Electronic supplementary information (ESI) for

Hierarchical superstructure aerogels for *in situ* biofluid metabolomics

Siyu Chen [1,†], Keyuan Pu [1,†], Yue Wang [1], Yang Su [1], Jiamin Qiu [2], Xin Wang [3], Kunbin Guo [3], Jun Hu [1], Huiwen Wei [1], Hongbiao Wang [3], Xiaolong Wei [3], Yuping Chen [3], Wen Lin [3]*, Wenxiu Ni [4,7], Yan Lin[5], Jiayang Chen [6], Samuel Kin-Man Lai [8] and Kwan-Ming Ng [1,7,8]*

[1] Department of Chemistry and Key Laboratory for Preparation and Application of Ordered Structural Materials of Guangdong Province, Shantou University, Guangdong, 515063, China

[2] Department of Biology, Shantou University, Shantou, Guangdong, 515063, China

[3] The Cancer Hospital of Shantou University Medical College, Guangdong, 515031, China

[4] Department of Medicinal Chemistry, Shantou University Medical College, Guangdong, 515041, China

[5] The Second Affiliated Hospital of Shantou University Medical College, Guangdong, 515041, China

[6] Instrumental Analysis & Testing Centre, Shantou University, Guangdong, 515063, China

[7] Chemistry and Chemical Engineering Guangdong Laboratory, Guangdong, 515063, China

[8] Laboratory for Synthetic Chemistry and Chemical Biology Limited, Units 1503-1511, 15/F., Building 17W, Hong Kong Science Park, New Territories, Hong Kong, China.

[†] **S.C., and K.P. contributed equally to this work.**

* Corresponding authors:
Kwan-Ming Ng (ngkm@hksccb.hk);
Wen Lin (1263811129@qq.com);

This part includes:

I. Supporting Methods

1. Chemicals and Materials

Preparation of gold-on-silica double-layered aerogel thin film: Sodium citrate (Macklin, 99%), hydrogen tetrachloroaurate(III) tetrahydrate (HAuCl$_4$·4H$_2$O, Macklin, 48~50% Au basis), sodium borohydride (NaBH$_4$, Xilong Scientific, AR), ammonium fluoride (NH$_4$F, Macklin, 98%), (3-glycidyloxypropyl)trimethoxysilane (GOPTS, Sigma Aldrich, 98.0%), sodium silicate (Macklin, powder, ≥18% Na (as Na$_2$O) basis, ≥60% Si(as SiO$_2$) basis) were used as received.

Standards utilized for determination of metabolite coverage and detection sensitivity: L-Histidine (L-His, Aladdin, 99%), verapamil hydrochloride (Macklin, 99%), angiotensin II human (Sigma-Aldrich, 93%), melatonin (Aladdin, 98%), α-cyclodextrin (Aladdin, 99%), L(+)-arginine (L-Arg, Aladdin, 99%), acetylcholine bromide (Macklin, 98%), Arg-Gly-Asp-Ser (RGDS, Xian xirui Bio-Technology, 98%), adenine (Shanghai yuanye Bio-Technology, 98%), berberine (Energy, 98%), sildenafil citrate (Shanghai TCI chemicals, 98%), leucine-enkephalin (Leu-enk, Macklin, BR), poly(ethylene glycol) (PEG 1000, Aladdin, 99%) were used as received.

2. Material preparation

GOPTS derivatization of glass coverslip: To perform the GOPTS derivatization, the coverslip (Bellco Glass Inc., 75mm x 25mm, 0.13 – 0.16mm thick) was firstly cleaned with piranha solution (H$_2$SO$_4$ (Xilong Scientific, AR) : H$_2$O$_2$ (Xilong Scientific, AR) = 3 : 1) followed by washing with D.I. water thoroughly. The cleaned coverslip was then derivatized with 2 % (*v/v*) GOPTS in anhydrous toluene for 19 hours, followed by drying with argon gas.

Preparation of silica aerogel thin film: Sodium silicate (Na$_2$SiO$_3$) solution was used as the silicon source for the preparation of silica aerogel by a sol-gel method [1]. 1.0 M NH$_4$F solution (0.32 mL) was added in the 0.4226 M sodium silicate solution PH = 11 - 12 (3.0 ml) and stirred (4 hours) for hydrolysis. 3.0 M Sulfuric acid (0.69 mL) was then added into the hydrolyzed solution to trigger the condensation reaction. The molar ratio of Na$_2$SiO$_3$: NH$_4$F : H$_2$SO$_4$ in the solution was kept constant at 1: 0.25:1.63. After stirring for 2 min, the gel solution (3 ml) was spin-coated (at the acceleration of 100 rpm/s to reach the final speed at 1800 rpm) by Spin-coater machine (LEBO Science, EZ4-S, China) on a GOPTS derivatized glass coverslip [2, 3]. The spin coating was repeated for three times. The spin-coated silica gel was immersed in DI water for aging (3 hours), and then dried under ambient conditions (heated at 160°C for 3 min).

Au nanoparticle synthesis: The aqueous solutions of sodium citrate (400 mM, 25 µl) and HAuCl$_4$·4H$_2$O (32.5 mM, 30.8 µl) were added into 4.93 ml water with stirring. After maintaining the stirring for 15 min, the freshly prepared NaBH$_4$ aqueous solution (200 mM, 20 µl) was added in the aqueous solution in a dropwise manner, and then stirring for 60 min. The relative amount of the gold salt, citrate ligand and the NaBH$_4$ was in the molar ratio of 1:10:4. The Au nanoparticle solution was then aged for one day before being utilized to synthesize the Au aerogel.

Preparation of gold-on-silica double-layered aerogel thin film: To make the thin layer of Au aerogel firmly attached on the surface of silica aerogel layer, GOPTS derivatization was

performed on the silica aerogel surface where the epoxy group can covalently bond to the Au surface (**Figure S6**). The silica aerogel was derivatized with 2 % (*v/v*) GOPTS in anhydrous toluene for 19 hours. To prepare a thin layer of Au aerogel on the top of silica aerogel, AuNPs solution (3-6 nm AuNP particle size, 6.20249E+19 NP/cm^3 particle density, 171 ml) were destabilized by the addition NH_4F (1.0 M, 19 ml to make up the total solution level of 3 cm, or specified elsewhere) to trigger the fusion of the AuNPs to form interconnected 3D networked sol-gel [**4**] which were sedimented on the surface of the silica aerogel coated coverslip. The solution was well mixed and then ground for 8 hours. After washing with DI water, the sol-gel was dried under ambient conditions to form the double-layered aerogel. In the sample analysis, the coverslip was fixed on the MALDI plate, and then rostered with the laser marking to generate the sample wells for biofluid application (**Figure S2**).

Functionalization of gold-on-silica double-layered aerogel thin film: To change the surface polarity of the Au aerogel, a variety of thiol ligands, including But, 3-Met-1-but, BM, 2-Eth-hex, 4-MPM, 2-Nap, Dec, Dodec, Octadec, ACT, 4-ATP, 3-MP, citrate, L-Cys, mPEG550-HS, mPEG750-HS, mPEG2000-HS, mPEG5000-HS, mPEG10000-HS were applied to perform the ligand exchange reaction on the Au aerogel surface. The ligands were dissolved by toluene or H_2O to react with Au aerogel for 8 hours. The details for the ligand exchange are summarized in **Table S2**.

3. Characterization of aerogel thin film

The aerogels were characterized with SEM (GeminiSEM - ZEISS Field Emission SEM, Germany), TEM (HR-TEM, JEM-F200, JEOL, Japan), XRD (Rigaku MiniFlex 600, Japan), UV-vis spectroscopy (Macy UV-1800PC, Shanghai, China), thickness measurement (Bruker Dektak XT, Germany), nitrogen adsorption (Micromeritics ASAP 2020, America), and contact angle measurement (JCY-1, Fangrui, Shanghai, China). The STEM tomography measurements were performed with TalosF200C (Thermo Scientific, USA).The screening parameters were: screen current 0.055 nA * 2.6 S, pixel size 0.276 nm, collecting angle 3-20 mrad. The 3D-STEM image was reconstructed by 3dmod (University of Colorado, version 4.11, USA). The reconstructed .vtk format file was converted into .stl file by Paraview (Kitware, version 5.11, USA). To simulate the E-field distribution on the photo-excited Au aerogel, finite-difference time-domain (FDTD) method (Ansys Lumerical, USA) was adopted. A model of the Au aerogel in the size of 276.0 * 207.0 * 96.6 nm was created from the reconstructed 3D-STEM image. The photo wavelength at 355 nm was incident along the Z-axis and the grids is 0.4 nm. The refractive index of the gold material was from Johnson and Christy [**5**].

4. Theoretical simulation

The Molecular dynamics (MD) simulation was performed by the large-scale atomic/molecular massively parallel simulator (LAMMPS) [6]. The computational domain has a length of 50 nm in both x and y directions, while the z direction has a size of 75 nm. The prepared computational domain has been separated into three regions, (1) the eluent region (z = 0 nm to 25 nm), (2) the Au aerogel region (z = 25 nm to 50 nm) and (3) the water region (z = 50 nm to 75 nm). The water region has been filled with a water slab of the same size (~ rigid 3000 SPC/E water molecules) while the eluent region has been kept empty. The central part of the 3D-STEM image with a size of 50 nm (l) x 50 nm (w) x 25 nm (h) has been extracted and converted to Tripos molecule structure format (MOL) using VEGA ZZ (Version 3.2.1.33) [7]. This Au aerogel model has then been fitted in the Au aerogel region of the simulation domain. Simulations were carried out with periodic boundary conditions applied in lateral directions while a fixed and reflective wall has been set for the z direction. Second-nearest-neighbor modified embedded-atom potentials (MEAM-2nn) [8] were employed for accounting for the interaction between Au atoms, while the interactions between water molecules or the interactions between water molecules and Au atoms have been computed using the standard 12/6 Lennard-Jones potential with a cutoff of 12 Å. The system was first pre-equilibrated with 50,000 steps until its maximum force was below 1,000.0 kJ/mol/nm, and then equilibrated using NVT at 298 K by coupling all atoms to a Berendsen-thermostat with a time constant of 0.1 ps and a simulation trajectory of 1 ns. The system was being kept at the same NVT ensemble while an additional gravitational acceleration has been imposed on each water molecule, this elution simulation was allowed to run for a further 5 ns. Simulation snapshots were saved every 100 fs to examine the reproducibility of key events while visualization was performed using the Open Visualization Tool (OVITO) (version 3.8.0) [9]. The number of water molecules present in the eluent region has also been monitored to keep track of the retention time of water molecules. To study the hydrophilicity effect of Au aerogel on the retention of water molecules, the above simulation has been repeated with single positive charges set on each of the Au atoms of the aerogel and compared with its neutral hydrophobic counterpart.

5. Mass spectrometric measurements

Determination of the metabolite coverage: Bruker Autoflex MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) was employed to perform all the mass spectrometric measurements. The parameters were: positive ion mode; mass range of 5 - 1500 Da in a linear mode; operated at random walk mode; shot number: 3000; frequency: 100 Hz; Laser fluence: 66 - 158 mJ/cm^2; delayed extraction time: 10 ns. The double-layered aerogel coated on the surface of coverslip was utilized as the SALDI substrate. A series of 5 × 14 aerogel wells (3 mm ID circle on the coverslip, **Figure S2**) were prepared by using a laser marking machine (DIAOTU B3 Smart, China). Different types of metabolites, including carbohydrate (α-Cyclodextrin), polymer (PEG-1000), amino acids (L-His , L-Arg), purines (Adenine), peptides (RGDS, Leu-enk, Angiotensin II), hormone (Melatonin), and drugs (Acetylcholine, Berberine, Verapamil, Sildenafil), were employed to investigate the metabolite coverage and detection sensitivity. The metabolite stock solutions (1E-2 M)

were prepared and diluted with H_2O. The diluted solutions (in the range of 2E-5 M – 5E-9 M) were employed for the analysis. 2 μl of each sample solution was applied in the aerogel well and dried under ambient conditions. All mass spectrometric measurements were performed in replicates (n = 3 – 6), and mean values and standard derivations were calculated for comparisons.

Determination of detection sensitivity and reproducibility: Leucine-enkephalin, Sildenafil, Berberine, L-Arg, Verapamil and Acetylcholine were dissolved with 30% methanol-water solution to prepare the stock solution for testing the detection sensitivity. The stock solutions (0.01 M) were further diluted with H_2O in the range of 2E-5 M – 5E-9 M to perform the MS measurement in quadruplicated. Moreover, to determine the reproducibility of the double-layered aerogel thin film, replicated measurements (n = 4) of the same standards at the concentration of 6E-5 M – 3E-8 M were performed. The standard deviations were calculated to evaluate the fluctuation of the detection.

Determination of the salt separation and storage capability: To investigate the effect of inorganic salt on the detection sensitivity, artificial urine [10] (**Table S3**) was utilized as the sample matrix. To perform control experiments, a single layer of Au aerogel coated on the surface GOPTS derivatized coverslip was utilized as the substrate. To prepare the single layer of Au aerogel, a GOPTS derivatized glass coverslip [**2, 3**] was immersed in the Au nanoparticle solution (171 ml), 19 ml freshly prepared 1M NH_4F was then rapidly added into solution (the total solution height in the container was 3.0 cm). The solution was well mixed and then ground for 8 hours. The Au aerogel was then washed with DI water thoroughly, and dried under ambient conditions.

Determination of the heat transfer and electron transfer: To investigate the effect of heat transfer and electron transfer of the photo-irradiated double-layered aerogel on the detection sensitivities, benzylpyridinium (BP) salt (2 μl of 1E-5 M BP salt in 40% methanol) as the chemical thermometer and juglone (2 μl of 3E-3 M juglone in 40% methanol) as the electron acceptor, were respectively applied on the surface of the double-layered aerogel. The total intensities and survival yield (SY) of BP ions, and the total intensities of juglone ions were measured at different laser fluence (66 - 158 mJ/cm^2). In a control experiment, the single layer of Au aerogel was used as the substrate to investigate the effect of the silica aerogel on the heat transfer and the electron transfer. All the measurements were replicated (n = 3 - 5). Mean values and standard derivations were calculated for comparisons.

