we divided all subjects into four cohorts for comparison: GU (BC+PC) vs HC, BC vs PC, BC vs HC, and PC vs HC. To ensure that meaningful comparisons could be made between the two groups within each cohort, we need that there is no statistical significance in the distribution of age and gender.

Matrix selection for MALDI-MS

To get the optimal signals for better mining diagnostic information from urine samples, we first examined the effectiveness of several common small-molecule matrices for urine analysis, including 2,5-dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CCA), and sinapinic acid (SA) in positive ion mode, while 1-naphthylhydrazine hydrochloride (NHHC), N-(1-naphthyl) ethylenediamine dihydrochloride (NEDC), and 9-aminoacridine (9AA) in negative ion mode. Urine samples from HC and GU cancer patients were tested respectively for evaluating the performance of matrices. As shown in Fig. S1, we found that SA in positive ion mode and NHHC in negative ion mode helped mass spectrometer capture the richest signals. However, the targeted signals were buried under strong background noise of SA (Fig. S1e), and therefore identification became difficult. Besides, due to its strong ultraviolet absorption and low background interference in the low molecular weight region, NHHC can sensitively analyse small molecules such as homogentisic acid in the urine, even in saturated NaCl solution according to the previous work of our group[1]. NHHC is a hydrochloride, whose crystal size is smaller, the "sweet spot" effect is weaker and the signal intensity is more stable. Moreover, urinary metabolites contain a large number of small-molecule organic acids [2], which are prone to be negatively charged in MS, so that using negative ion mode can obtain more signals of them. Eventually, we chose NHHC as the matrix of MALDI-MS for urine sample metabolic analysis in negative ion mode.

Mass spectrometry analysis

3-Chloro-L-phenylalanine was introduced as the internal standard for quantification use in urine MALDI MS. 1 μL urine sample was mixed with 1 μL 10mg/mL NHHC (50%MeOH) and 1 μL of internal standard aqueous solution, then 1 μL of the mixture was deposited on a MALDI plate and allowed to dry at room temperature for consequent MALDI- MS analysis.

Mass spectra of urine samples were collected in negative ion mode over the mass range of 0–1000 Da. A total of 351 mass spectra collected were summarized and *m/z* were aligned. All the data were normalized using the signal intensity of $[IS\text{-}H]^-$ (*m/z* 198.0). The number of occurrences of each *m/z* in all spectra was calculated, and those less than 234 (351*2/3) were eliminated. Subsequently, the matrix background was checked and tested 30 times in parallel,

and the *m/z* features with the top eight highest abundance were eliminated. And then 318 *m/z* features were obtained, which were used as a basis for subsequent statistical analysis.

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was performed on an Ultraflextreme MALDI-TOF/TOF-MS (Bruker Daltonics, Billerica, MA) equipped with a smartbeam (Nd: YAG 355 nm) laser. Oligosaccharides were used for mass calibration in negative ion mode before each run, including D-glucose (MW: 180.16), maltose (MW: 342.30), 1-kestose (MW: 504.4), nystose (MW: 666.6), 1,1,1-kestopentaose (MW: 828.7), fructo-oligosaccharide DP6 (MW: 990.86). 1-Naphthylhydrazine hydrochloride (NHHC, 10mg/mL in 50%MeOH) was used as a matrix. The mass spectra were recorded at an accelerating voltage of 19 kV, a reflection voltage of 20 kV, and the laser spot and laser pulse energy were adjusted to the optimal condition to obtain the best signal. Data acquisition was carried out using the FlexImaging 4.0 software provided by Bruker Daltonics.

Matrix-assisted laser desorption/ionization fourier transform ion cyclotron resonance mass spectrometer (MALDI-FTICR-MS, SolariX; 15 T, Bruker Daltonics, Bremen, Germany) was utilized to accurate mass measurements and chemical formula assignments of each peak in negative ion mode.

LC-MS/MS was utilized to verify the structures of some of the metabolites. Instruments: ultra-high pressure liquid phase (UltiMate 3000, THERMO), high-resolution mass spectrometry (5600 QTOF, AB SCIEX), chromatographic column (ACQUITY UPLC HSS T3 1.8 μ m 2.1 × 100 mm, Waters).

Statistical analysis

The two-tailed test was performed to calculate the p values with FDR-corrected using the MetaboAnalyst 5.0. FC analysis and OPLS-DA were also performed using the MetaboAnalyst 5.0. Heatmaps were constructed using the OmicStudio tools at https://www.omicstudio.cn/tool. Machine learning analysis was conducted using Orange (Version 3.27.1).

Equations

The AUC, Accuracy, F1-score, Precision, and Recall were used to evaluate the classifier, which could be calculated with true positive (TP), false positive (FP), true negative (TN), and

false negative (FN). The F1-score is a measure of a model's accuracy on a dataset, which is used to evaluate binary classification systems. The formula for the standard F1-score is the harmonic mean of the precision and recall.

The equations are as follows:

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

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

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

$$Recall = Sensitivity = \frac{TP}{TP + FN}$$

Ethics statement

All experiments in this study were performed in compliance with relevant laws or guidelines. This study was carried out following relevant ethical regulations and guidelines of the Hospital Scientific Research Ethics Committee of First Affiliated Hospital of Gannan Medical University. This project was reviewed by the Hospital Scientific Research Ethics Committee of First Affiliated Hospital of Gannan Medical University and agreed to apply for research projects in line with the relevant provisions on biological human trials in the "measures for Ethical Review of Biomedical Research involving Human beings" of the Health Commission and the relevant provisions of the Ministry of Science and Technology's "guiding opinions on being kind to Experimental Animals". And informed consent was obtained from all human subjects.

Data availability

The data that support the findings of this study and Python codes are available from the corresponding author upon reasonable request.

Author contributions

Z.N. initiated the proposal; X.W., Y.L., H.L. and Z.N. designed research; X.W., J.F. performed research; X.W., L.H., and J.C. analyzed data; X.W., H.L. and Z.N. wrote the paper.

Acknowledgements

This work was supported by grants from the National Natural Sciences Foundation of China (Grant Nos. 21625504, 21827807, and 21790390/21790392) and Chinese Academy of Sciences.

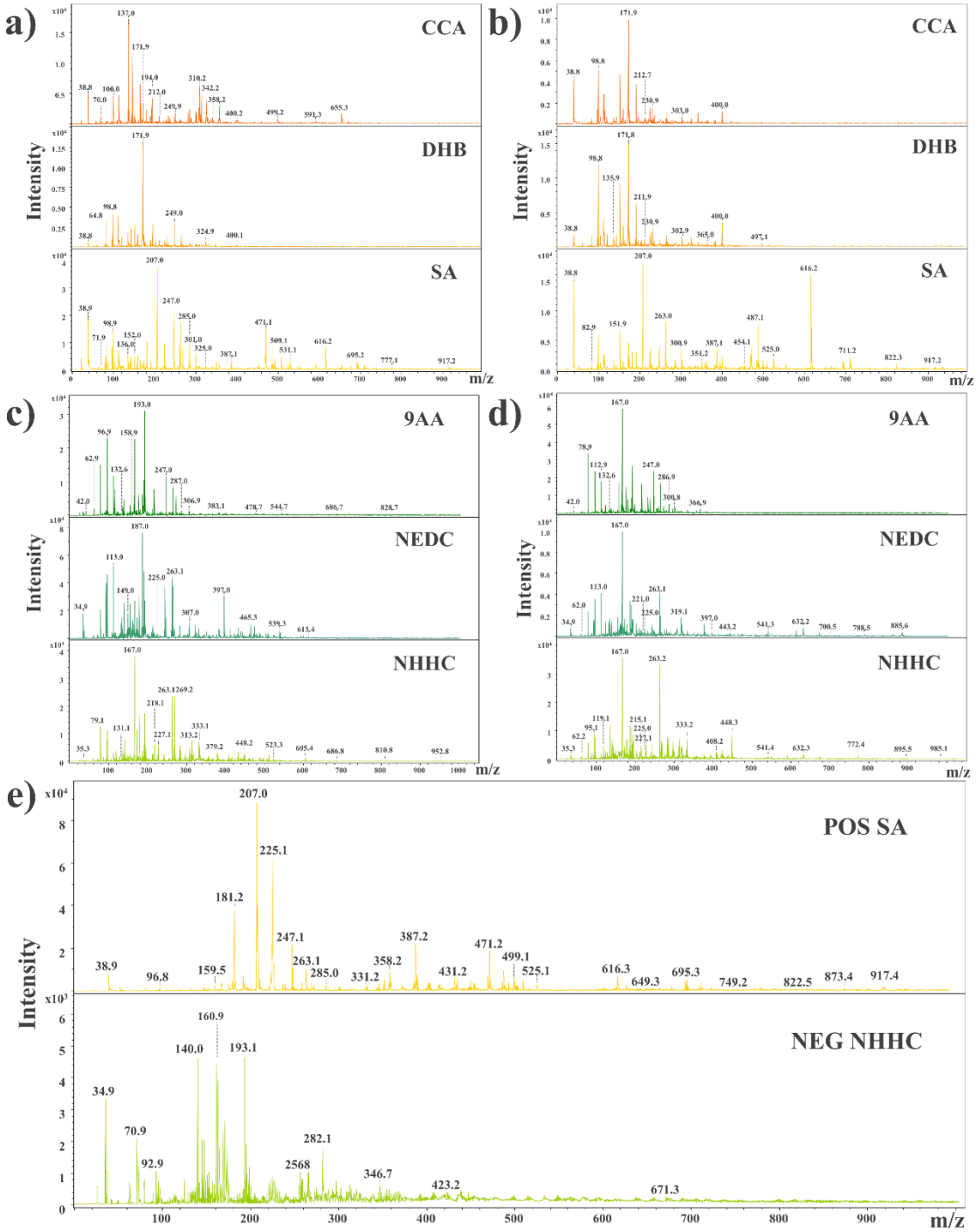

Fig. S1 Comparison of different matrices. Mass spectra of urine samples from **a)** healthy control (HC) individuals and **b)** genitourinary (GU) cancer patients using CCA, DHB, and SA in positive ion mode. Mass spectra of urine samples from **c)** healthy control (HC) individuals and **d)** genitourinary (GU) cancers using 9AA, NEDC, and NHHC in negative ion mode. **e)** The background mass spectra of SA (10mg/mL, 50% MeOH) in positive ion mode and NHHC (10mg/mL, 50% MeOH) in negative ion mode, respectively.

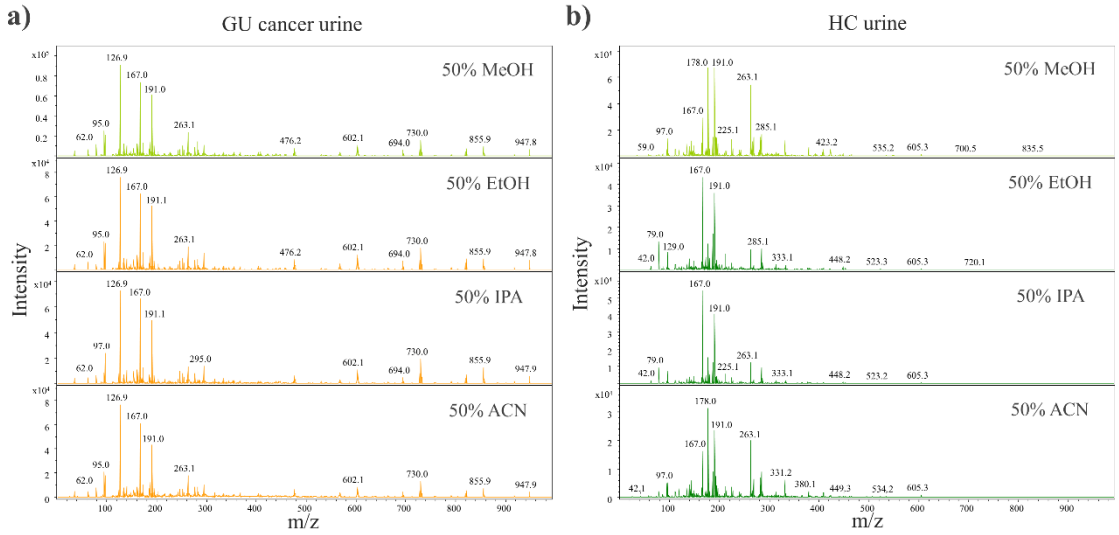

Fig. S2 Comparison of different solvents for preparing NHHC solutions. Mass spectrum of urine samples from **a)** genitourinary (GU) cancers and **b)** healthy control (HC) individuals using methanol, ethanol, isopropanol, acetonitrile, and water as solvents of NHHC matrices solutions in negative ion mode.

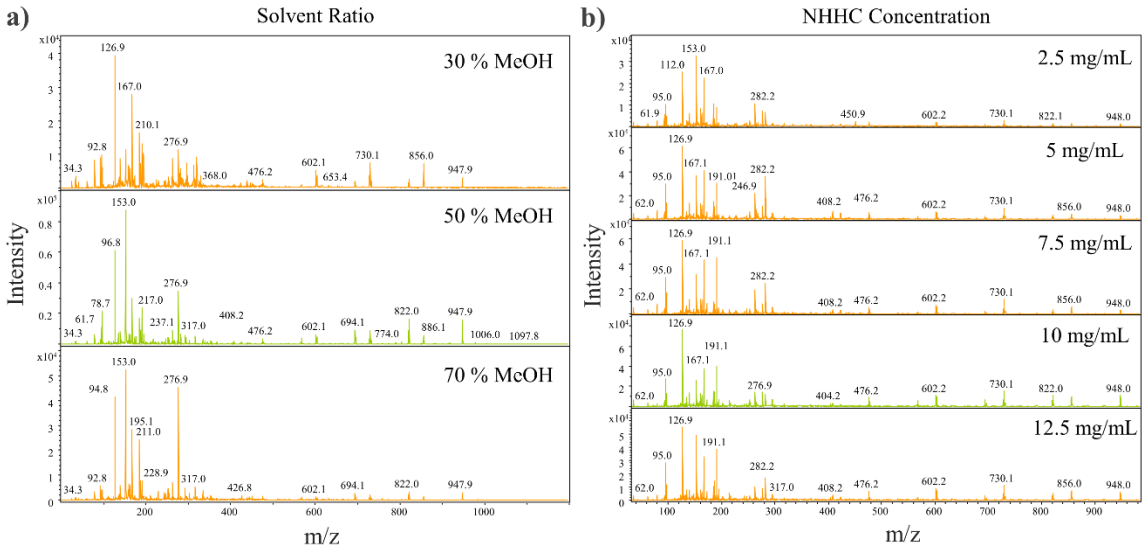

Fig. S3 Comparison of different ratios of MeOH/H_2O and different NHHC concentrations. Mass spectrum of a randomly selected GU urine sample using different **a)** ratio of MeOH/H_2O and **b)** NHHC concentrations (50% MeOH as solvent).

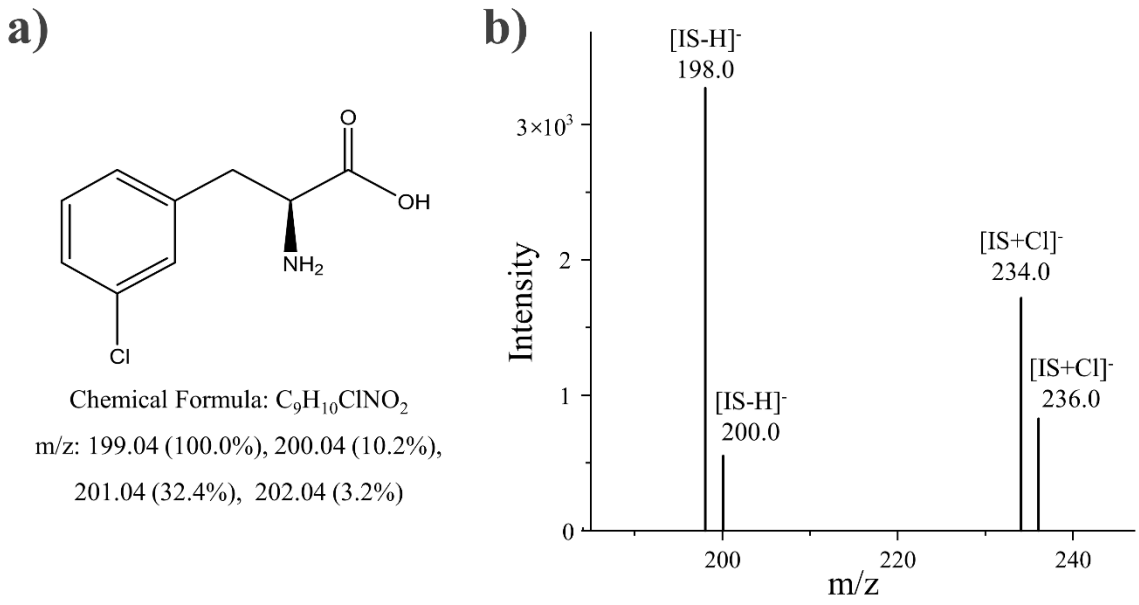

Fig. S4 a) Structure of L-Phe(3-Cl)-OH. **b)** Mass spectrum of 8mM L-Phe(3-Cl)-OH (50% MeOH) in negative ion mode. In negative ion mode mass spectrometry analysis, L-Phe(3-Cl)-OH was detected at *m/z* 198.0 / 200.0 as $[M-H]^-$ and *m/z* 234.0 / 236.0 as $[M+Cl]^-$ respectively.

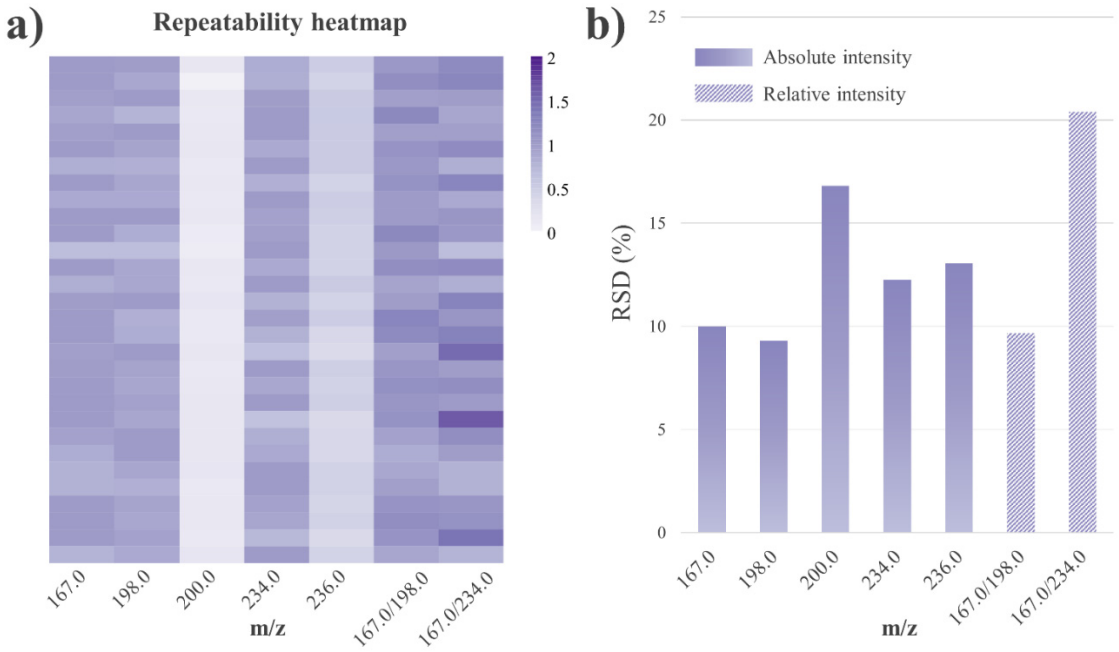

Fig. S5 The relatively quantitative ability of internal standard. **a)** Repeatability heatmap of uric acid (UA) and internal standard (IS) signal intensities. **b)** The relative standard deviation (RSD) of signal intensities. *m/z* 167.0: $[UA-H]^-$, *m/z* 198.0 / 200.0: $[IS-H]^-$, *m/z* 234.0 / 236.0: $[IS+Cl]^-$. This experiment was repeated 30 times.

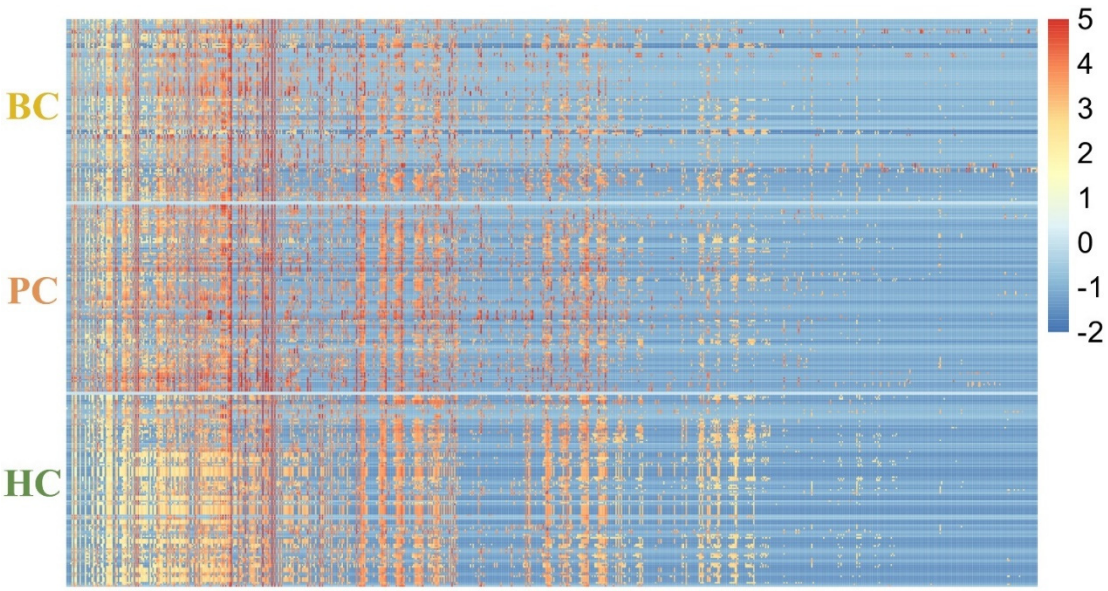

Fig. S6 Heatmap of 117 urine samples (38 BC, 39 PC, and 40 HC) with 351 mass spectra data.