6. Discovery of Urine and Serum Metabolic Biomarkers for Lung Cancer Patients

Clinical samples collection: A total number of 190 serum and urine samples were collected from 102 participants, including 67 cancer patients (27 early-stage and 40 late-stage) and 35 healthy controls, at the Cancer Hospital of Shantou University Medical College according to the study protocol approved by the Medical Ethics Committee of Cancer Hospital of Shantou University Medical College (Permission No. 202050). Written consents have been obtained from the participants. The demographic information of the participants is provided in **Table S4**. The peripheral venous blood and midstream urine of participants were collected in the morning.

For the blood collection, vacuum serum tubes without additives (Chengdu Rich Science Industry Co., Ltd, Sichuan, China) were utilized. After standing for 1 h at room temperature for blood coagulation, the samples were centrifuged at 1500 g for 15 min. The serum (supernatant layer) was aliquoted and stored at -86 °C. To remove the proteins in the serum samples, a commonly adopted organic solvent mixture, acetonitrile/methanol (1:1, *v/v*), was used in the current study. [**11,12**] For each serum sample, 50 μl of serum was deproteinized with 150 μl ice-cold acetonitrile/methanol (1:1, *v/v*). After vortexing for a few seconds, the deproteinized samples were kept at 4 °C for 15 min, and then centrifuged at 15000 g for 15 min. Then, 1.5 μl of each sample was pipetted onto the array wells of the aerogel fixed on the MALDI plate. To reduce the duration of sample preparation, batch mode operation was adopted.

For urine collection, 30 mL urine sample was collected in an urine specimen container (Chengdu Rich Science Industry Co., Ltd, Sichuan, China). The sample was aliquoted and stored at -86 °C, and 1 ml of each sample (after centrifugation at 4 °C, at 8000g for 15 min) was used for the refractive index (refractometry) measurement [**13**] and dissolved salt measurement. To determine the amount of dissolved salt in urine samples, 1.0 ml of each urine sample was centrifugated at 8000 *g* at 4 °C for 15 min. 500 μl supernatant layer was collected and freeze-dried to determine the weight of the dissolved salt according to the weight difference method. The specific gravity (SG) measurement was performed with a refractometer. The linear correlation of dissolve salt concentration with specific gravity (as reflected from reflective index) was determined (**Figure S26**), and was utilized in the urine concentration normalization (i.e., 1.25 mg/ml with the SG of 1.002) for a fair comparison in urine metabolomics studies [**14**].

Metabolomic SALDI-MS measurements: To confirm the consistence of the quality of the slides, UV-vis absorption in the range of 300 nm to 900 nm (**Figure S29**) of the prepared slides were measured. Only those slides with the absorbance variation (at 355 nm) less than 5 % were selected for the SALDI-MS measurement. Prior to the analysis, the serum samples were thawed at 4°C, and were then mixed with ice-cold acetonitrile/methanol (1:1, *v/v*) to perform protein precipitation and extract the metabolites (with polar to semi-polar properties) for further SALDI-MS analysis. After vortexing for a few seconds, the deproteinized samples were kept at 4 °C for 15 min, and then centrifuged at 15000 *g* for 15 min. Next, about 160 μl of clarified supernatants were collected. A quality control (QC) serum sample was prepared by mixing equal aliquots of all the serum samples into a pooled sample. The urine samples were thawed at 4 °C and diluted with deionized water to the same concentration of dissolved salts

(i.e., 1.25 mg/ml for sample concentration normalization). Moreover, a QC urine sample was prepared by mixing equal aliquots of all the urine samples into a pooled sample. Then, 1.5 µl of each sample was pipetted onto the array wells of the aerogel fixed on the MALDI-plate (**Figure S2**). After the ambient drying, the samples were analyzed by the mass spectrometer operated in the linear positive mode to maximize the detection sensitivity and record as many features as possible. The parameters were: positive ion mode; mass range of 5 - 1500 Da in a linear mode; operated at random walk mode; shot number: 10000; frequency: 333.3 Hz; Laser fluence: 158 mJ/cm^2; delayed extraction time: 10 ns. Replicate measurements (n = 5) were performed for each sample, and the QC samples were designed to run after every 10 sample wells analyzed. In the current study, to make the best use of enhanced E-field promoted ionization property of the aerogel, positive ion mode was adopted to perform the metabolomics characterization of biofluids of lung cancer patients.

SALDI-MS Data Processing: All SALDI-MS mass spectra data were converted to the mzML format. The MALDIquant package [**15**] of R programming were utilized for data processing and peak extraction, including log2 transformation, SavitzkyGolay smoothing, SNIP baseline correction and mass value alignment, followed by the averaging of five technical replicate data. The parameters of peak detection are: signal-to-noise ratio = 2; half window size = 60; peak bin tolerance=0.00083. The normalization was performed to the obtained data matrices by a MS total useful signal method [**16**] with a "home-built" macro in Excel. Finally, the data matrices were utilized for the following analysis.

Feature Selection: The samples were randomly assigned to (1) a discovering cohort of 19 early-stage patients, 26 late-stage patients and 25 healthy controls; (2) a validation cohort of 6 early-stage patients, 10 late-stage patients and 9 healthy controls (**Figure 6a, Figure S27**). The principal component analysis (PCA, **Figure 6b**, **Figure S19**, **Figure S20**) was performed with MetaboAnalyst 5.0 (McGill University, Montreal, Canada) [**17**]. The sparse partial least squares discriminant analysis (sPLS-DA, **Figure S19**, **Figure S20**) was performed by the mixOmics package [**18**] in the R programming and the parameters of the sPLS-DA were tuned by cross-validation. The orthogonal partial least squares discriminant analysis (OPLS-DA, **Figure S19**, **Figure S20**) was performed with SIMCA-P 14.1 software (version 14.1; Umetrics AB, Umeå, Sweden) and pareto scaling was employed in combination with mean-centering. The elasticnet was performed using Scikit-learn 1.20 package [**19**] in Python 3.9 and StandardScaler was applied in combination with mean-centering. The t-test was performed with MetaboAnalyst 5.0. The important features were screened according to (1) Loading value >0.1 in sPLS-DA; (2) VIP $\geq$ 1, |p(corr)| $\geq$ 0.5 and |p| $\geq$ 0.05 in OPLS-DA; (3) model selected frequency with repeat occurrence over 50% in 100 models of elasticnet; (4) p value < 0.05 according to two-sided t-test; (5) the reliable signal to noise ratio (S/N > 2); (6) the presence of the pair co-existence relationships of [M+H]$^+$, [M+Na]$^+$ or [M+K]$^+$. In the compound identification, the accurate mass of distinctive biomarkers was determined in the reflectron mode. Mass calibration was performed with the standard calibration mixture at the mass precision of 50 ppm, and the mass resolution (at m/z 494) was approximately 9000 (FWHM). Moreover, the high-resolution mass spectrometer (Q Exactive Focus Orbitrap, Thermo Fisher Scientific) was utilized to perform the accurate mass and tandem MS measurement, and the MS/MS patterns were

compared with standard references and/or databases (Human Metabolome Database) for the biomarker identity confirmation (**Table S5**). For those biomarkers without the reference standard comparison, the MS/MS fragment ions were manually assigned with reference to the structure of the biomarkers (**Figure S30**).

Model Validation: The support vector machine (SVM), random forest (RF) and logistic regression (LR) were performed using Scikit-learn 1.20 package in Python 3.9. The parameter of SVM was kernel = rbf. The parameters of RF were n_estimators = 8, max_depth = None, min_samples_split = 2. The 200 rounds cross-validation was performed for the evaluation of the classification capability of the characteristic metabolites. Moreover, 26.3% data were assigned as an independent cohort (validation cohort) for the validation of the stability and generalization capability for lung cancer screening of the characteristic metabolites. Additionally, the biological and clinical significance of the characteristic metabolites discovered in serum (**Figure 6c-i,ii**) and biofluids (serum and urine) (**Figure 6c-iii,iv**) were analyzed by the enrichment analysis via MetaboAnalyst 5.0, respectively.

II. Supporting Figures
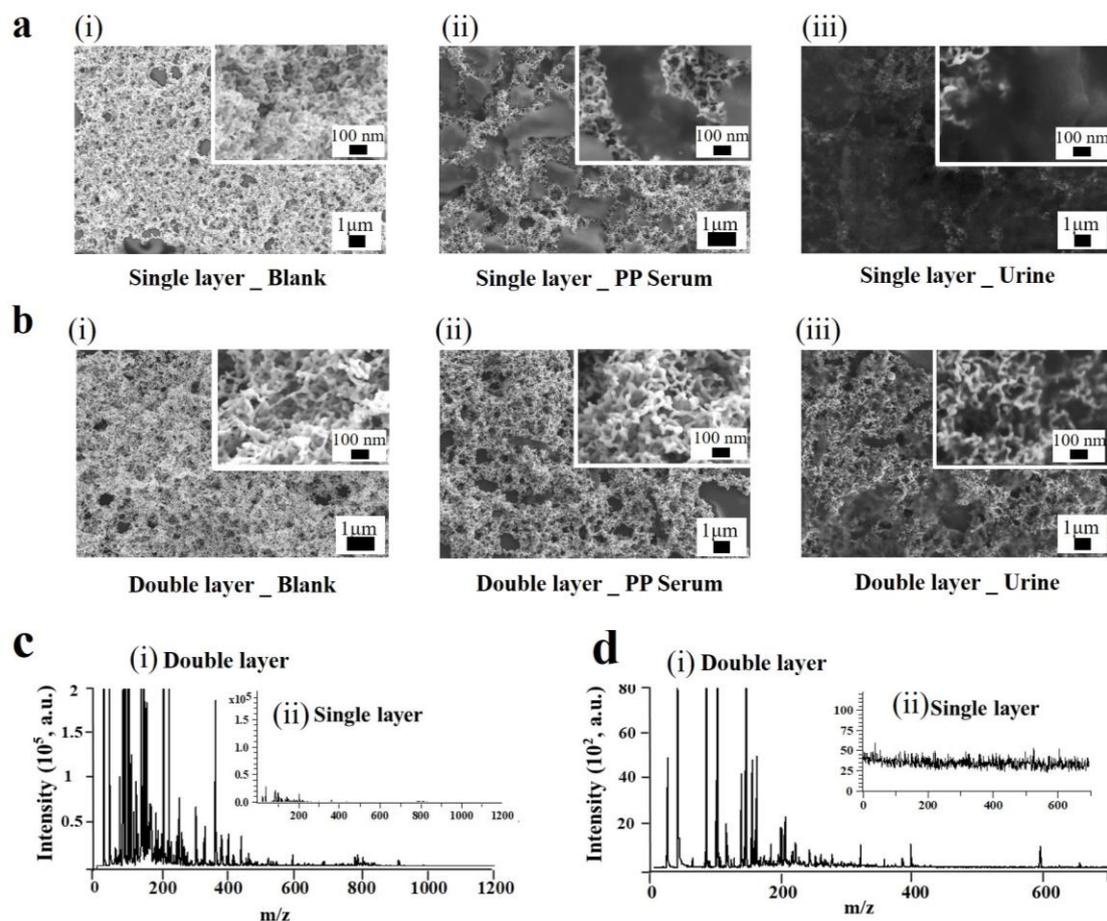
Figure S1. SEM images of (**a**) Au single-layered aerogel and (**b**) Au/SiO$_2$ double-layered aerogel: (i) without applying any sample, (ii) with the application of protein-precipitated serum, (iii) with the application of urine; (**c**) SALDI-MS spectra of protein precipitated serum on (i) Au/SiO$_2$ double-layered aerogel and (ii) Au single-layered aerogel; (**d**) SALDI-MS spectra of urine on (i) Au/SiO$_2$ double-layered aerogel and (ii) Au single-layered aerogel.

Au/SiO$_2$ aerogel coverslip stuck on MALDI plate

Without laser marking
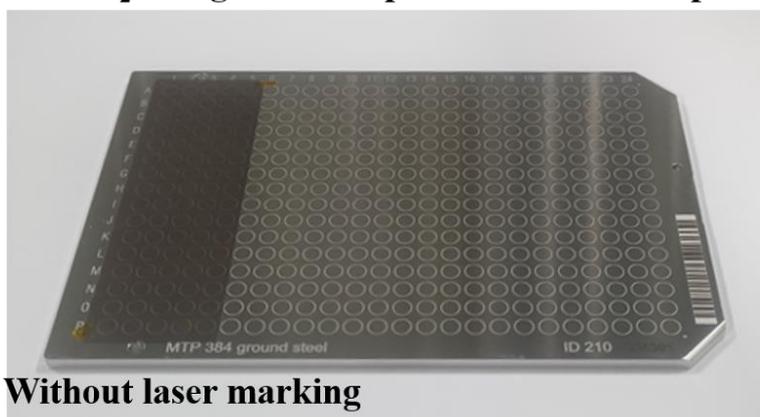
laser
marking

Application of biofluid on the sample well

Biofluids
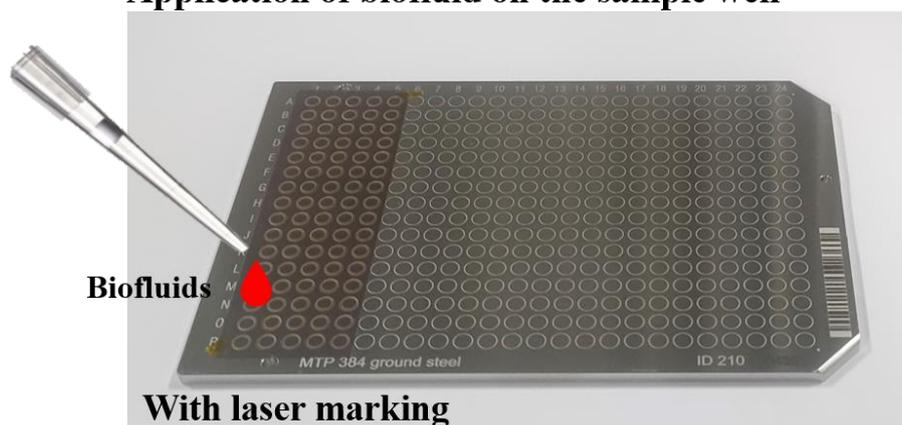
With laser marking

Figure S2. Au/SiO$_2$ double-layered aerogel coverslip stuck on the MALDI plate for the laser marking to generate the sample wells, and application of biofluid on the sample well.