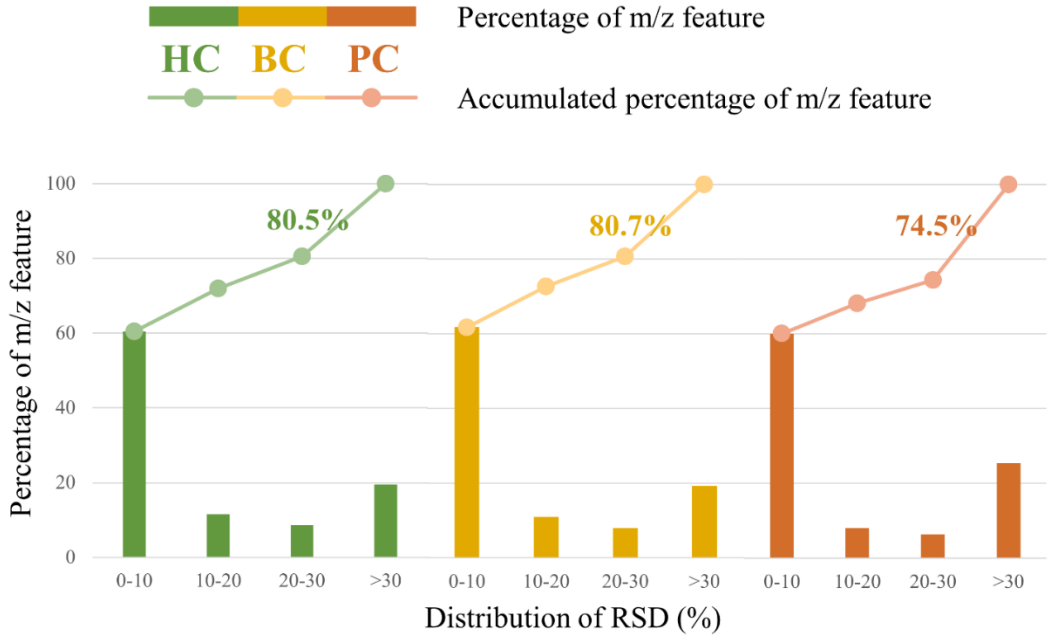

Fig. S7 RSD distributions of *m/z* features in BC, PC, and HC groups.

Table S1. Metabolites in urine detected in negative ion mode using MALDI-FTICR-MS.

Name	Formula	Theoretical *m/z*	Experimental *m/z*	Delta (ppm)	Adduct	HMDB ID
Ribonic acid	$C_5H_{10}O_6$	165.04046	165.041	3.272	$[M-H]^-$	HMDB0000867
Quinolinic acid‡	$C_7H_5NO_4$	166.01458	166.013	9.517	$[M-H]^-$	HMDB0000232
Uric acid*‡	$C_5H_4N_4O_3$	167.02106	167.021	0.359	$[M-H]^-$	HMDB0000289
Citric acid*†‡	$C_6H_8O_7$	173.00861	173.009	2.254	$[M-H_2O-H]^-$	HMDB0000094
		191.01973	191.020	1.413	$[M-H]^-$	
Ascorbic acid*†‡	$C_6H_8O_6$	175.02481	175.025	1.086	$[M-H]^-$	HMDB0000044
Hippuric acid†‡	$C_9H_9NO_3$	178.05097	178.051	0.168	$[M-H]^-$	HMDB0000714
Glucose*†‡	$C_6H_{12}O_6$	179.05611	179.056	0.614	$[M-H]^-$	HMDB0000122
Inositol*†‡						HMDB0000211
1-Methyluric acid‡	$C_6H_6N_4O_3$	181.03671	181.037	1.602	$[M-H]^-$	HMDB0003099
Hydroxyphenyllactic acid‡	$C_9H_{10}O_4$	181.05063	181.051	2.044	$[M-H]^-$	HMDB0000755
L-Tryptophan‡	$C_{11}H_{12}N_2O_2$	185.07149	185.072	2.756	$[M-H_2O-H]^-$	HMDB0000929
p-Tolyl Sulfate‡	$C_7H_8O_4S$	187.00705	187.007	0.267	$[M-H]^-$	/
3-Hydroxyanthranilic acid†‡	$C_7H_7NO_3$	188.01200	188.010	10.638	$[M+Cl]^-$	HMDB0001476
2-O-Methylascorbic acid†	$C_7H_{10}O_6$	189.04046	189.040	2.433	$[M-H]^-$	HMDB0240294
Glucuronic acid*‡	$C_6H_{10}O_7$	193.03538	193.035	1.969	$[M-H]^-$	HMDB0000127
3-Hydroxyhippuric acid*‡	$C_9H_9NO_4$	194.04588	194.046	0.618	$[M-H]^-$	HMDB0006116
4-Hydroxyhippuric acid*‡					$[M-H]^-$	HMDB0013678
Salicyluric acid*‡					$[M-H]^-$	HMDB0000840

Gluconic acid*‡	$C_6H_{12}O_7$	195.05103	195.051	0.154	$[M-H]^-$	HMDB0000625
Tryptophol†‡	$C_{10}H_{11}NO$	196.05347	196.054	2.703	$[M+Cl]^-$	HMDB0003447
6-amino-5[N-methylformylamino]-1-methyluracil†	$C_7H_{10}N_4O_3$	197.06801	197.068	0.051	$[M-H]^-$	HMDB0059771
O-methoxycatechol-O-sulphate	$C_7H_8O_5S$	203.00197	203.002	0.148	$[M-H]^-$	HMDB0060013
2-Methylcitric acid†	$C_7H_{10}O_7$	205.03538	205.035	1.853	$[M-H]^-$	HMDB0000379
5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone*	$C_{11}H_{12}O_4$	207.06628	207.066	1.352	$[M-H]^-$	HMDB0029185
N-Benzyloxycarbonylglycine‡	$C_{10}H_{11}NO_4$	208.06153	208.062	2.259	$[M-H]^-$	HMDB0000735
Indoxyl sulfate†‡	$C_8H_7NO_4S$	212.00230	212.002	1.415	$[M-H]^-$	HMDB0000682
8-Hydroxy-7-methylguanine†	$C_6H_7N_5O_2$	216.02938	216.028	6.388	$[M+Cl]^-$	HMDB0006037
Porphobilinogen†‡	$C_{10}H_{14}N_2O_4$	225.08808	225.088	0.355	$[M-H]^-$	HMDB0000245
Prolylhydroxyproline†‡	$C_{10}H_{16}N_2O_4$	227.10373	227.104	1.189	$[M-H]^-$	HMDB0006695
Paracetamol sulfate	$C_8H_9NO_5S$	230.01287	230.013	0.565	$[M-H]^-$	HMDB0059911
methyl 3-(1H-indol-3-yl) propanoate‡	$C_{12}H_{13}NO_2$	238.06403	238.066	8.275	$[M+Cl]^-$	HMDB0240623
Ibuprofen†‡	$C_{13}H_{18}O_2$	241.10008	241.098	8.627	$[M+Cl]^-$	HMDB0001925
Pseudouridine‡	$C_9H_{12}N_2O_6$	243.06226	243.062	1.070	$[M-H]^-$	HMDB0000767
Uridine‡					$[M-H]^-$	HMDB0000296
N-acetyltryptophan‡	$C_{13}H_{14}N_2O_3$	245.09317	245.093	0.694	$[M-H]^-$	HMDB0013713
Propenoylcarnitine*	$C_{10}H_{17}NO_4$	250.08516	250.084	4.638	$[M+Cl]^-$	HMDB0013124
Ribothymidine*‡	$C_{10}H_{14}N_2O_6$	257.07791	257.078	0.350	$[M-H]^-$	HMDB0000884
3-Methyluridine*‡						HMDB0004813

3-hydroxy-3-(3-hydroxyphenyl)propanoic acid-O-sulphate	$C_9H_{10}O_7S$	261.00745	261.007	1.724	$[M-H]^-$	HMDB0059967
Phenylacetylglutamine‡	$C_{13}H_{16}N_2O_4$	263.10373	263.104	1.026	$[M-H]^-$	HMDB0006344
Melatonin‡	$C_{13}H_{16}N_2O_2$	267.09058	267.093	9.061	$[M+Cl]^-$	HMDB0001389
2-(4-hydroxyphenyl)chromenylium-3,5,7-triol†	$C_{15}H_{11}O_5$	269.04610	269.046	0.372	$[M-H]^-$	HMDB0003263
Aspartylphenylalanine*‡	$C_{13}H_{16}N_2O_5$	279.09865	279.099	1.254	$[M-H]^-$	HMDB0000706
S-3-oxodecanoyl cysteamine	$C_{12}H_{23}NO_2S$	280.11435	280.113	4.819	$[M+Cl]^-$	HMDB0059773
(R)-3-hydroxybutyrylcarnitine	$C_{11}H_{21}NO_5$	282.11138	282.112	2.198	$[M+Cl]^-$	HMDB0062735
p-Cresol glucuronide	$C_{13}H_{16}O_7$	283.08233	283.082	1.166	$[M-H]^-$	HMDB0011686
7-Hydroxyetodolac† 6-Hydroxyetodolac†	$C_{17}H_{21}NO_4$	284.12867	284.127	5.878	$[M-H_2O-H]^-$	HMDB0060544 HMDB0060543
Diphenol glucuronide†	$C_{12}H_{14}O_8$	285.06159	285.062	1.438	$[M-H]^-$	HMDB0059998
5'-(3',4'-Dihydroxyphenyl)-gamma-valerolactone sulfate	$C_{11}H_{12}O_7S$	287.02310	287.023	0.348	$[M-H]^-$	HMDB0029191
Canavaninosuccinate	$C_9H_{16}N_4O_7$	291.09462	291.093	5.565	$[M-H]^-$	HMDB0012197
Didemethylcitalopram	$C_{18}H_{17}FN_2O$	295.12522	295.124	4.134	$[M-H]^-$	HMDB0060472
5'-Methylthioadenosine*‡	$C_{11}H_{15}N_5O_3S$	296.08228	296.080	7.701	$[M-H]^-$	HMDB0001173
7C-aglycone†	$C_{18}H_{18}O_4$	297.11323	297.112	4.140	$[M-H]^-$	HMDB0004808
Nilutamide‡	$C_{12}H_{10}F_3N_3O_4$	298.04395	298.043	3.187	$[M-H_2O-H]^-$	HMDB0014803
beta-D-3-Ribofuranosyluric acid	$C_{10}H_{12}N_4O_7$	299.06332	299.063	1.070	$[M-H]^-$	HMDB0029920
N-Acetylaspartylglutamic acid*†‡	$C_{11}H_{16}N_2O_8$	303.08339	303.083	1.287	$[M-H]^-$	HMDB0001067

4-Hydroxy-5-(dihydroxyphenyl)-valeric acid-O-sulphate	$C_{11}H_{14}O_8S$	305.03366	305.034	1.115	$[M-H]^-$	HMDB0059978
Resveratrol 3-sulfate†	$C_{14}H_{12}O_6S$	307.02818	307.028	0.586	$[M-H]^-$	HMDB0041772
N-Acetylneuraminic acid*‡	$C_{11}H_{19}NO_9$	308.09871	308.099	0.941	$[M-H]^-$	HMDB0000230
N2,N2-Dimethylguanosine‡	$C_{12}H_{17}N_5O_5$	310.11569	310.116	1.000	$[M-H]^-$	HMDB0004824
Beta-D-Glucopyranuronic acid†	$C_{13}H_{14}O_9$	313.05651	313.057	1.565	$[M-H]^-$	HMDB0010314
Nicotine glucuronide†	$C_{16}H_{22}N_2O_6$	319.12940	319.130	1.880	$[M-H_2O-H]^-$	HMDB0001272
N4-Acetylcytidine‡	$C_{11}H_{15}N_3O_6$	320.06549	320.063	7.780	$[M+Cl]^-$	HMDB0005923
trans-isoeugenol-O-glucuronide*	$C_{16}H_{20}O_8$	321.09743	321.096	4.453	$[M-H_2O-H]^-$	HMDB0060021
Gliclazide‡	$C_{15}H_{21}N_3O_3S$	322.12309	322.124	2.825	$[M-H]^-$	HMDB0015252
Galactosylhydroxylysine*	$C_{12}H_{24}N_2O_8$	323.14599	323.146	0.031	$[M-H]^-$	HMDB0000600
Dihyroxy-1H-indole glucuronide I	$C_{14}H_{15}NO_8$	324.07249	324.073	1.574	$[M-H]^-$	HMDB0059997
Hydroxyhexamide*	$C_{15}H_{22}N_2O_4S$	325.12275 361.09943	325.122 361.102	2.307 7.117	$[M-H]^-$ $[M+Cl]^-$	HMDB0060610
Acetaminophen glucuronide†	$C_{14}H_{17}NO_8$	326.08814	326.088	0.429	$[M-H]^-$	HMDB0010316
mono(2-ethyl-5-hydroxyhexyl) phthalate†	$C_{16}H_{22}O_5$	329.11613	329.114	6.472	$[M+Cl]^-$	HMDB0094679
Neomenthol-glucuronide†	$C_{16}H_{28}O_7$	331.17623	331.176	0.694	$[M-H]^-$	HMDB0060012
4'-O-Methyl-(-)-epicatechin-7-O-sulphate 4'-O-Methyl-(-)-epicatechin-5-O-sulphate 3'-O-Methyl-(-)-epicatechin-5-O-sulphate	$C_{16}H_{16}O_7S$	333.04328	333.044	2.162	$[M-H_2O-H]^-$	HMDB0029184 HMDB0029182 HMDB0029176
4-Hydroxy-5-(dihydroxyphenyl)-valerate-O-methyl-O-sulfate*†	$C_{12}H_{16}O_9S$	335.04423	335.043	3.671	$[M-H]^-$	HMDB0059977
Enterodiol‡	$C_{18}H_{22}O_4$	337.12121	337.122	2.343	$[M+Cl]^-$	HMDB0005056

11-Ketoetiocholanolone	$C_{19}H_{28}O_3$	339.17325	339.172	3.685	$[M+Cl]^-$	HMDB0006031
Epinephrine glucuronide*	$C_{15}H_{21}NO_9$	340.10324	340.104	2.235	$[M-H_2O-H]^-$	HMDB0010336
Trehalose*†‡	$C_{12}H_{22}O_{11}$	341.10843 377.08561	341.108 377.086	1.261 1.034	$[M-H]^-$ $[M+Cl]^-$	HMDB0000975
Testosterone sulfate	$C_{19}H_{28}O_5S$	349.14736	349.146	3.895	$[M-H_2O-H]^-$	HMDB0002833
Chlorogenic acid*	$C_{16}H_{18}O_9$	353.08781	353.091	9.035	$[M-H]^-$	HMDB0003164
Succinylaminoimidazole carboxamide riboside	$C_{13}H_{18}N_4O_9$	355.08899	355.088	2.788	$[M-H_2O-H]^-$	HMDB0240295
Riboflavin*†‡	$C_{17}H_{20}N_4O_6$	357.11989	357.119	2.492	$[M-H_2O-H]^-$	HMDB0000244
3-Methoxy-4-hydroxyphenylglycol glucuronide	$C_{15}H_{20}O_{10}$	359.09837	359.099	1.754	$[M-H]^-$	HMDB0000496
2'-alpha-Mannosyl-L-tryptophan*	$C_{17}H_{22}N_2O_7$	401.11210	401.109	7.729	$[M+Cl]^-$	HMDB0240296
Androsterone sulfate†‡	$C_{19}H_{30}O_5S$	369.17412	369.174	0.325	$[M-H]^-$	HMDB0002759
Dihydroferulic acid 4-O-glucuronide*†	$C_{16}H_{20}O_{10}$	371.09837	371.099	1.698	$[M-H]^-$	HMDB0041723
Casticin‡	$C_{19}H_{18}O_8$	373.09289	373.090	7.746	$[M-H]^-$	HMDB0030660
D-Maltose	$C_{12}H_{22}O_{11}$	377.08561	377.086	1.034	$[M+Cl]^-$	HMDB0000163
hesperetin 3'-O-sulfate	$C_{16}H_{14}O_9S$	381.02858	381.030	3.727	$[M-H]^-$	HMDB0029202
UDP-D-Xylose†	$C_{16}H_{20}N_2O_7$	387.09645	387.093	8.913	$[M+Cl]^-$	HMDB0001013
N-Acetylserotonin glucuronide	$C_{18}H_{22}N_2O_8$	393.13034	393.129	3.409	$[M-H]^-$	HMDB0060833
Aldosterone*	$C_{21}H_{28}O_5$	395.16308	395.164	2.328	$[M+Cl]^-$	HMDB0000037
5-(3',5'-Dihydroxyphenyl)-gamma-valerolactone-O-glucuronide-O-methyl	$C_{18}H_{22}O_{10}$	397.11402	397.114	0.050	$[M-H]^-$	HMDB0060030

5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-4'-O-methyl-3'-O-glucuronide						HMDB0059990
5-(3',4'-Dihydroxyphenyl)-gamma-valerolactone-3'-O-methyl-4'-O-glucuronide						HMDB0059988
5-(3',4'-dihydroxyphenyl)-gamma-valerolactone-3'-O-glucuronide						HMDB0029190
Melatonin glucuronide	$C_{19}H_{24}N_2O_8$	407.14599	407.146	0.025	$[M-H]^-$	HMDB0060830
Pyridinoline*	$C_{18}H_{28}N_4O_8$	409.17232	409.171	3.226	$[M-H_2O-H]^-$	HMDB0000851
Bicalutamide*‡	$C_{18}H_{14}F_4N_2O_4S$	411.04265	411.043	0.851	$[M-H_2O-H]^-$	HMDB0015260
5-(3',4',5'-trihydroxyphenyl)-gamma-valerolactone-O-methyl-5'-O-glucuronide†	$C_{18}H_{22}O_{11}$	413.10894	413.109	0.145	$[M-H]^-$	HMDB0060028
5-(3',4',5'-trihydroxyphenyl)-gamma-valerolactone-O-methyl-4'-O-glucuronide†	$C_{18}H_{22}O_{11}$	413.10894	413.109	0.145	$[M-H]^-$	HMDB0060027
beta-1,4-Mannosyl-N-acetylglucosamine†	$C_{14}H_{25}NO_{11}$	418.11216	418.110	5.166	$[M+Cl]^-$	HMDB0006535
Gemcitabine diphosphate	$C_9H_{13}F_2N_3O_{10}P_2$	421.99715	421.999	4.384	$[M-H]^-$	HMDB0060639
Alfuzosin†‡	$C_{19}H_{27}N_5O_4$	424.17571	424.177	3.041	$[M+Cl]^-$	HMDB0014490
Flavoxate*	$C_{24}H_{25}NO_4$	426.14776	426.150	5.256	$[M+Cl]^-$	HMDB0015279
Estrone glucuronide*	$C_{24}H_{30}O_8$	427.17568	427.178	5.431	$[M-H_2O-H]^-$	HMDB0004483
4-Hydroxy-5-(3',4'-dihydroxyphenyl)-valerate-O-methyl-O-glucuronide*†	$C_{18}H_{26}O_{12}$	433.13515	433.135	0.346	$[M-H]^-$	HMDB0059972
4-Hydroxy-5-(3',5'-dihydroxyphenyl)-valerate-O-methyl-O-glucuronide*†						HMDB0059974
Estramustine	$C_{23}H_{31}Cl_2NO_3$	438.16082	438.161	0.411	$[M-H]^-$	HMDB0015327
5-Methyltetrahydrofolic acid	$C_{20}H_{25}N_7O_6$	440.16824	440.166	5.089	$[M-H_2O-H]^-$	HMDB0001396

17-Hydroxypregnenolone sulfate	$C_{21}H_{32}O_6S$	447.16136	447.162	1.431	$[M+Cl]^-$	HMDB0000416
Glycoursodeoxycholic acid†‡	$C_{26}H_{43}NO_5$	448.30685	448.307	0.335	$[M-H]^-$	HMDB0000708
17-Hydroxyandrostane-3-glucuronide	$C_{25}H_{40}O_8$	449.25393	449.255	2.382	$[M-H_2O-H]^-$	HMDB0010359
Estriol-3-glucuronide	$C_{24}H_{32}O_9$	463.19736	463.193	9.413	$[M-H]^-$	HMDB0010335
Abiraterone sulfate*	$C_{24}H_{31}NO_4S$	464.16678	464.164	5.989	$[M+Cl]^-$	HMDB0060584
Androsterone glucuronide	$C_{25}H_{38}O_8$	465.24939	465.249	0.838	$[M-H]^-$	HMDB0002829
Enterolactone glucuronide	$C_{24}H_{26}O_{10}$	473.14532	473.146	1.437	$[M-H]^-$	HMDB0240377
Codeine-6-glucuronide	$C_{24}H_{29}NO_9$	474.17696	474.172	10.460	$[M-H]^-$	HMDB0060464
2-Methoxyestrone 3-glucuronide	$C_{25}H_{32}O_9$	475.19736	475.195	4.966	$[M-H]^-$	HMDB0004482
2-Methoxy-estradiol-17b 3-glucuronide	$C_{25}H_{34}O_9$	477.21301	477.209	8.403	$[M-H]^-$	HMDB0006765
DG(16:0/24:0/0:0)*	$C_{25}H_{36}O_9$	479.22866	479.229	0.709	$[M-H]^-$	HMDB0010338
Tetrahydrofolic acid‡	$C_{19}H_{23}N_7O_6$	480.14038	480.139	2.874	$[M+Cl]^-$	HMDB0001846
6-Hydroxymelatonin†	$C_{25}H_{38}O_9$	481.24431	481.245	1.434	$[M-H]^-$	HMDB0010351
Aldosterone 18-glucuronide*	$C_{27}H_{36}O_{11}$	535.21849	535.218	0.916	$[M-H]^-$	HMDB0010345
Cortolone-3-glucuronide*†	$C_{27}H_{42}O_{11}$	541.26544 577.24211	541.266 577.240	1.035 3.655	$[M-H]^-$ $[M+Cl]^-$	HMDB0010320
Tetrahydroaldosterone-3-glucuronide	$C_{27}H_{40}O_{11}$	575.22646	575.225	2.538	$[M+Cl]^-$	HMDB0010357
3'-Sialyllactose*†‡	$C_{23}H_{39}NO_{19}$	632.20435	632.205	1.028	$[M-H]^-$	HMDB0000825
3-Sialyl-N-acetyllactosamine†	$C_{25}H_{42}N_2O_{19}$	673.23090	673.232	1.634	$[M-H]^-$	HMDB0006581

* Differential metabolites between GU and HC, † Differential metabolites between BC and PC, ‡ Metabolites validated by LC-MS/MS.