Spin coating times	Thickness of SiO$_2$ aerogel (nm)
1	100 ± 18.1
2	202 ± 22.0
3	448 ± 17.1
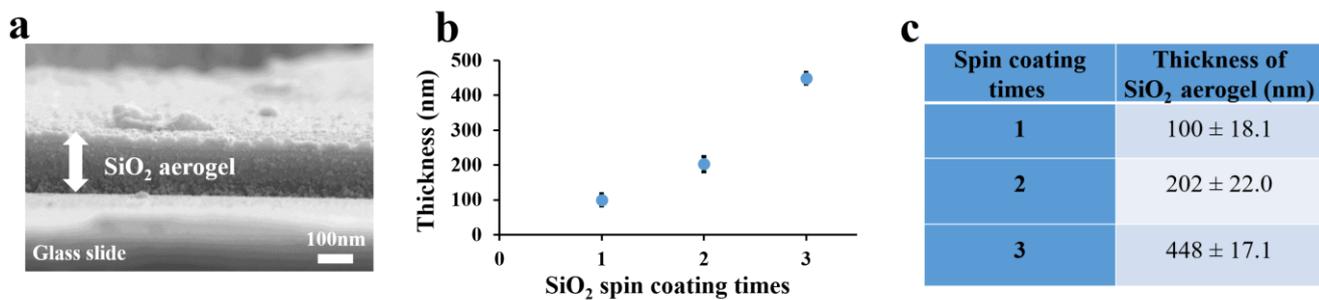
Figure S3. (**a**) SEM image of SiO$_2$ aerogel layer coated on a glass slide, (**b** and **c**) effect of spin coating times on the thickness of SiO$_2$ aerogel.

(a)

(b)
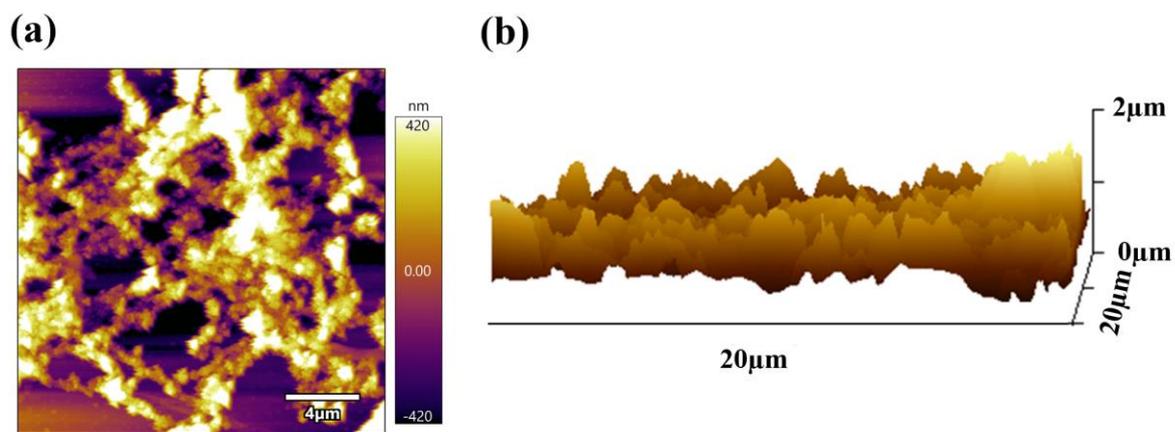
(c)

Aerogel	Surface Roughness (*Sa*)			
	1st Measurement	2nd Measurement	3rd Measurement	Mean
Au-SiO$_2$	265.4 nm	243.3 nm	212.2 nm	240.3 nm

Figure S4. Surface roughness of AuSiO$_2$ aerogel: (**a**, **b**) AFM images showing the topology; (a) summary of surface roughness (*Sa*), derived from the arithmetic mean of the absolute deviations from the mean surface level of the aerogel.

a SEM image of SiO$_2$ aerogel

(i)
Cracking of SiO$_2$ aerogel on glass slide without GOPTS derivatization

(ii)
Intact of SiO$_2$ aerogel on glass slide with GOPTS derivatization

b Photography of Au/SiO$_2$ aerogel

(i)
Cracking of Au aerogel on SiO$_2$ aerogel without GOPTS derivatization

(ii)
Intact of Au aerogel on SiO$_2$ aerogel with GOPTS derivatization
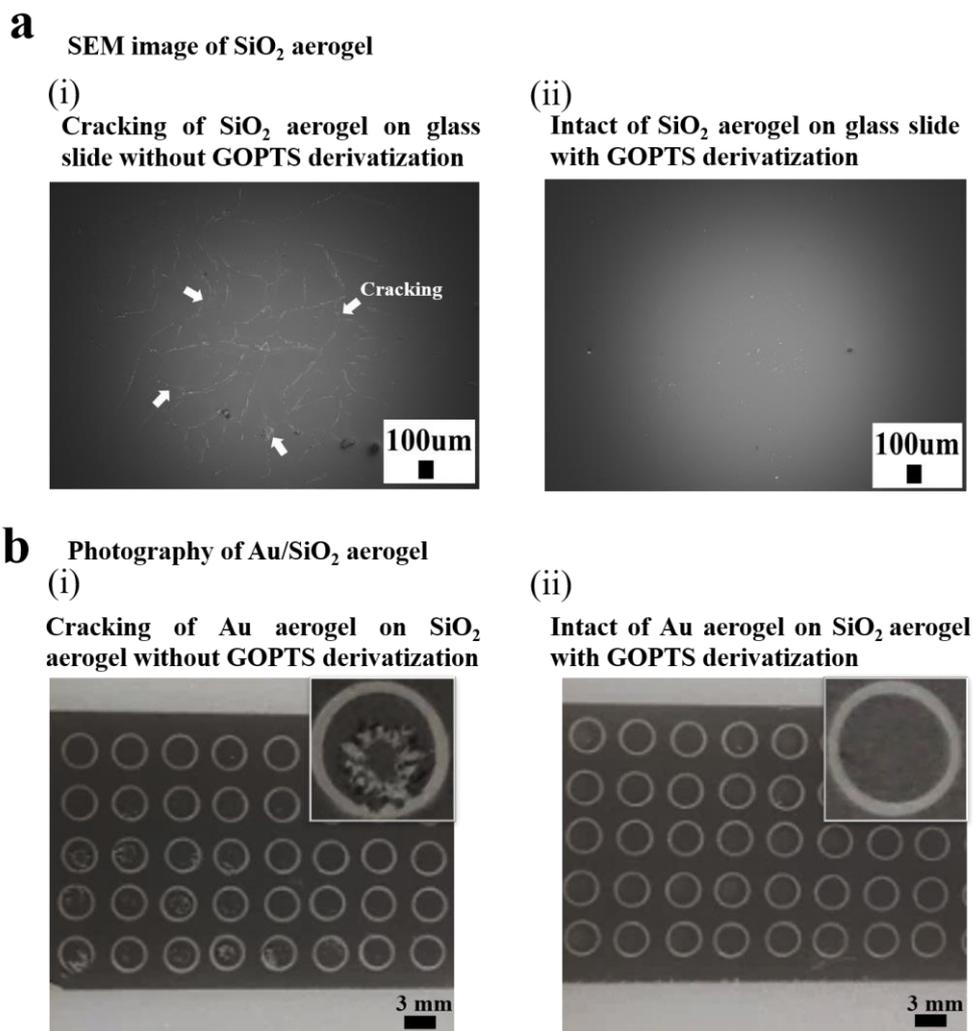
Figure S5. (**a**) SEM images of the SiO$_2$ aerogel coated on the (i) normal glass coverslip and (ii) GOPTS derivatized glass coverslip, showing the integrity of SiO$_2$ aerogel on the coverslip; (**b**) the photograph of the Au aerogel coated on the (i) normal SiO$_2$ aerogel and (ii) GOPTS derivatized SiO$_2$ aerogel, showing its integrity after the drying of the applied sample solution.
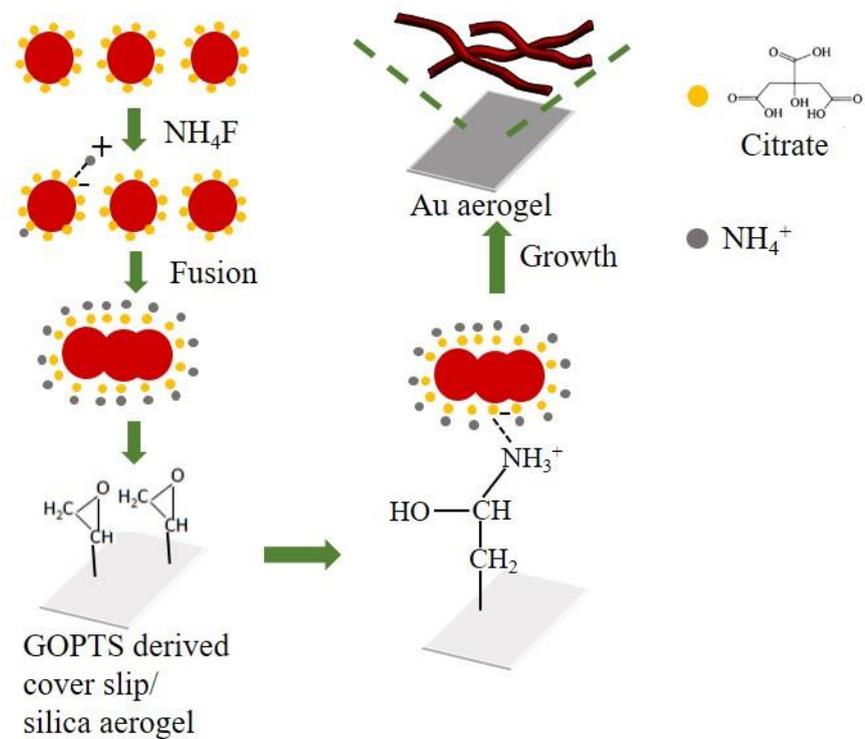
Figure S6. Schematic diagram showing the proposed mechanism of GOPTS surface modification for binding the Au aerogel layer.

a (i)

(ii)

Height of Au solution (cm)	Thickness of Au aerogel (nm)
1	565 ± 85
2	714 ± 168
3	1032 ± 181
4	1216 ±178

b

c

d (i)   (ii)   (iii)   (iv)

565 nm   714 nm   1032 nm   1216 nm
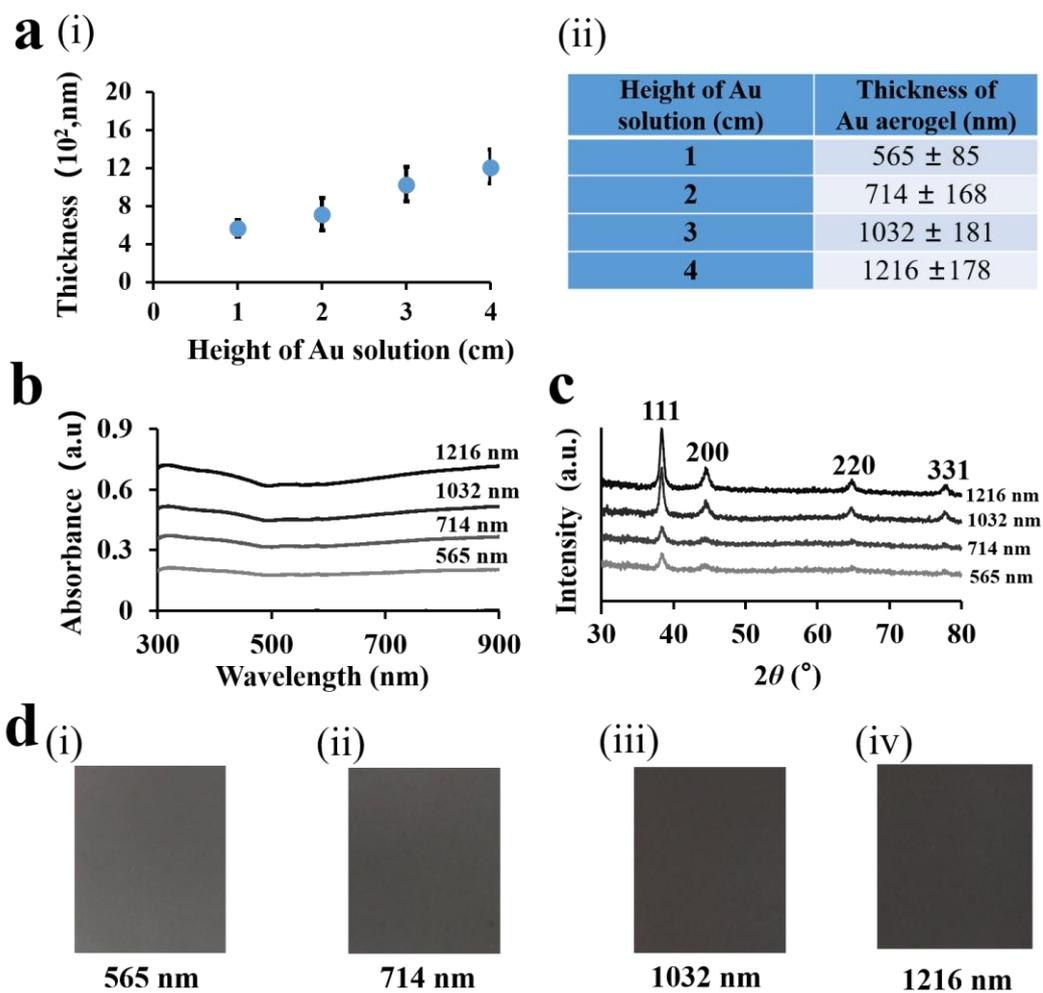
Figure S7. Characterization of Au aerogel. **(a)** Effect of Au solution height on the thickness of Au aerogel; **(b)** UV-visible absorption spectra; **(c)** XRD spectra; and **(d)** photographs.
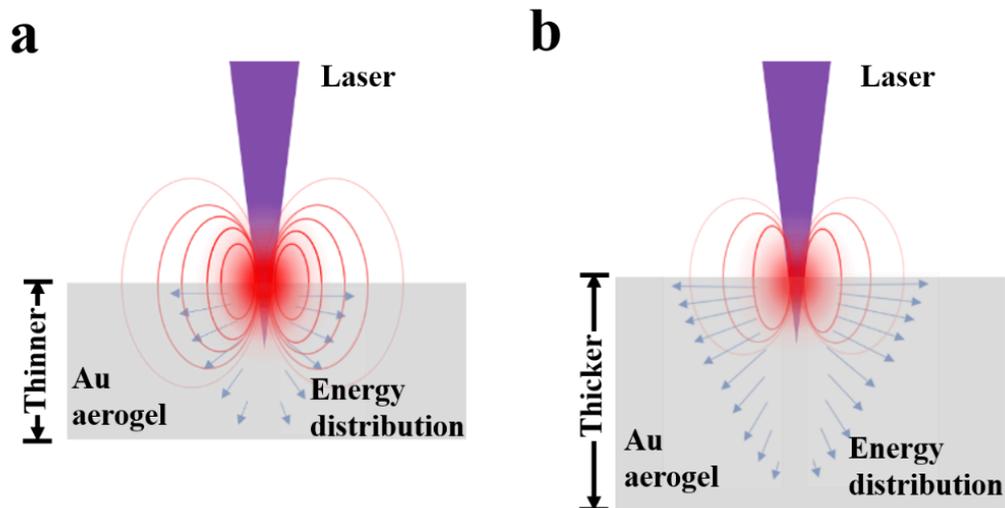
Figure S8. Schematic diagram showing the distribution of absorbed energy along the depth of the Au aerogel with different thickness: (**a**) thinner and (**b**) thicker Au aerogel. The further thickening of the Au aerogel can make the absorbed energy distributed along the depth, thus making the energy less localized on the surface of the Au aerogel, and weakening the ionization efficiency.
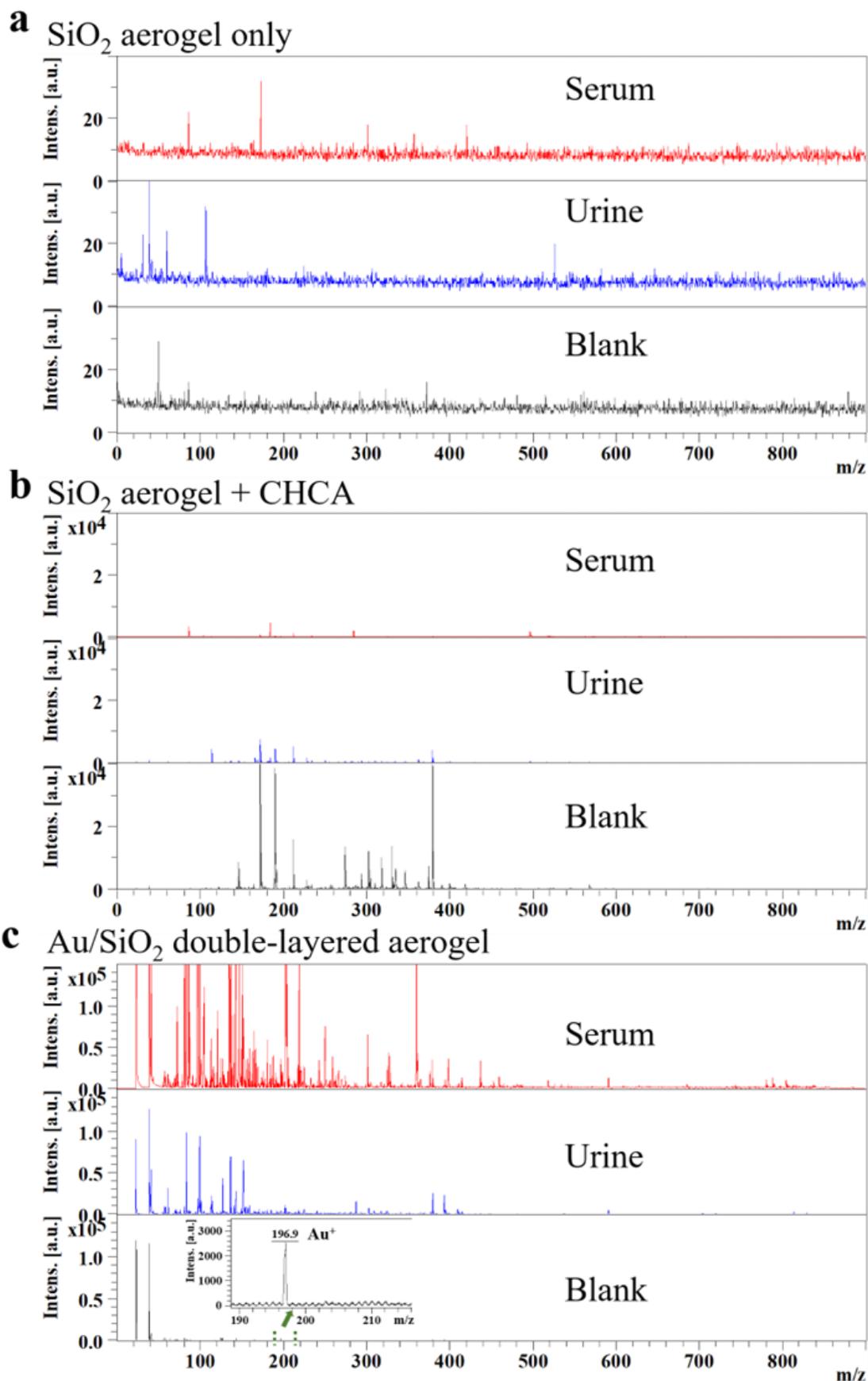
Figure S9. SALDI-MS spectra of protein precipitated serum, urine and blank control on (**a**) single-layered SiO_2 aerogel, (**b**) the organic matrix (CHCA) added SiO_2 aerogel and (**c**) the double-layered Au/SiO_2 aerogel.
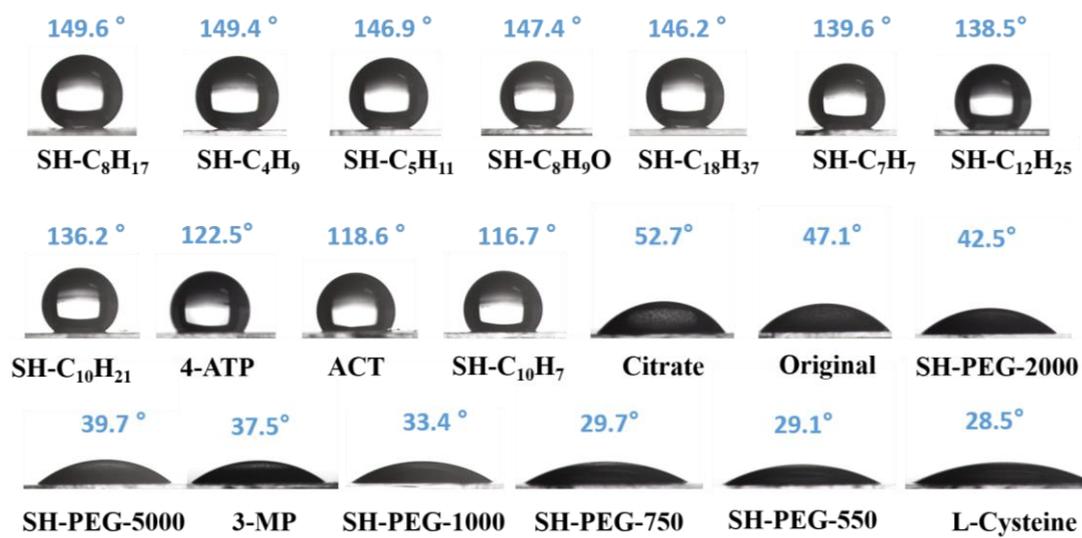
Figure S10. Characterization of the contact angles of surface functionalized Au aerogel.
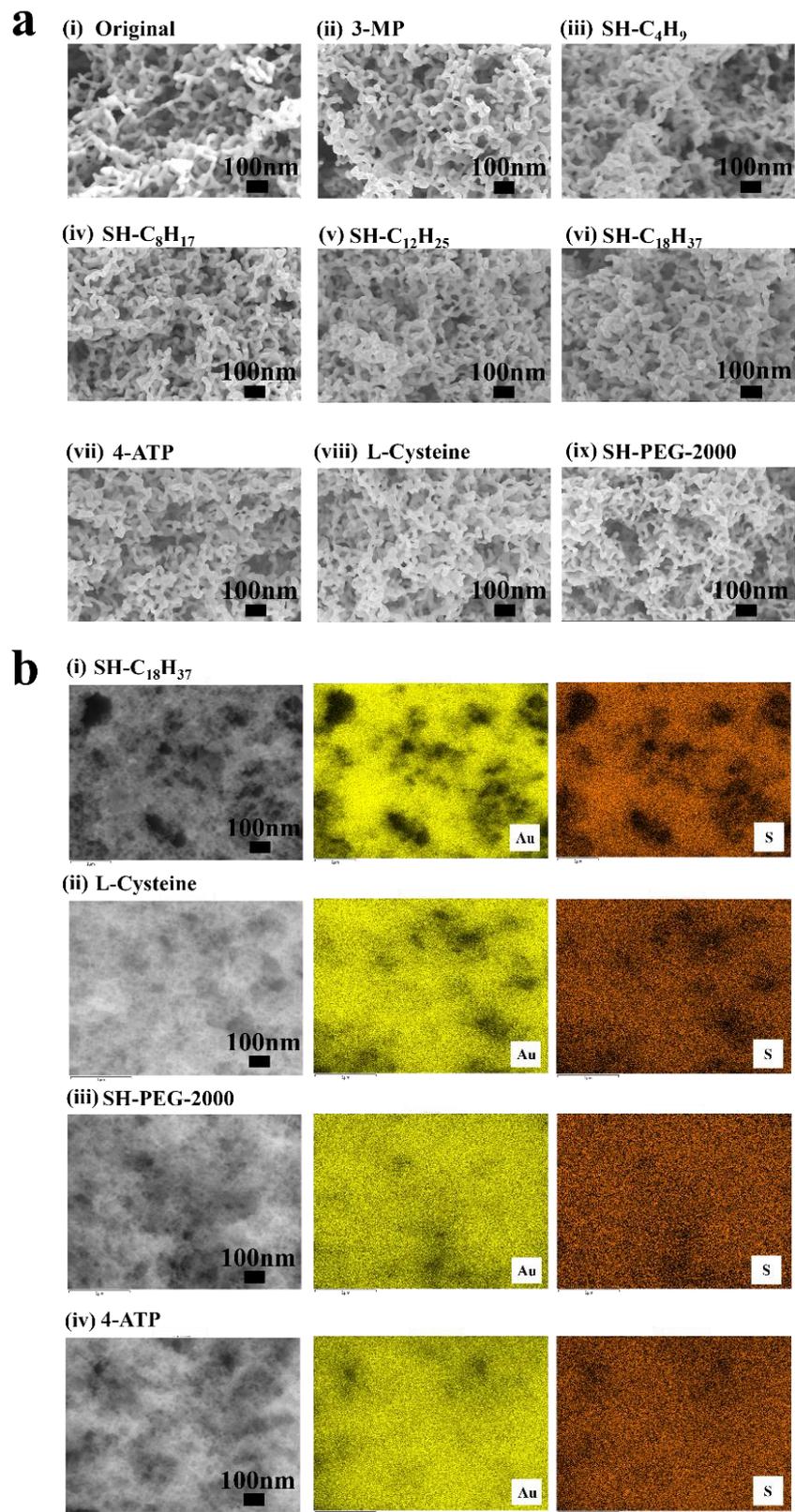
Figure S11. Characterization of surface functionalized Au aerogel. (**a**) SEM images; and (**b**) EDS elemental mapping.
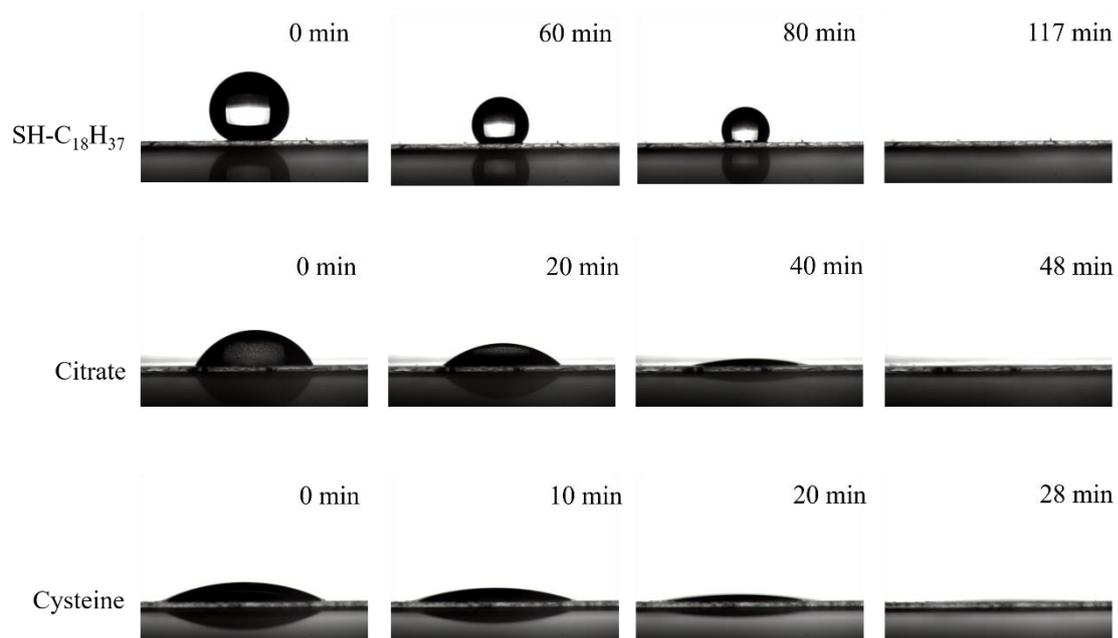
Figure S12. Photograph of fluid droplet showing the higher surface hydrophobicity can maintain the fluid droplet on the Au aerogel surface for a longer period.
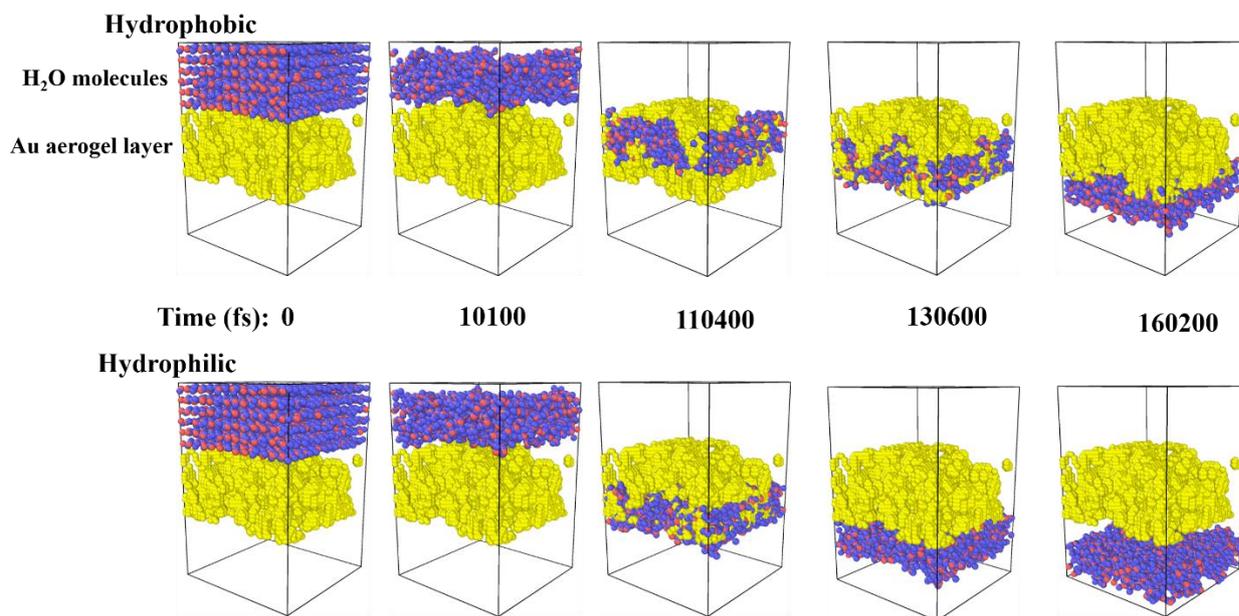
Figure S13. The LAMMP molecular dynamics (MD) simulation showing the water molecules passing through the hydrophobic surface/pore is slower than that passing through the hydrophilic surface/pore.
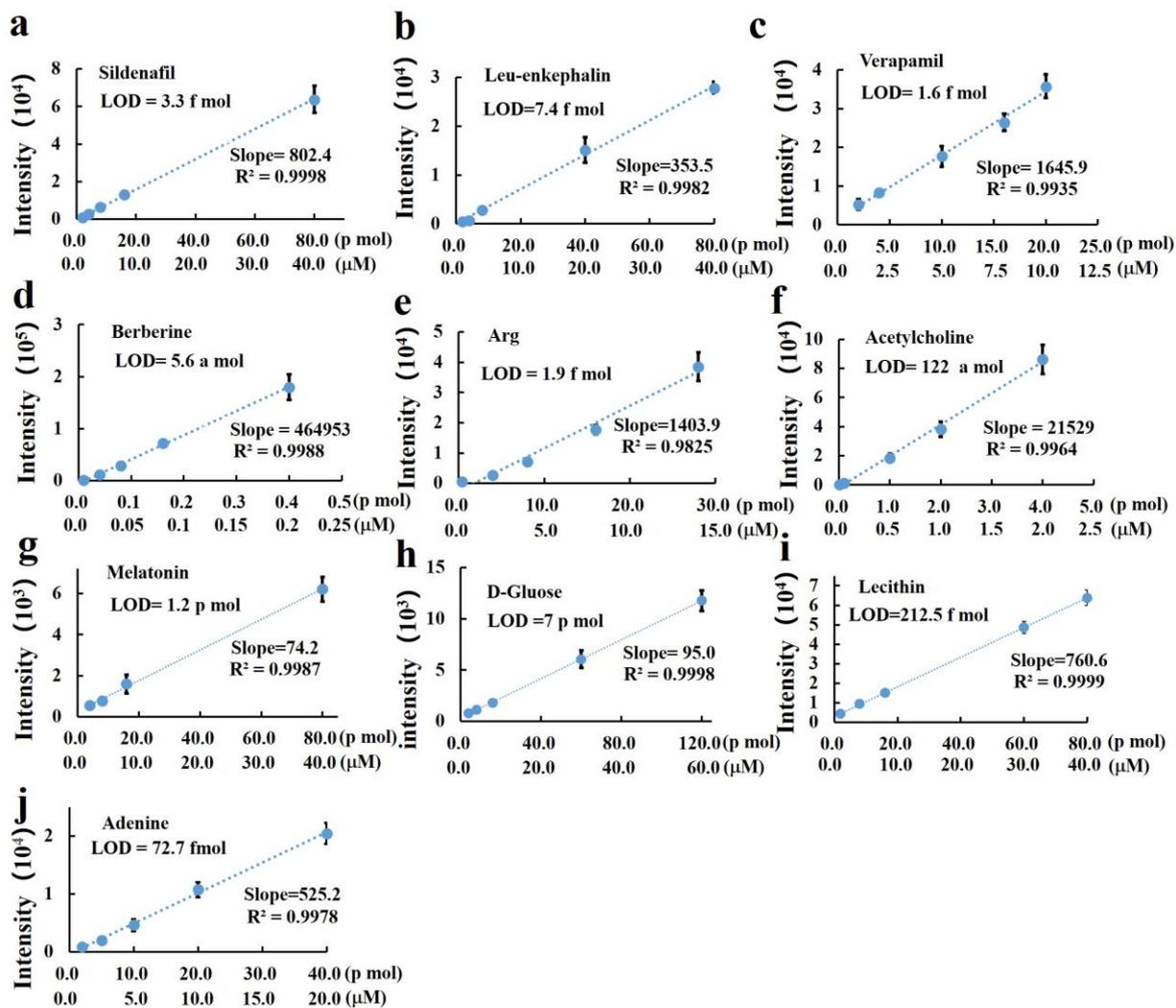
Figure S14. Detection limits of metabolite standards determined at the laser fluence of 106 mJ/cm^2 using Au/SiO$_2$ double-layered aerogel as the substrate.
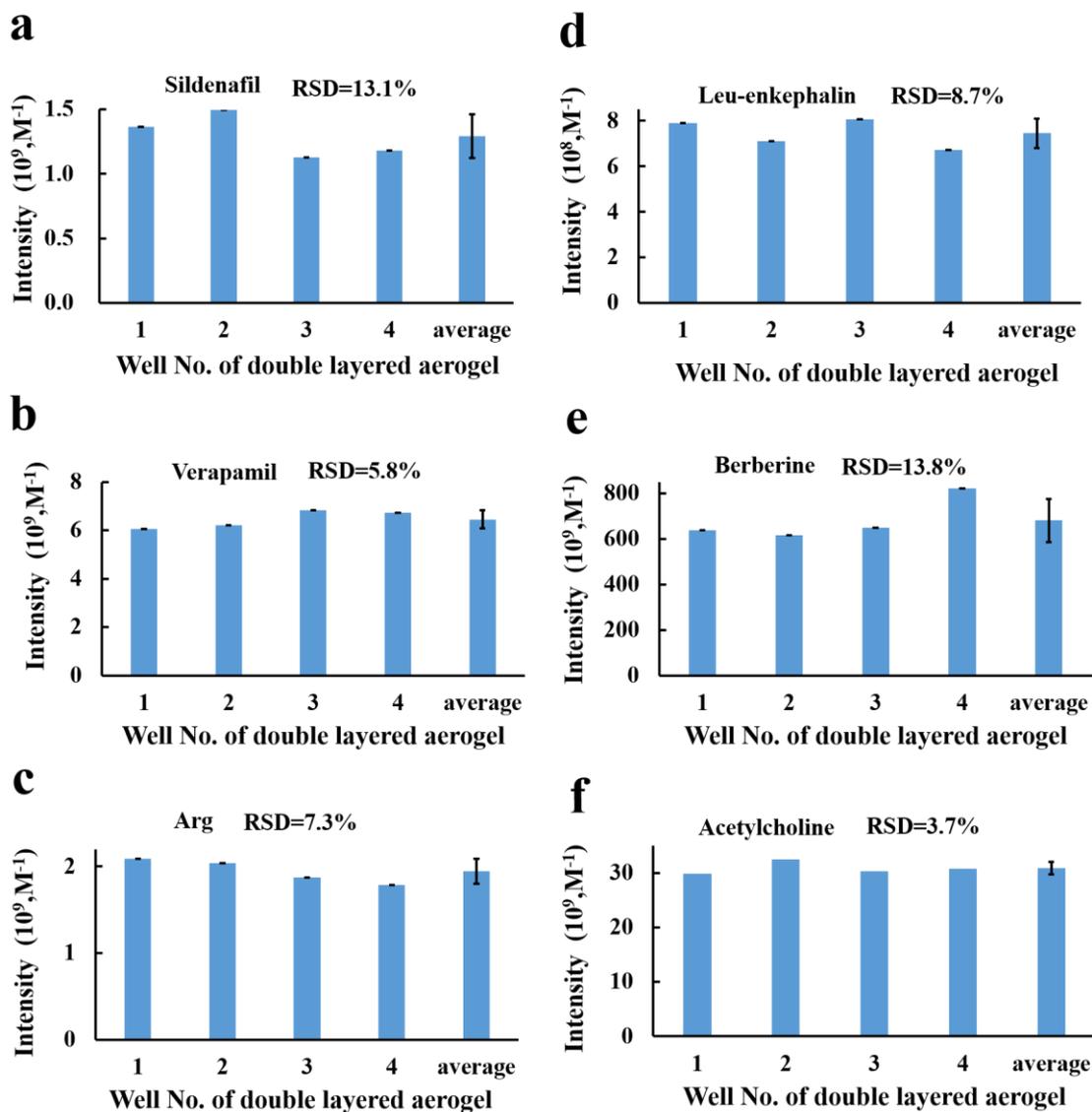
Figure S15. Well-to-well detection reproducibility of the Au/SiO$_2$ double-layered aerogel measured at the laser fluence of 106 mJ/cm^2. 2.0 µl aqueous solution of the analytes, including sildenafil citrate (8E-6 M), verapamil (1E-6 M), L-Arg (8E-6 M), leu-enkephalin (6E-5 M), berberine (3E-8 M), and acetylcholine (2E-6 M) were applied onto the array wells of the aerogel for determining the well-to-well detection reproducibility.
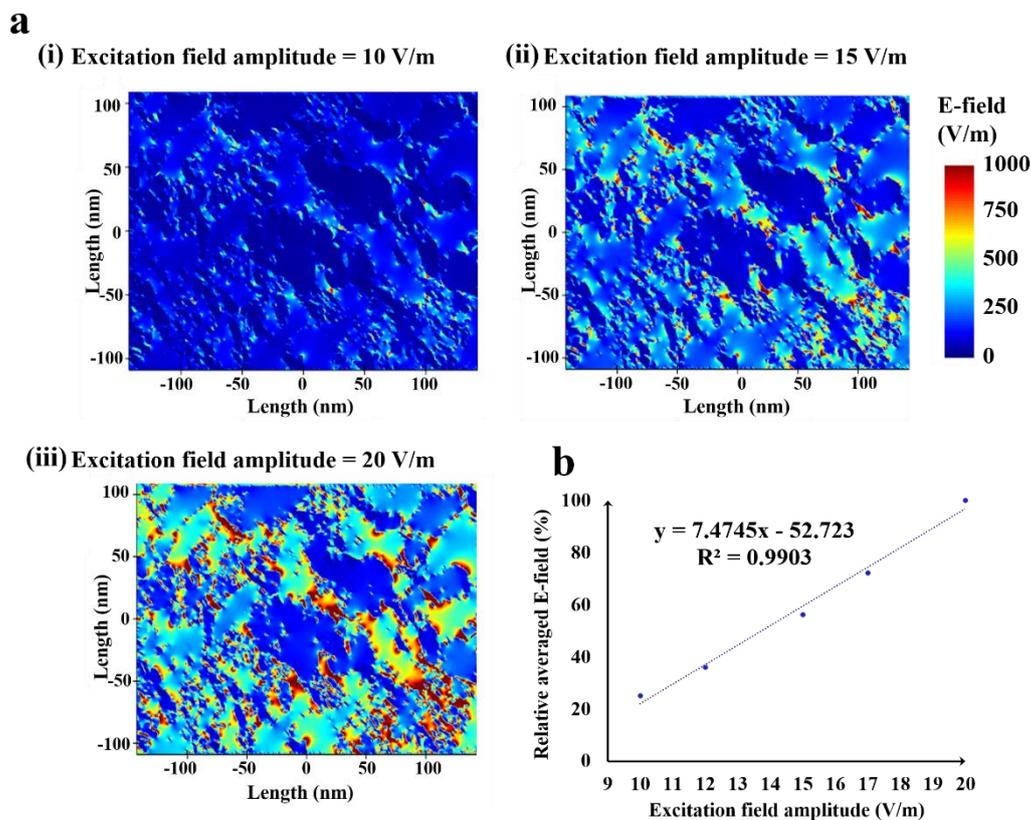