Table S2. Some of the metabolites in positive ion mode and negative ion mode using LC-MS/MS.

Name	Formula	Theoretical *m/z*	Experimental *m/z*	Delta (ppm)	Adduct
Tryptophol	$C_{10}H_{11}NO$	160.076	160.076	0.000	$[M-H]^-$
N-Benzyloxycarbonylglycine	$C_{10}H_{11}NO_4$	210.07600	210.080	19.041	$[M+H]^+$
		208.06911	208.062	34.171	$[M-H]^-$
Porphobilinogen	$C_{10}H_{14}N_2O_4$	227.10260	227.101	7.045	$[M+H]^+$
		227.10260	227.102	2.642	$[M+Na]^+$
Ribothymidine	$C_{10}H_{14}N_2O_6$	259.09180	259.090	6.947	$[M+H]^+$
		257.07819	257.078	0.739	$[M-H]^-$
3-Methyluridine		259.08289	259.080	11.155	$[M+H]^+$
		257.07541	257.076	2.295	$[M-H]^-$
Prolylhydroxyproline	$C_{10}H_{16}N_2O_4$	229.11830	229.117	5.674	$[M+H]^+$
		227.10373	227.104	1.189	$[M-H]^-$
L-Tryptophan	$C_{11}H_{12}N_2O_2$	407.17001	407.169	2.481	$[2M-H]^-$
		188.07001	188.070	0.053	$[M+H]^+$
N-Acetylcytidine	$C_{11}H_{15}N_3O_6$	203.08260	203.083	1.970	$[M-H]^-$
		286.10309	286.105	6.676	$[M+H]^+$
		284.08890	284.087	6.688	$[M-H]^-$
5'-Methylthioadenosine	$C_{11}H_{15}N_5O_3S$	298.09683	298.095	6.139	$[M+H]^+$
		296.08228	296.085	9.187	$[M-H]^-$
N-Acetylaspartylglutamic acid	$C_{11}H_{16}N_2O_8$	303.08368	303.089	17.553	$[M-H]^-$

N-Acetylneuraminic acid	$C_{11}H_{19}NO_9$	308.09818	308.094	13.567	$[M-H]^-$
Nilutamide	$C_{12}H_{10}F_3N_3O_4$	316.05511	316.051	13.004	$[M-H]^-$
methyl 3-(1H-indol-3-yl)propanoate	$C_{12}H_{13}NO_2$	204.10190	204.101	4.410	$[M+H]^+$
N2,N2-Dimethylguanosine	$C_{12}H_{17}N_5O_5$	312.12970	312.131	4.165	$[M+H]^+$
		310.11591	310.115	2.934	$[M-H]^-$
Trehalose	$C_{12}H_{22}O_{11}$	365.10541	365.110	12.572	$[M+Na]^+$
		377.13000	377.129	2.652	$[M+Cl]^-$
N-Acetyltryptophan	$C_{13}H_{14}N_2O_3$	247.10809	247.108	0.364	$[M+H]^+$
		245.09309	245.093	0.367	$[M-H]^-$
Melatonin	$C_{13}H_{16}N_2O_2$	233.00000	232.999	4.292	$[M+H]^+$
		231.11430	231.123	37.644	$[M-H]^-$
Phenylacetylglutamine	$C_{13}H_{16}N_2O_4$	265.11951	265.118	5.696	$[M+H]^+$
		263.10251	263.104	5.663	$[M-H]^-$
Aspartylphenylalanine	$C_{13}H_{16}N_2O_5$	281.10001	281.100	0.036	$[M+H]^+$
Ibuprofen	$C_{13}H_{18}O_2$	207.13795	207.137	4.586	$[M+H]^+$
Gliclazide	$C_{15}H_{21}N_3O_3S$	324.13760	324.142	13.574	$[M+H]^+$
		322.12311	322.129	18.285	$[M-H]^-$
Riboflavin	$C_{17}H_{20}N_4O_6$	377.14557	377.146	1.140	$[M+H]^+$
		375.13049	375.128	6.638	$[M-H]^-$
Bicalutamide	$C_{18}H_{14}F_4N_2O_4S$	431.06830	431.067	3.016	$[M+H]^+$
		429.05380	429.051	6.526	$[M-H]^-$
Enterodiol	$C_{18}H_{22}O_4$	285.14871	285.147	5.997	$[M+H]^+$

Casticin	$C_{19}H_{18}O_8$	375.10718	375.109	4.852	$[M+H]^+$
Tetrahydrofolic acid	$C_{19}H_{23}N_7O_6$	444.16370	444.163	1.576	$[M-H]^-$
Alfuzosin	$C_{19}H_{27}N_5O_4$	388.19904	388.209	25.657	$[M-H]^-$
Androsterone sulfate	$C_{19}H_{30}O_5S$	369.17410	369.173	2.980	$[M-H]^-$
3'-Sialyllactose	$C_{23}H_{39}NO_{19}$	632.20441	632.206	2.515	$[M-H]^-$
Glycodeoxycholic acid	$C_{26}H_{43}NO_5$	450.10001	450.104	8.865	$[M+H]^+$
Uric acid	$C_5H_4N_4O_3$	167.02100	167.021	0.000	$[M-H]^-$
		169.03561	169.034	9.525	$[M+H]^+$
Glucuronic acid	$C_6H_{10}O_7$	193.03537	193.036	3.264	$[M-H]^-$
Glucose	$C_6H_{12}O_6$	181.07066	181.070	3.645	$[M+Na]^+$
Inositol		179.05540	179.056	3.351	$[M-H]^-$
		203.05260	203.049	17.729	$[M+Na]^+$
Gluconic acid	$C_6H_{12}O_7$	195.05099	195.051	0.051	$[M-H]^-$
1-Methyluric acid	$C_6H_6N_4O_3$	183.05127	183.049	12.401	$[M+H]^+$
		181.03670	181.036	3.867	$[M-H]^-$
Ascorbic acid	$C_6H_8O_6$	175.02480	175.023	10.284	$[M-H]^-$
Citric acid	$C_6H_8O_7$	191.01900	191.019	0.000	$[M-H]^-$
		193.03429	193.034	1.502	$[M+H]^+$
Quinolinic acid	$C_7H_5NO_4$	166.01486	166.016	6.867	$[M-H]^-$
3-Hydroxyanthranilic acid	$C_7H_7NO_3$	154.04980	154.049	5.193	$[M+H]^+$
p-Tolyl Sulfate	$C_7H_8O_4S$	187.00710	187.007	0.535	$[M-H]^-$
3-Indoxyl sulfate	$C_8H_7NO_4S$	212.00230	212.003	3.302	$[M-H]^-$

4-Hydroxyphenyllactic acid	$C_9H_{10}O_4$	181.04980	181.050	1.105	$[M-H]^-$
DL-p-Hydroxyphenyllactic acid		181.05011	181.049	6.131	$[M-H]^-$
Pseudouridine	$C_9H_{12}N_2O_6$	243.06210	243.062	0.411	$[M-H]^-$
Uridine		245.07680	245.076	3.264	$[M+H]^+$
Hippuric acid	$C_9H_9NO_3$	180.06551	180.065	2.832	$[M+H]^+$
		178.05099	178.051	0.056	$[M-H]^-$
4-Hydroxyhippuric acid	$C_9H_9NO_4$	196.05920	196.060	4.080	$[M+H]^+$
		194.04520	194.046	4.123	$[M-H]^-$
Salicyluric acid		194.04590	194.046	0.515	$[M-H]^-$

Table S3. Age and gender distribution.

	BC (n = 38)	PC (n = 39)	HC (n = 40)	p value
	GU (n = 77)			
Cohort 1				
Number of individuals	56		33	
Age, mean ± SE	66.1 ± 0.9		62.9 ± 1.5	0.054
Age range	55-77		53-86	-
Gender (f/m)	3/53		5/28	0.239
Cohort 2				
Number of individuals	38	39	-	
Age, mean ± SE	65.5 ± 1.9	70.1 ± 1.4	-	0.052
Age range	41-81	56-83	-	-
Gender (f/m)	4/34	0/39	-	0.117
Cohort 3				
Number of individuals	38	-	40	
Age, mean ± SE	65.5 ± 1.9	-	61.0 ± 1.4	0.057
Age range	41-81	-	50-86	-
Gender (f/m)	4/34	-	5/35	1.000
Cohort 4				
Number of individuals	-	32	31	
Age, mean ± SE	-	67.3 ± 1.2	63.5 ± 1.5	0.054
Age range	-	56-82	54-86	-
Gender (f/m)	-	0/32	0/31	1.000
Cohort 5				
Number of individuals	27	-	40	
Age, mean ± SE	63.9 ± 2.2	-	61.0 ± 1.4	0.057
Age range	41-81	-	50-86	-
Gender (f/m)	3/24	-	5/35	1.000

The two-tailed test was used to calculate the statistical significance in age distribution between the groups in each cohort. A chi-square (χ^2) test was applied to calculate the statistical significance in gender composition. BC: bladder cancer group, PC: prostate cancer group, HC: healthy control group. Cohort 1: GU (BC+PC) vs HC, Cohort 2: BC vs PC, Cohort 3: BC vs HC, Cohort 4: PC vs HC and Cohort 5: initial-diagnosed BC vs HC.

Table S4. Detailed information of the patients.

No.	Gender	Age	Group	Therapy	Cohort			
					1	2	3	4
1	male	50	HC				√	
2	male	51	HC				√	
3	male	51	HC				√	
4	male	52	HC				√	
5	male	53	HC				√	
6	male	53	HC				√	
7	male	53	HC				√	
8	male	53	HC		√		√	
9	male	54	HC		√		√	
10	male	54	HC		√		√	√
11	male	55	HC		√		√	√
12	male	55	HC		√		√	√
13	male	56	HC		√		√	√
14	male	56	HC		√		√	√
15	male	56	HC		√		√	√
16	male	56	HC		√		√	√
17	male	57	HC		√		√	√
18	male	57	HC		√		√	√
19	female	58	HC		√		√	√
20	male	58	HC		√		√	√
21	male	58	HC		√		√	√
22	male	58	HC		√		√	√
23	female	59	HC		√		√	√
24	male	60	HC		√		√	√
25	male	60	HC		√		√	√
26	male	61	HC		√		√	√
27	female	63	HC		√		√	√
28	male	63	HC		√		√	√
29	male	63	HC		√		√	√
30	male	65	HC		√		√	√
31	male	66	HC		√		√	√
32	female	71	HC		√		√	√
33	female	71	HC		√		√	√
34	male	72	HC		√		√	√
35	male	73	HC		√		√	√
36	male	73	HC		√		√	√
37	male	75	HC		√		√	√
38	male	76	HC		√		√	√
39	male	79	HC		√		√	√
40	male	86	HC		√		√	√

41	male	41	BC			√	√
42	male	43	BC			√	√
43	male	45	BC			√	√
44	male	47	BC	postoperative, chemotherapy		√	√
45	male	47	BC			√	√
46	male	55	BC		√	√	√
47	male	55	BC		√	√	√
48	male	56	BC	postoperative	√	√	√
49	male	57	BC		√	√	√
50	male	57	BC		√	√	√
51	male	57	BC		√	√	√
52	male	57	BC		√	√	√
53	male	61	BC		√	√	√
54	male	62	BC	postoperative	√	√	√
55	male	63	BC	postoperative, chemotherapy	√	√	√
56	female	64	BC		√	√	√
57	female	65	BC		√	√	√
58	male	66	BC		√	√	√
59	male	67	BC		√	√	√
60	male	69	BC		√	√	√
61	male	69	BC		√	√	√
62	male	70	BC		√	√	√
63	male	70	BC	chemotherapy	√	√	√
64	male	70	BC	postoperative	√	√	√
65	male	71	BC		√	√	√
66	male	71	BC		√	√	√
67	male	74	BC		√	√	√
68	male	75	BC		√	√	√
69	male	75	BC	chemotherapy	√	√	√
70	female	76	BC		√	√	√
71	male	76	BC		√	√	√
72	male	76	BC		√	√	√
73	male	80	BC	postoperative, chemotherapy		√	√
74	male	80	BC	postoperative		√	√
75	male	80	BC			√	√
76	female	80	BC	postoperative, chemotherapy		√	√
77	male	80	BC	postoperative, chemotherapy		√	√
78	male	81	BC			√	√

79	male	56	PC		√	√	√
80	male	59	PC		√	√	√
81	male	59	PC		√	√	√
82	male	60	PC		√	√	√
83	male	60	PC	radiotherapy	√	√	√
84	male	60	PC	radiotherapy	√	√	√
85	male	60	PC		√	√	√
86	male	60	PC		√	√	√
87	male	60	PC		√	√	√
88	male	65	PC		√	√	√
89	male	65	PC		√	√	√
90	male	65	PC		√	√	√
91	male	65	PC		√	√	√
92	male	65	PC	radiotherapy	√	√	√
93	male	65	PC		√	√	√
94	male	66	PC		√	√	√
95	male	66	PC		√	√	√
96	male	66	PC		√	√	√
97	male	69	PC		√	√	√
98	male	69	PC		√	√	√
99	male	69	PC		√	√	√
100	male	69	PC		√	√	√
101	male	72	PC		√	√	√
102	male	72	PC		√	√	√
103	male	74	PC		√	√	√
104	male	74	PC		√	√	√
105	male	74	PC		√	√	√
106	male	74	PC		√	√	√
107	male	77	PC		√	√	√
108	male	77	PC			√	√
109	male	81	PC			√	√
110	male	82	PC			√	√
111	male	82	PC			√	
112	male	82	PC			√	
113	male	83	PC			√	
114	male	83	PC			√	
115	male	83	PC			√	
116	male	83	PC			√	
117	male	83	PC			√	

BC: bladder cancer patients, PC: prostate cancer patients, HC: healthy control volunteers, √: used in the cohort.

Table S5. Human biospecimens characteristics.

characteristics	
Biospecimen type	Urine
Clinical diagnosis of patients	Bladder cancer, Prostate cancer, Healthy control
Collection mechanism	Fasting morning urine
Constitution of preservative	None
Type of long-term preservation	Freezing
Storage temperature	-80℃
Storage duration	0 to 7 months

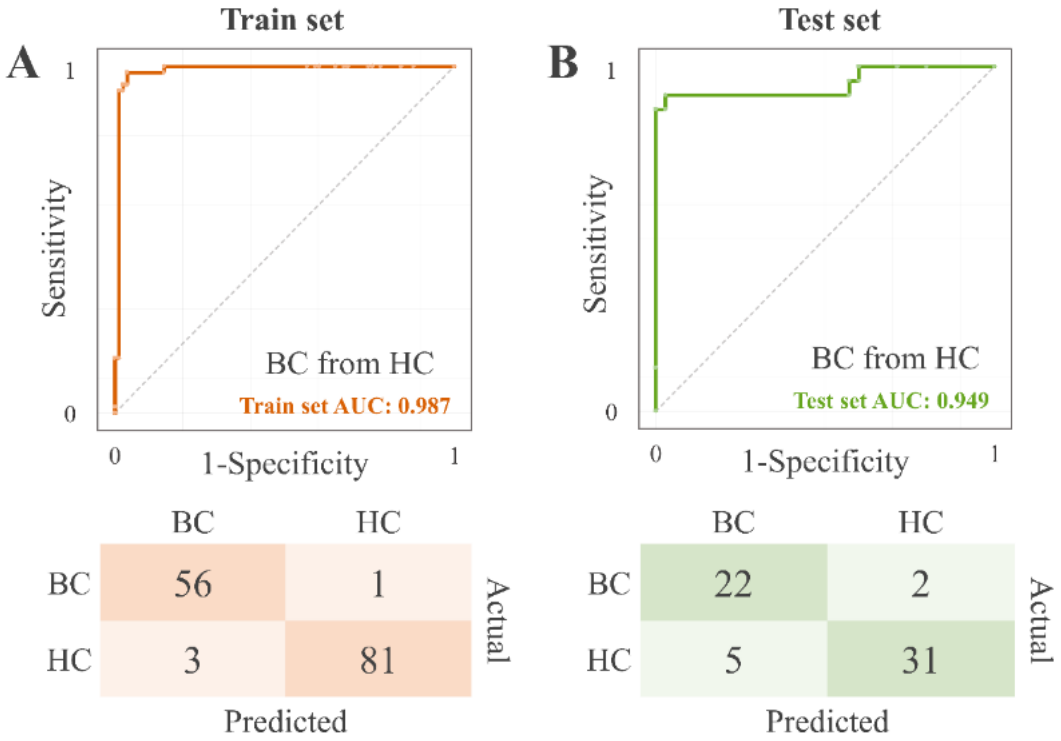

Fig. S8 ROC curves of diagnostic performances and confusion matrices based on (A) train set and (B) test set of cohort 5 using Neural Network model.

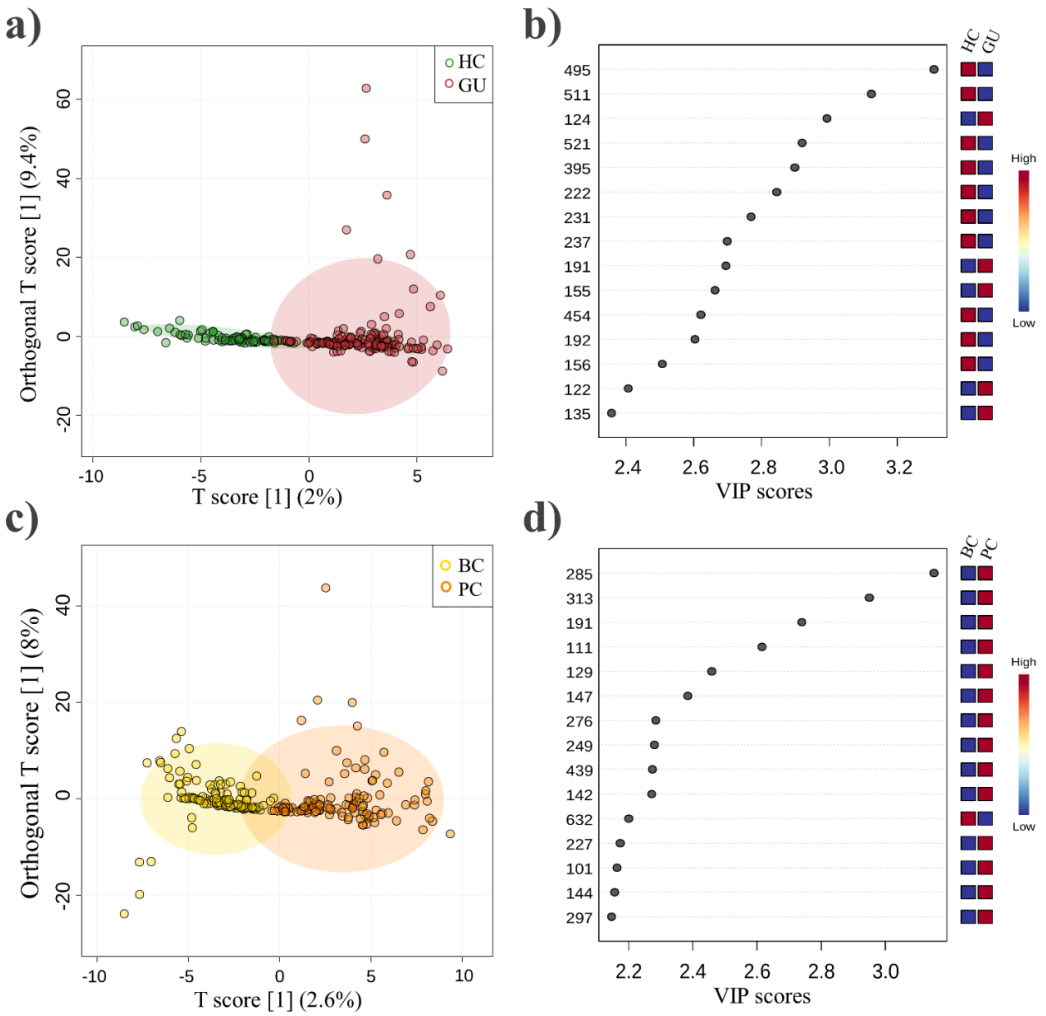

Fig. S9 Orthogonal partial least squares discrimination analysis (OPLS-DA). OPLS-DA score plots of **a)** cohort 1 and **c)** cohort 2. Variable importance in the projection (VIP) score plot of **b)** cohort 1 and **d)** cohort 2.

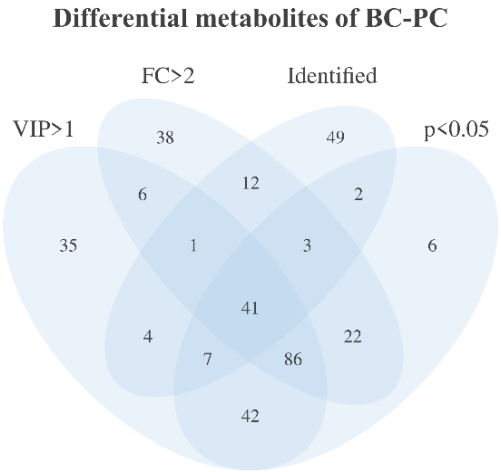

Fig. S10 Venn diagram of differential metabolites between BC and PC. VIP: variable importance of projection values. FC: fold change values (the metabolite content ratios of BC to PC in cohort 2). Identified: metabolites with annotation. p: p values with FDR-corrected using the two-tailed test.

Fig. S11 The Pearson correlation heatmap of 35 differential metabolites between GU and HC in cohort 1. Pink indicates a positive correlation, while blue indicates a negative correlation. *: <0.05, **: <0.01, ***: <0.001. GU: genitourinary cancers group including bladder cancer (BC) and prostate cancer (PC). HC: healthy control group.