Figure S16. The FDTD simulation of E-field distribution on the photoexcited Au aerogel at (**a**) different excitation field amplitude of (i) 10, (ii) 15 and (iii) 20 V/m; (**b**) the correlation plot showing the enhanced E-field increased with the excitation field amplitude.
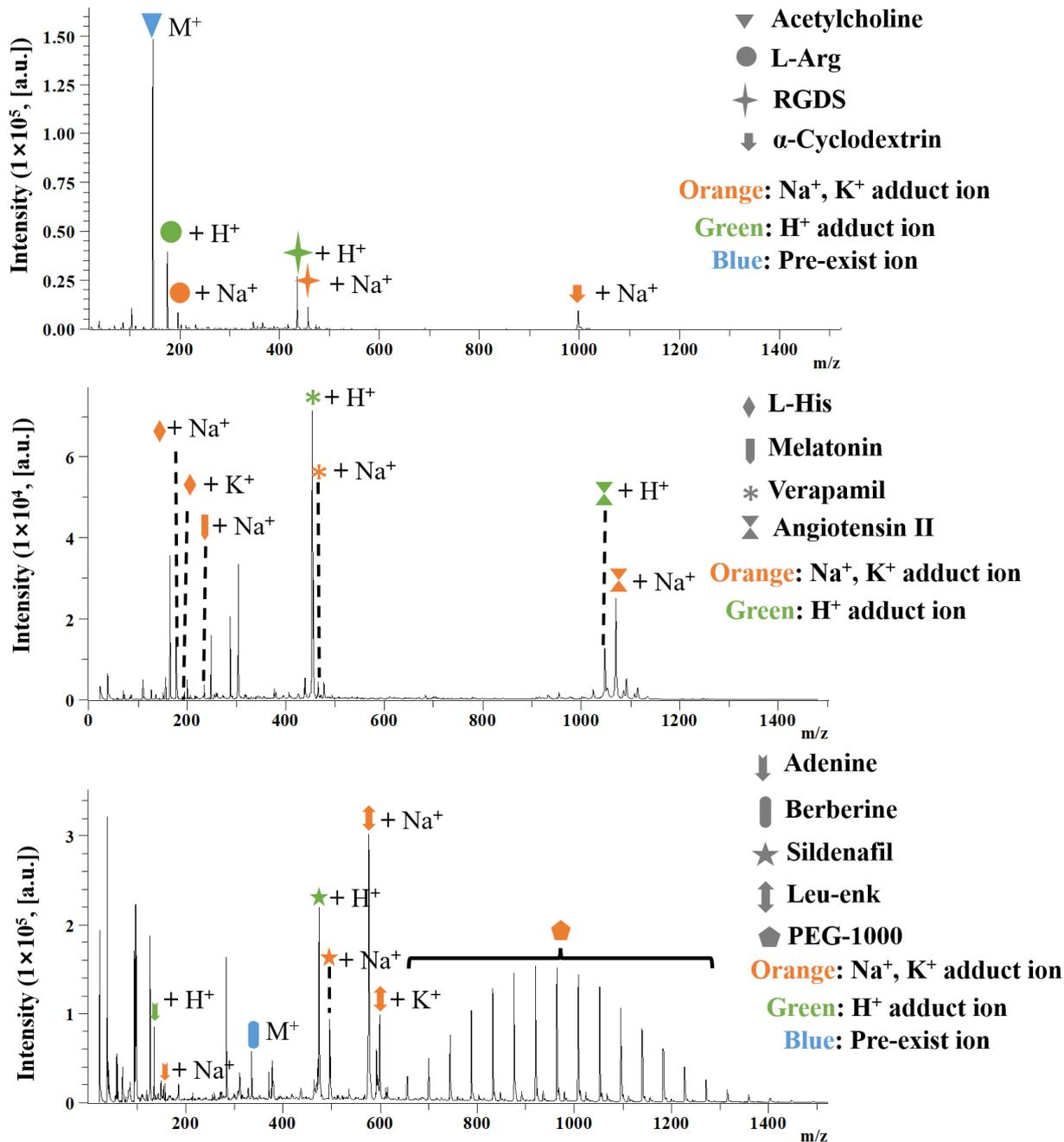
Figure S17. SALDI-MS spectra recorded by using the C_{18} functionalized Au single-layered aerogel as substrate, showing the protonation and metal cationization (Na$^+$, K$^+$) of analytes.
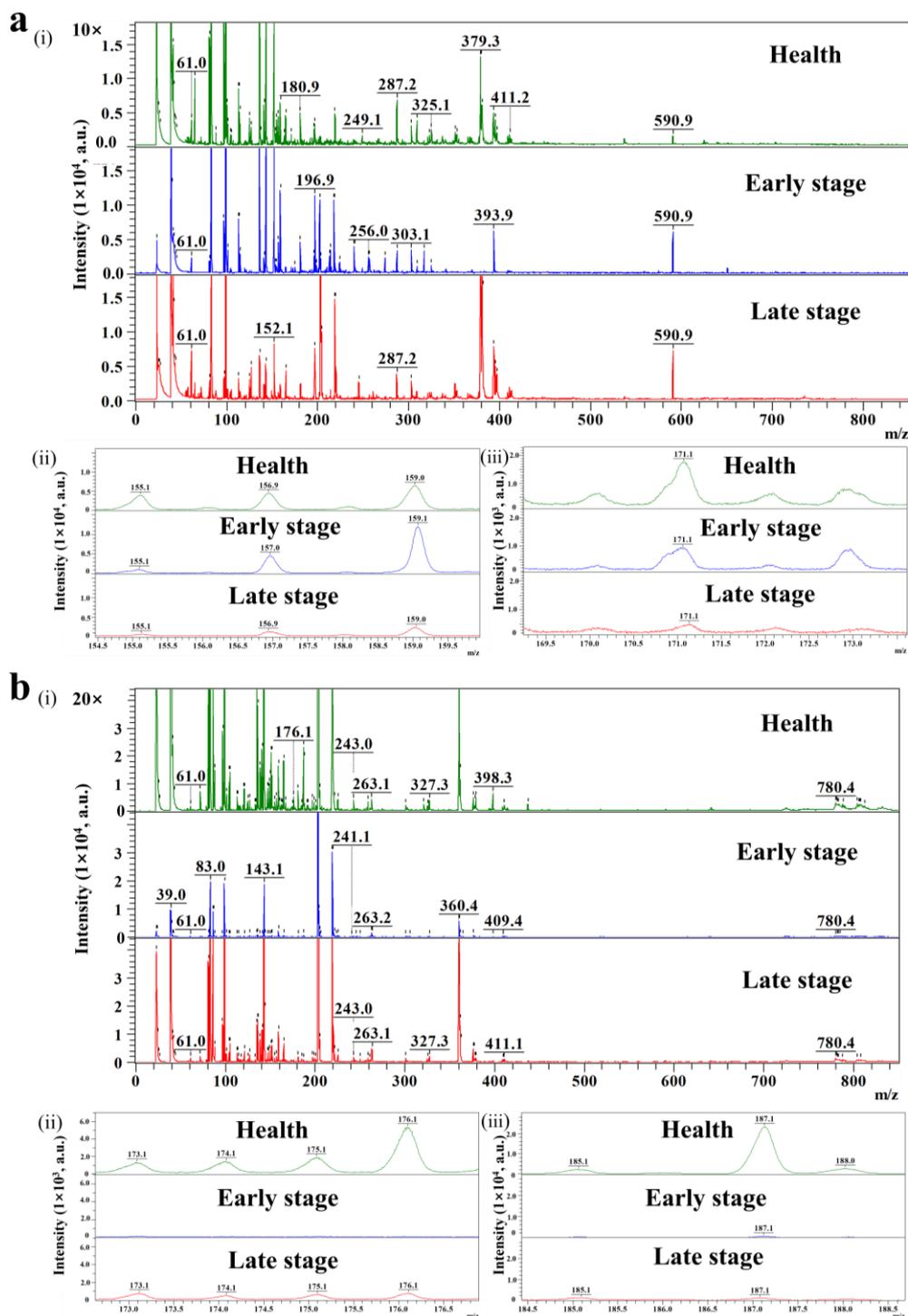
Figure S18. Representative SALDI-MS spectra of clinical samples from early-stage, late-stage lung cancer and the health individual. (**a**) Urine samples from early-stage, late-stage lung cancer and the health: (i) overview of SALDI-MS spectra, the intensity of the mass spectra is magnified by 10× for better visualization, (ii) some selected feature peaks for differentiating the health and early-stage patient, (iii) some selected feature peaks for differentiating the health and late-stage patient. (**b**) Serum samples from early-stage, late-stage lung cancer and the health: (i) overview of SALDI-MS spectra, the intensity of the mass spectrum is magnified by 20× for better visualization, (ii) some selected feature peaks for differentiating the health and early-stage patient, (iii) some selected feature peaks for differentiating the health and late-stage patient.
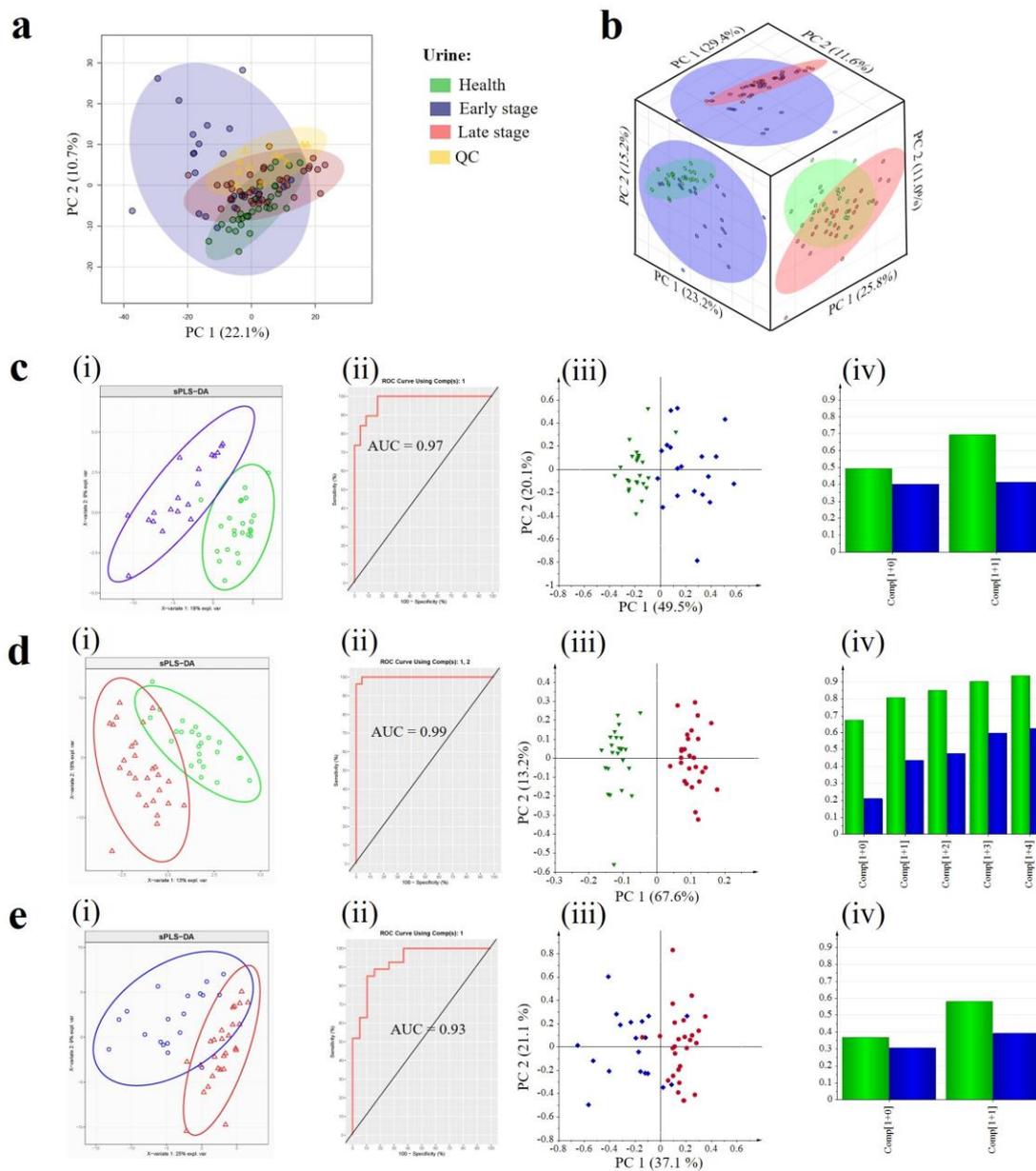
Figure S19. (**a**) PCA score plot of the urine samples and QC samples with all features; (**b**) PCA score plots show the tendency of the separation between the early-stage patients, late-stage patients and healthy group with all features from urine samples; (**c**) the characteristic metabolites for the differentiation of health individuals and early-stage patients: (i) PCA score plot and (ii) cross-validation ROC curve of the sPLS-DA; (iii) PCA score plot and (iv) summary of fit of the OPLS-DA; (**d**) the characteristic metabolites for the differentiation of health individuals and late-stage patients: (i) PCA score plot and (ii) cross-validation ROC curve of the sPLS-DA; (iii) PCA score plot and (iv) summary of fit of the OPLS-DA; (**e**) the characteristic metabolites for the differentiation of early-stage patients and late-stage patients: (i) PCA score plot and (ii) cross-validation ROC curve of the sPLS-DA; (iii) PCA score plot and (iv) summary of fit of the OPLS-DA.