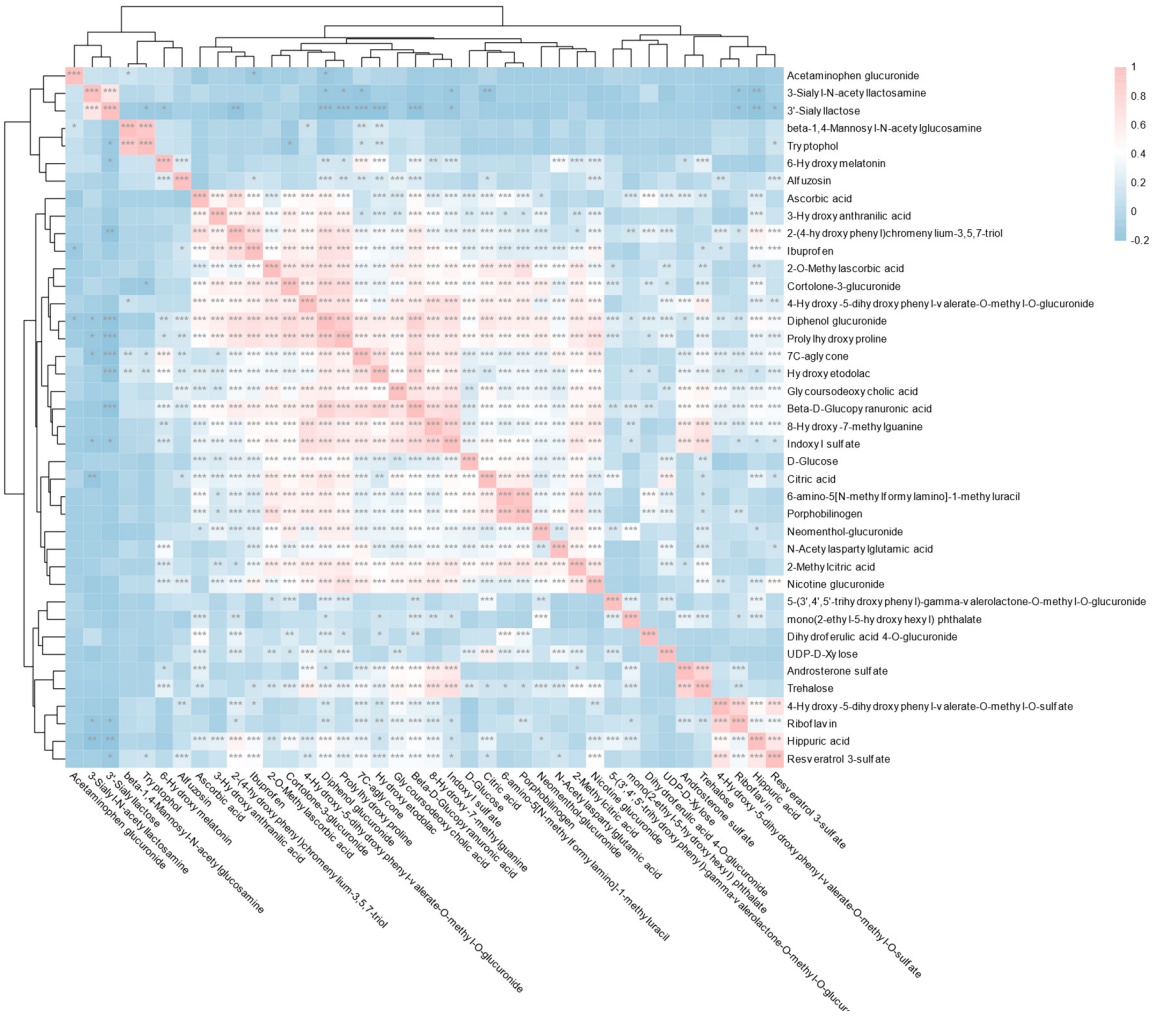

Fig. S12 The Pearson correlation heatmap of 41 differential metabolites between BC and PC in cohort 2. Pink indicates a positive correlation, while blue indicates a negative correlation. *: <0.05, **: <0.01, ***: <0.001. BC: bladder cancer group, PC: prostate cancer group.

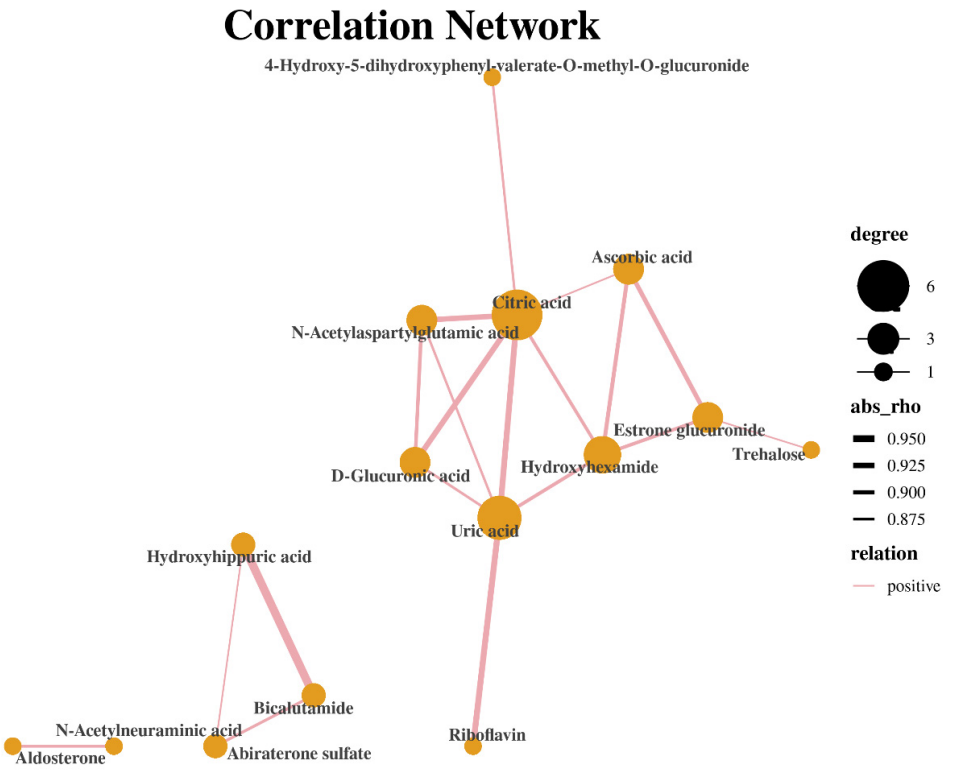

Fig. S13 Correlation network (Pearson correlation > 0.85) of differential metabolites between GU and HC.

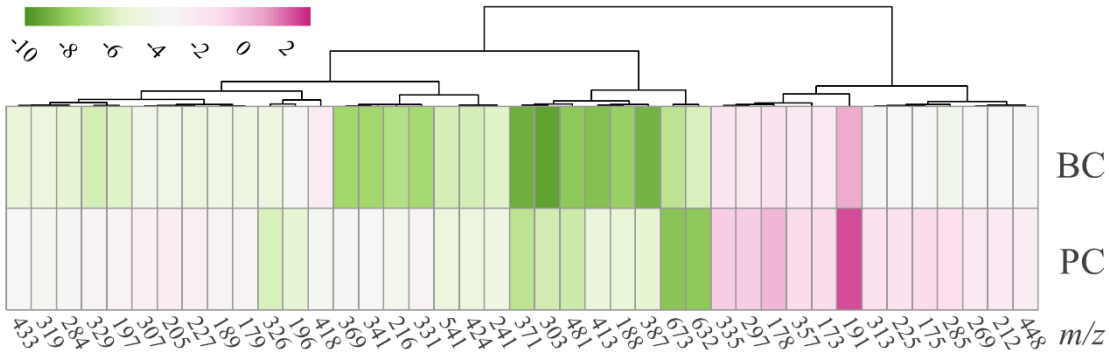

Fig. S14 The heatmap constructed from the average normalized intensity values of 41 differential metabolites between BC and PC.

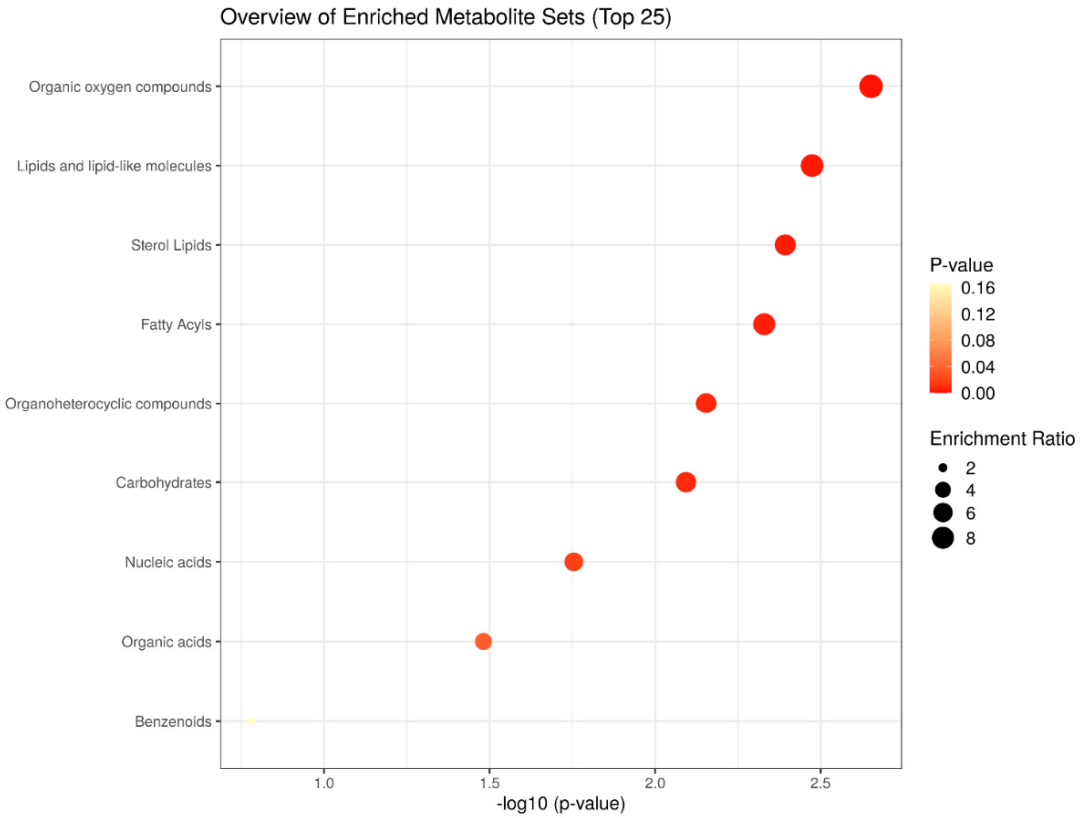

Fig. S15 Quantitative metabolite sets enrichment analysis (qMSEA) based on chemical structures of differential metabolites between GU and HC.