Figure S20. (**a**) PCA score plot of the serum samples and QC samples with all features; (**b**) PCA score plots show the tendency of the separation between the early-stage patients, late-stage patients and healthy group with all features from serum samples; (**c**) the characteristic metabolites for the differentiation of health individuals and early-stage patients: (i) PCA score plot and (ii) cross-validation ROC curve of the sPLS-DA; (iii) PCA score plot and (iv) summary of fit of the OPLS-DA; (**d**) the characteristic metabolites for the differentiation of health individuals and late-stage patients: (i) PCA score plot and (ii) cross-validation ROC curve of the sPLS-DA; (iii) PCA score plot and (iv) summary of fit of the OPLS-DA; (**e**) the characteristic metabolites for the differentiation of early-stage patients and late-stage patients: (i) PCA score plot and (ii) cross-validation ROC curve of the sPLS-DA; (iii) PCA score plot and (iv) summary of fit of the OPLS-DA.

Figure S21. Venn diagram of the important features of urine samples discovered by sPLS-DA, OPLS-DA and elasticnet for the differentiation of (**a**) health individuals and early-stage patients; (**b**) health individuals and late-stage patients; (**c**) early-stage patients and late-stage patients.
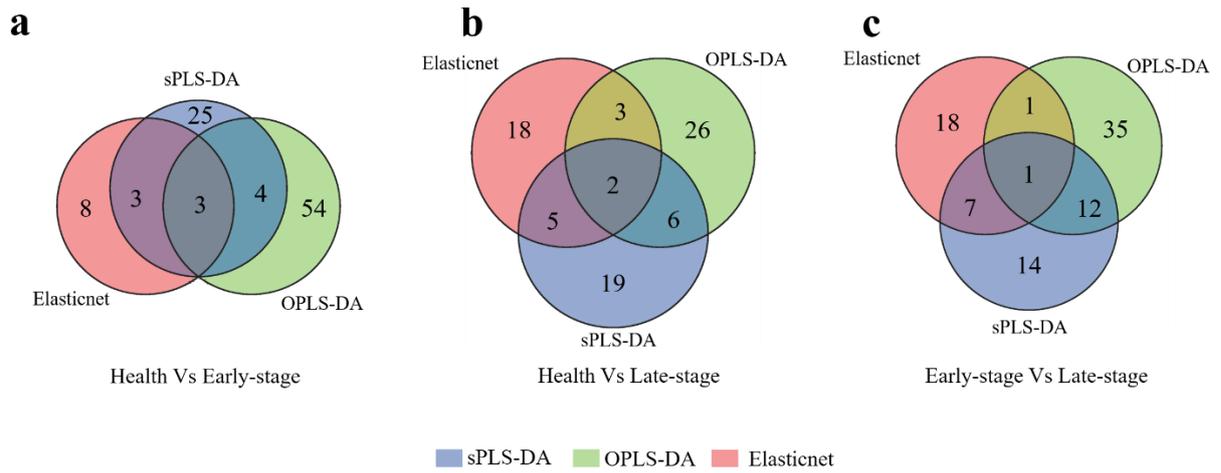
Figure S22. Venn diagram of the important features of serum samples discovered by sPLS-DA, OPLS-DA and elasticnet for the differentiation of (**a**) health individuals and early-stage patients; (**b**) health individuals and late-stage patients; (**c**) early-stage patients and late-stage patients.
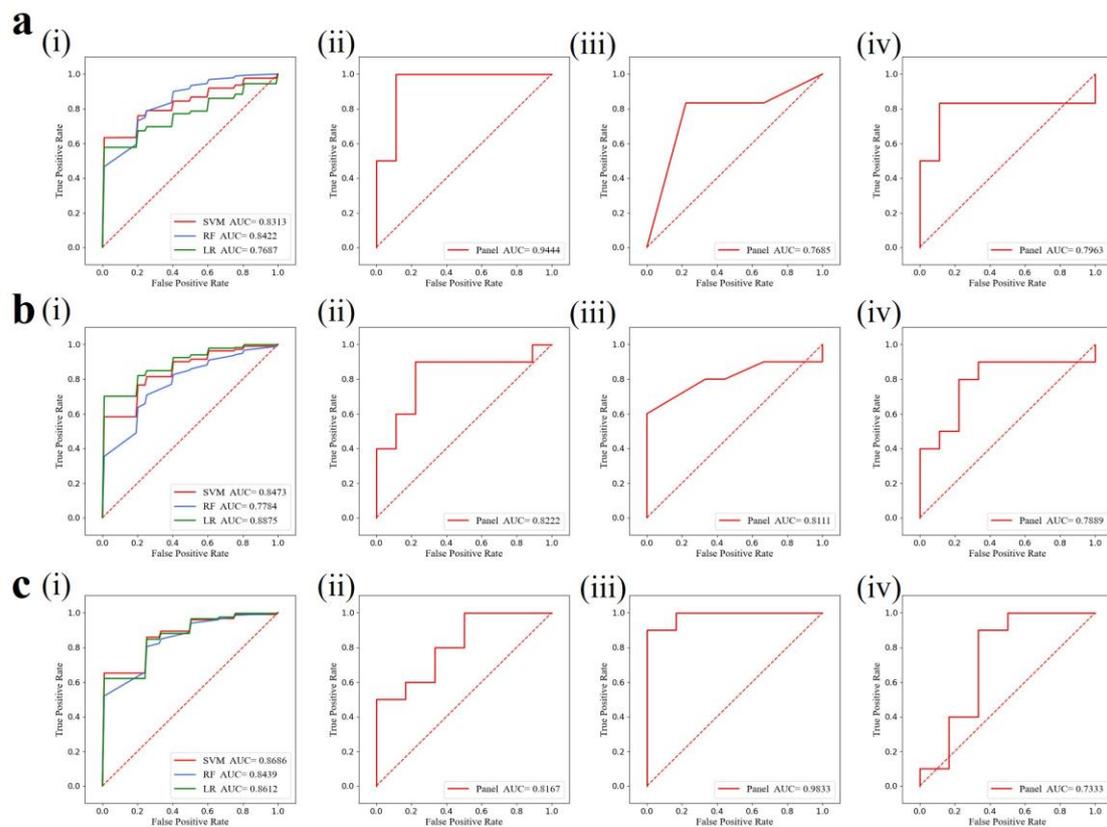
Figure S23. ROC curve for the differentiation of the early-stage patients, late-stage patients and health individuals with the 14 identified characteristic metabolites from urine samples: (**a**) the model performance for the classification of the health individuals and early-stage patients: (i) the 200 rounds cross-validation in the training cohort by SVM, RF and LR; the prediction for the validation cohort by (ii) SVM, (iii) RF and (iv) LR; (**b**) the model performance for the classification of the health individuals and late-stage patients: (i) the 200 rounds cross-validation in the training cohort by SVM, RF and LR; the prediction for the validation cohort by (ii) SVM, (iii) RF and (iv) LR; (**c**) the model performance for the classification of the early-stage patients and late-stage patients: (i) the 200 rounds cross-validation in the training cohort by SVM, RF and LR; the prediction for the validation cohort by (ii) SVM, (iii) RF and (iv) LR.