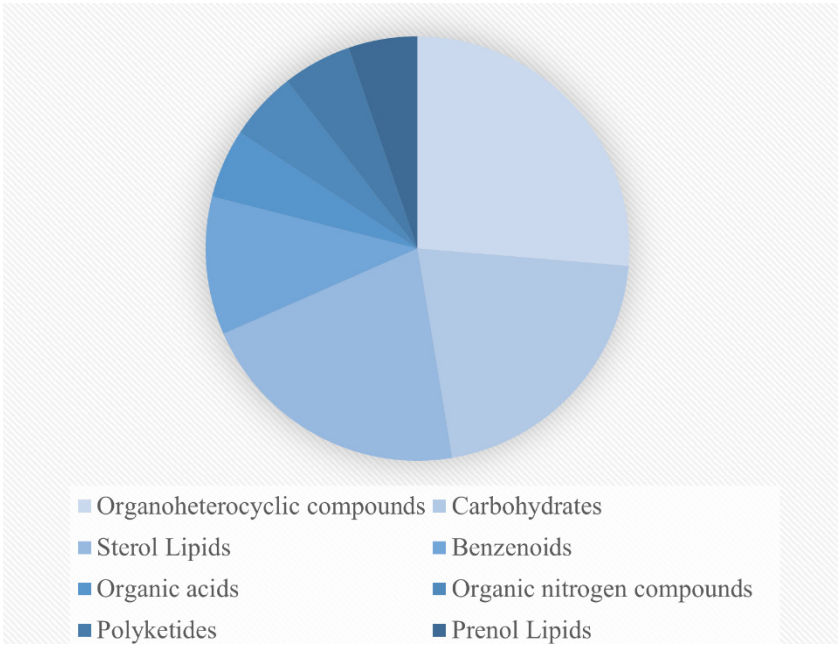

Fig. S16 Distribution pie chart of major metabolite sets of differential metabolites between BC and PC.

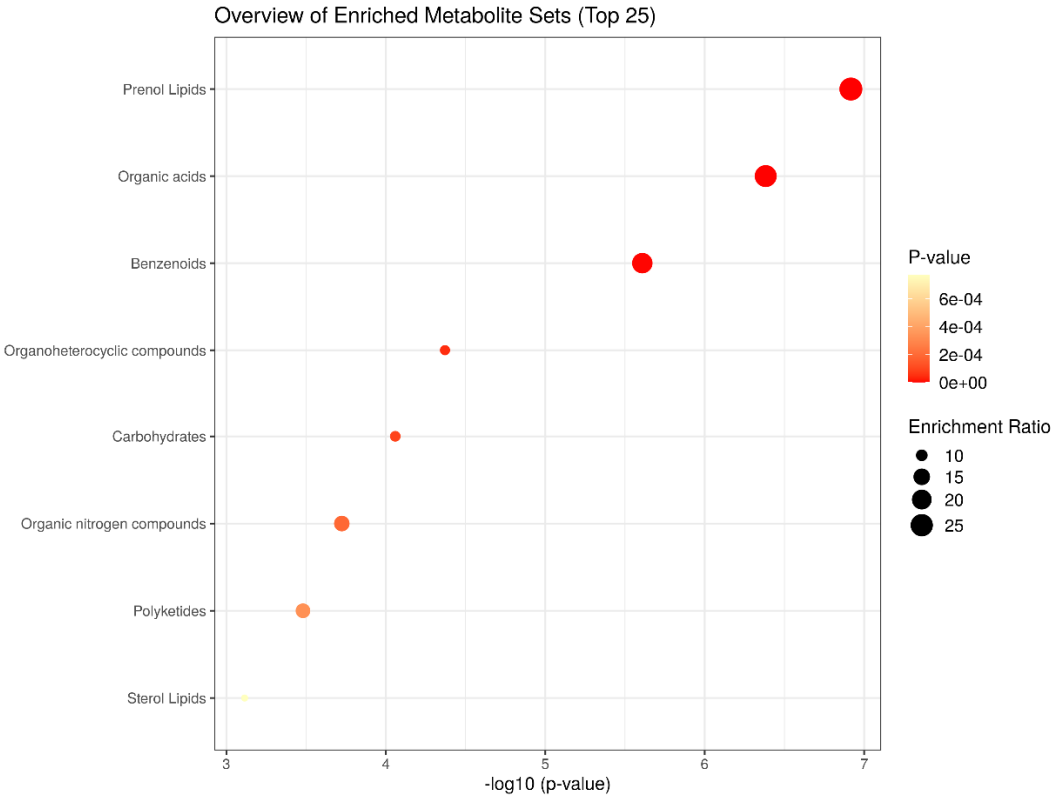

Fig. S17 Quantitative metabolite sets enrichment analysis (qMSEA) based on chemical structures of differential metabolites between BC and PC.

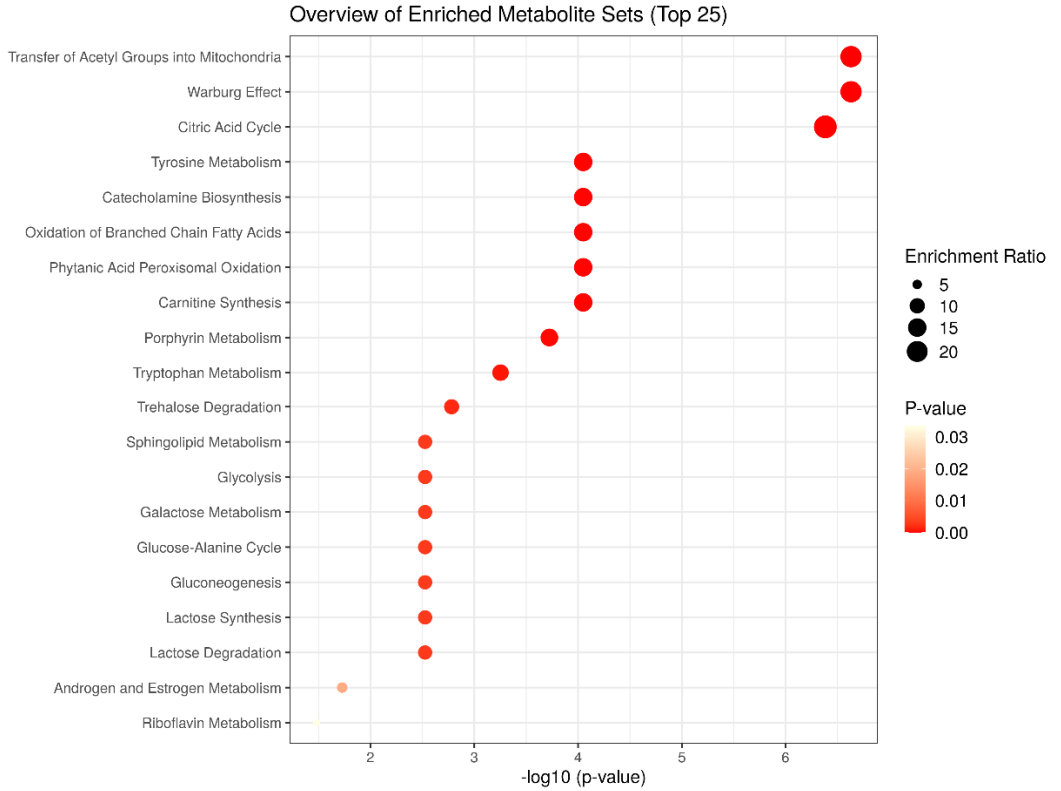

Fig. S18 Dysregulated metabolic pathways between BC and PC based on quantitative metabolite set enrichment analysis (qMSEA) of differential metabolites..

Table S6. Parameters of seven different machine learning models using all *m/z* features of test sets.

	Model	AUC	Accuracy	F1	Precision	Recall
HC vs GU	NB	0.933	0.827	0.818	0.839	0.827
	RF	0.929	0.864	0.860	0.870	0.864
	SVM	0.903	0.765	0.729	0.829	0.765
	LR	0.853	0.790	0.788	0.787	0.790
	kNN	0.843	0.765	0.768	0.777	0.765
	NN	0.814	0.765	0.769	0.784	0.765
	AdaBoost	0.703	0.704	0.708	0.720	0.704
BC vs PC	SVM	0.949	0.826	0.825	0.830	0.826
	LR	0.936	0.841	0.841	0.841	0.841
	NN	0.928	0.884	0.884	0.884	0.884
	RF	0.921	0.884	0.883	0.895	0.884
	NB	0.919	0.841	0.839	0.847	0.841
	kNN	0.909	0.855	0.855	0.861	0.855
	AdaBoost	0.823	0.826	0.825	0.830	0.826
BC vs HC	SVM	0.845	0.768	0.765	0.795	0.768
	NN	0.804	0.826	0.824	0.835	0.826
	RF	0.803	0.696	0.692	0.713	0.696
	NB	0.741	0.522	0.399	0.761	0.522
	AdaBoost	0.663	0.667	0.664	0.668	0.667
	LR	0.651	0.696	0.694	0.696	0.696
	kNN	0.604	0.565	0.565	0.567	0.565
PC vs HC	SVM	0.943	0.807	0.805	0.829	0.807
	LR	0.928	0.719	0.710	0.771	0.719
	NB	0.916	0.842	0.842	0.843	0.842
	RF	0.909	0.807	0.804	0.843	0.807
	NN	0.893	0.807	0.802	0.863	0.807
	kNN	0.806	0.684	0.677	0.716	0.684
	AdaBoost	0.759	0.754	0.753	0.769	0.754

RF (Random Forest), SVM (Support Vector Machines), NN (Neural Network), kNN (k-Nearest Neighbor), NB (Naive Bayes), LR (Logistic Regression). The equations of the relevant parameters are given in ESI† (Equations).

Table S7. Parameters of seven different machine learning models using differential metabolites features of test sets.

	Model	AUC	Accuracy	F1	Precision	Recall
HC vs GU	NN	0.940	0.901	0.903	0.905	0.901
	SVM	0.925	0.864	0.865	0.867	0.864
	NB	0.892	0.802	0.805	0.809	0.802
	LR	0.879	0.815	0.813	0.812	0.815
	RF	0.726	0.691	0.694	0.696	0.691
	AdaBoost	0.725	0.753	0.762	0.779	0.753
	kNN	0.666	0.593	0.618	0.692	0.593
BC vs PC	NN	0.971	0.899	0.898	0.901	0.899
	LR	0.969	0.812	0.811	0.853	0.812
	RF	0.936	0.826	0.824	0.828	0.826
	kNN	0.934	0.768	0.763	0.849	0.768
	SVM	0.913	0.812	0.812	0.822	0.812
	NB	0.897	0.797	0.798	0.804	0.797
	AdaBoost	0.76	0.768	0.767	0.767	0.768

RF (Random Forest), SVM (Support Vector Machines), NN (Neural Network), kNN (k-Nearest Neighbor), NB (Naive Bayes), LR (Logistic Regression). The equations of the relevant parameters are given in ESI† (Equations).

References

1. Q. He, S. Chen, J. Wang, J. Hou, J. Wang, S. Xiong and Z. Nie, *Clin Chim Acta*, 2013, **420**, 94-98.

2. C. Burton and Y. Ma, *Current Medicinal Chemistry*, 2019, **26**, 5-28; Q. Wen, P. Boshier, A. Myridakis, I. Belluomo and G. B. Hanna, *Metabolites*, 2021, **11**.