Figure S24. ROC curve for the differentiation of the early-stage patients, late-stage patients and health individuals with the 14 identified characteristic metabolites from serum samples: (**a**) the model performance for the classification of the health individuals and early-stage patients: (i) the 200 rounds cross-validation in the training cohort by SVM, RF and LR; the prediction for the validation cohort by (ii) SVM, (iii) RF and (iv) LR; (**b**) the model performance for the classification of the health individuals and late-stage patients: (i) the 200 rounds cross-validation in the training cohort by SVM, RF and LR; the prediction for the validation cohort by (ii) SVM, (iii) RF and (iv) LR; (**c**) the model performance for the classification of the early-stage patients and late-stage patients: (i) the 200 rounds cross-validation in the training cohort by SVM, RF and LR; the prediction for the validation cohort by (ii) SVM, (iii) RF and (iv) LR.

Figure S25. Box-plots illustrate the relative abundances of 11 characteristic metabolites discovered in (**a**) serum and (**b**) urine associated with lung cancer. The averaged CV of the characteristic metabolites in serum samples was 24.4%, while that in urine samples was 34.2%.
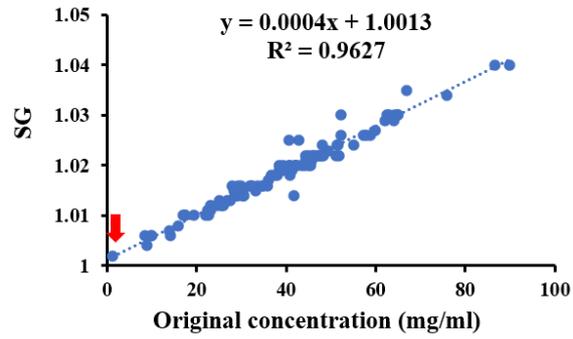
Figure S26. The linear correlation of dissolved salt concentration with specific gravity of urine samples. (*Red arrow: The reference point for the normalization of sample concentration.*)

190 biofluid samples

Discovery cohort
(140 samples,
527 ion peaks)

Validation cohort
(50 samples)

validating: ROC curve

Models for feature peaks discovery

sPLS-DA, OPLS-DA,
elasticnet

Prediction model validation

sPLS-DA: Loading value >0.1
OPLS-DA: p value>0.05, p(corr) >0.5, VIP>1
Elasticnet: Frequency>50%

t-test: P<0.05
S/N >2

200 rounds cross-
validation: ROC curve

126 (urine) and
137 (serum) feature peaks

SVM, RF, LR

The presence of the pair
co-existence relationships
of [M+H]$^+$, [M+Na]$^+$ or
[M+K]$^+$

Accurate MS
Tandem MS

Models for lung
cancer diagnosis

14 (urine) and 14 (serum)
identified characteristic
metabolites

Training cohort
(140 samples, 14 (urine) and 14
(serum) identified characteristic
metabolites

Figure S27. Schematic workflow of the metabolic biomarker discovery in biofluid samples for lung cancer diagnosis, including feature peaks discovery, identification of metabolic biomarkers, cross-validation and out-of-bag validation of the diagnosis model derived from the metabolic biomarkers.
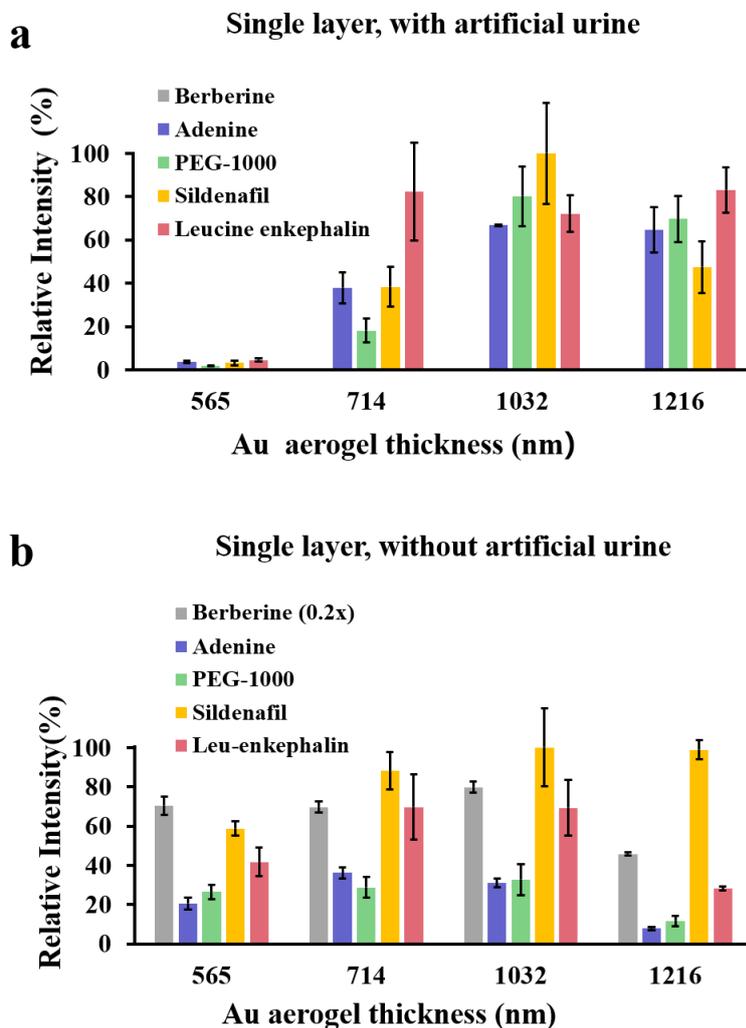
Figure S28. Effect of thickness of Au single-layered aerogel on the detection sensitivity (**a**) with and (**b**) without the interference of artificial urine. *The metabolite standards were spiked in artificial urine (in 1:1 ratio) to investigate the interference of artificial urine.*
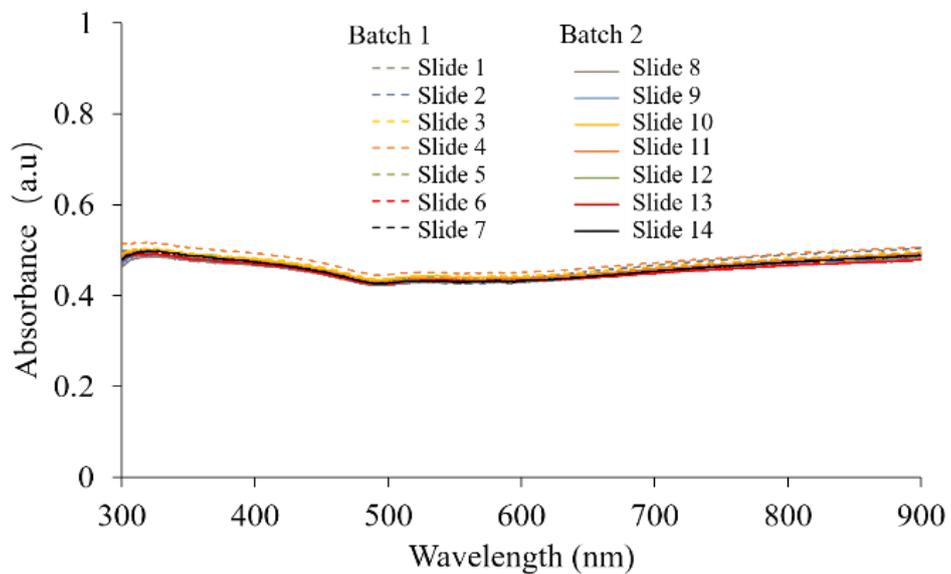
Figure S29. UV-visible spectra of 14 slides of Au/SiO$_2$ double-layered aerogel prepared in two batches. The absorbance at 355 nm was in the range of 0.485 ± 0.009 a.u..

Serum:
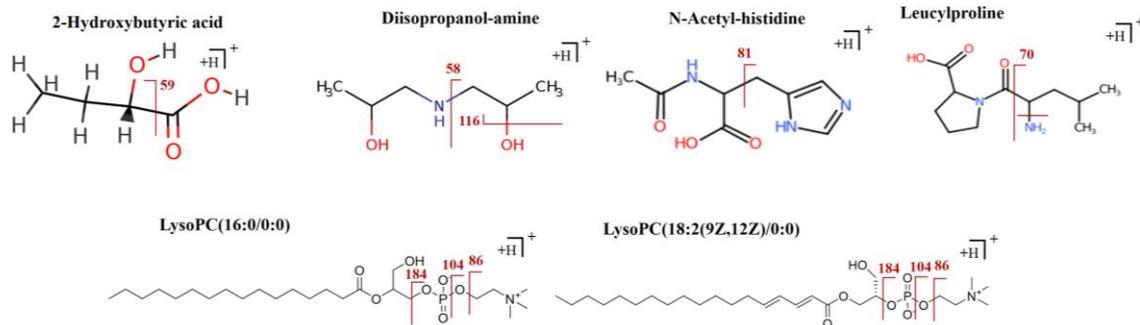
Urine:
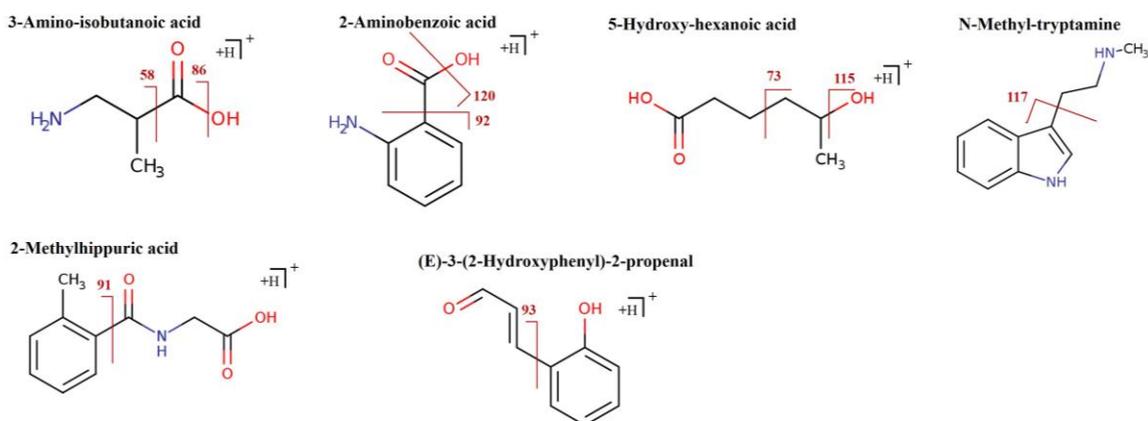
Figure S30. The manually assigned key MS/MS fragment ions with reference to the structure of the biomarker.

III. Supporting Tables

Table S1. The average pore size, specific surface area and total pore volume of aerogel materials.

Aerogel materials	Average pore size (nm)	Specific surface area (m²/g)	Total pore volume (cm³/g)
Au aerogel	45.5	5.46	0.025
SiO_2 aerogel	19.5	584.49	2.208

Table S2. The ligands utilized for the functionalization of Au aerogel.

Name of Ligands	Abbreviation	Chemical Formula	Conc. (M)	Solvent
1-Butanethiol (Macklin, 97%)	But	$C_4H_{10}S$	0.677	Toluene
3-Methyl-1-butanethiol (Macklin, 97%)	3-Met-1-but	$C_5H_{12}S$	0.677	Toluene
Benzyl mercaptan (Aladdin, 98%)	BM	C_7H_8S	0.677	Toluene
2-Ethyl-1-hexanethiol (Energy,98.5%)	2-Eth-1-hex	$C_8H_{18}S$	0.677	Toluene
4-Methoxyphenyl)methanethiol (Aladdin, 98%)	4-MPM	$C_8H_{10}OS$	0.677	Toluene
2-Naphthalenethiol (Macklin, 99%)	2-Nap	$C_{10}H_8S$	0.677	Toluene
1-Decanethiol (Aladdin, 96%)	Dec	$C_{10}H_{22}S$	0.677	Toluene
1-Dodecanethiol (Macklin,98%)	Dodec	$C_{12}H_{26}S$	0.677	Toluene
Octadecanethiol (Macklin, 97%)	Octadec	$C_{18}H_{38}S$	0.677	Toluene
2-Amino-4-chlorothiophenol (Aladdin, 96%)	ACT	C_6H_6ClNS	0.677	Toluene
4-Aminothiophenol (Sigma-Aldrich, 97%)	4-ATP	C_6H_7NS	9E-4	50% Ethanol
3-aminopropanethiol (Macklin, 98%)	3-MP	C_3H_9NS	9E-4	0.05 M HCl
Sodium citrate (Macklin, 99%)	citrate	$C_6H_9NaO_7$	9E-4	H_2O
L-Cysteine (Aladdin, 99%)	L-Cys	$C_3H_7NO_2S$	9E-4	H_2O
mPEG550-HS (Xi'an ruixi Bio. Tech. Co., Ltd, 95%)	mPEG550		9E-4	H_2O
mPEG750-HS (Xi'an ruixi Bio. Tech. Co., Ltd, 95%)	mPEG750		9E-4	H_2O
mPEG2000-HS (Xi'an ruixi Bio. Tech. Co., Ltd, 95%)	mPEG2000		9E-4	H_2O
mPEG5000-HS (Xi'an ruixi Bio. Tech. Co., Ltd, 95%)	mPEG5000		9E-4	H_2O
mPEG10000-HS (Xi'an ruixi Bio. Tech. Co., Ltd, 95%)	mPEG10000		9E-4	H_2O

Table S3. The composition of artificial urine [**10**] utilized in the study.

Composition of artificial urine	Final Conc. (mM)
Lactic acid (Macklin, 20% H_2O)	1.1
Sodium citrate (Macklin, 99%)	2.0
Sodium bicarbonate (Aladdin, AR＞99.8%)	25.0
Urea (Macklin, 99%)	170.0
Calcium chloride (Macklin, 99.9% metal basis)	2.5
Sodium chloride (Aladdin, AR＞99.5%)	100.0
Magnesium sulfate (Aladdin anhydrous, ≥99.5%)	2.0
Sodium sulfate anhydrous (Aladdin, AR 99%)	10.0
Potassium dihydrogen phosphate (Aladdin, AR＞99%)	7.0
Potassium phosphate dibasic (Macklin, >99.0%)	7.0
Ammonium chloride (Macklin, GR,99.8%)	25.0

Table S4. Demographic information and clinical feature of lung cancer patients and healthy controls.

Characteristics		Early stage (Phase I, II) (n=27)		Late stage (Phase III, IV) (n=40)		Health (n=35)	
Male	Female	14	13	26	14	18	17
Age (mean±SD, years)		58.2±8.1		61.8±6.9		56.8±9.5	
BMI (mean±SD, years)		23.5±3.7		22.4±3.9		23.3±2.9	
Smoker		9		13		8	
Subtype	Adenocarcinoma	22		34		/	
	Squamous cell carcinoma	5		6		/	

Table S5. Metabolites discovered and identified in serum or urine samples as biomarkers to differentiate the lung cancer patients and healthy controls. The fragment patterns of the potential biomarkers were compared with the MS/MS of the reference standards or the online database (HMDB) for the identification.[a]

Serum Metabolites	HMDB ID	Monoisotopic Molecular Weight (Da)	Adduct Ions	Detected m/z	Data Source	Collision Energy (eV)	Fragment Ions: m/z (Relative Intensity)	Biological Significance
Isoleucine	HMDB0000172	131.0946	[M+H]$^+$	132.1019	Measured	30	55.019 (20), 56.050 (57), 57.058 (76), 69.071 (100), 86.061 (76)	The exact isomer of Isoleucine is L-Alloisoleucine. The Isoleucine has been demonstrated as a biomarker in colorectal cancer, while only few researches about L-Alloisoleucine are recorded in HMDB.
					Standard	30	55.021 (12), 56.055 (100), 57.070 (81), 69.070 (80), 86.066 (34)	
			[M+Na]$^+$	154.0839	/	/	/	
			[M+K]$^+$	/	/	/	/	
L-Glutamine	HMDB0000641	146.0691	[M+H]$^+$	147.0764	Measured	30	56.050 (61), 74.024 (1), 84.081 (100), 85.029 (1)	The exact isomer of L-Glutamine is D-Glutamine. The L-Glutamine has been demonstrated as a biomarker in colorectal cancer, leukemia and pancreatic cancer, while only few researches about D-Glutamine are recorded in HMDB.
					Standard	30	56.055 (66), 74.048 (2), 84.054 (100), 85.027 (2)	
			[M+Na]$^+$	169.0584	/	/	/	
			[M+K]$^+$	/	/	/	/	

3-Methyl-2-oxovaleric acid	HMDB0000491	130.0629	[M+H]+	131.0705	Measured	45	52.027 (6), 55.055 (6), 57.054 (82), 57.936 (8), 63.024 (17), 65.039 (100), 67.055 (5), 68.058 (6), 69.054 (11), 77.039 (36), 91.055 (36), 97.008 (8), 98.985 (5), 115.054 (17)	/
					Standard	40	52.003 (19), 55.085 (22), 57.044 (33), 57.984 (100), 63.060 (11), 65.092 (15), 67.035 (18), 68.018 (28), 69.014 (24), 77.042 (10), 91.043 (11), 96.996 (14), 98.960 (25), 115.044 (28)	
			[M+Na]+	153.0524	/	/	/	
			[M+K]+	169.0261	/	/	/	
L-Histidine	HMDB0000177	155.0695	[M+H]+	156.0768	Measured	30	56.050 (36), 81.045 (18), 82.053 (37), 83.061 (100), 93.045 (55), 110.072 (49)	Histidine can be a biomarker common to many types of cancer like lung cancer, breast cancer and so on.[20] Determined by UHPLC-Q-Orbitrap-HRMS analysis.
					Standard	30	56.088 (50), 81.032 (48), 82.034 (46), 83.019 (100), 93.012 (71), 110.034 (28)	
			[M+Na]+	178.0587	/	/	/	
			[M+K]+	/	/	/	/	
L-Carnitine	HMDB0000062	161.1052	[M+H]+	162.1124	Measured	30	57.034 (33), 58.066 (53), 59.074 (28), 60.082 (100), 85.029 (73), 102.092 (54), 103.039 (28)	The high energy expenditure of lung cancer cells may disrupt carnitine homeostasis, which is involved in metabolism by mediating long-chain fatty acid transport in the mitochondrial membrane. [21] Determined by air flow-assisted desorption electrospray ionization-mass spectrometry imaging.
					Standard	30	57.068 (53), 58.099 (100), 59.029 (47), 60.111 (88), 85.020 (87), 102.075 (81), 103.016 (27)	
			[M+Na]+	184.0945	/	/	/	
			[M+K]+	200.0684	/	/	/	

					Measured	30	65.039 (100), 91.055 (47)	
Dopamine	HMDB0000073	153.0789	[M+H]+	154.0849	Standard	30	65.049 (100), 91.023 (57)	/
			[M+Na]+	176.0668	/	/	/	
			[M+K]+	192.0395	/	/	/	
L-Tyrosine	HMDB0000158	181.0738	[M+H]+	182.0812	Measured	30	65.039 (4), 91.055 (100), 95.050 (39), 107.049 (8), 119.049 (21), 123.044 (11), 136.076 (6)	In lung cancer, the tyrosine levels abnormally higher than the health controls because of derangement of protein metabolism. [22] Determined by high resolution-liquid chromatography mass spectrometry.
					Standard	30	65.048 (9), 91.019 (100), 95.050 (21), 107.074 (14), 119.065 (13), 123.075 (4), 136.088 (2)	
			[M+Na]+	204.0631	/	/	/	
			[M+K]+	/	/	/	/	
L-Tryptophan	HMDB0000929	204.0898	[M+H]+	205.0969	Measured	30	91.055 (16), 115.055 (18), 117.058 (9), 118.065 (100), 130.065 (9), 132.081 (16), 143.073 (20), 146.060 (25), 170.060 (5)	In lung cancer, the tryptophan levels change comparing with the health because of derangement of protein metabolism. [22, 23] Determined by high-resolution liquid chromatography mass spectrometry and GC-MS.
					Standard	25	91.026 (14), 115.016 (20), 117.070 (15), 118.088 (100), 130.070 (8), 132. 050 (14), 143.083 (15), 146.096 (33), 170.078 (6)	
			[M+Na]+	227.0790	/	/	/	
			[M+K]+	243.0530	/	/	/	

					Measured/Database	CE	m/z (intensity)	Note
2-Hydroxy-butyric acid	HMDB0000008	104.0473	[M+H]⁺	105.0551	Measured	30	57.050 (15), 59.050 (100), 69.034 (19), 87.000 (19)	The exact isomer of 2-Hydroxybutyric acid is 2-Hydroxybutanoic acid. The 2-Hydroxybutyric acid has been demonstrated as a biomarker in colorectal cancer, while only few researches about 2-Hydroxybutanoic acid are recorded in HMDB.
					Database	40	57.034 (4), 59.050 (100), 69.034 (43), 87.044 (66)	
			[M+Na]⁺	127.0367	/	/	/	
			[M+K]⁺	/	/	/	/	
Diisopropanol-amine	HMDB0251354	133.1103	[M+H]⁺	134.1178	Measured	30	56.050 (5), 58.066 (100), 59.074 (8), 70.066 (23), 116.058 (16)	/
					Database	40	56.050 (14), 58.066 (100), 59.055 (42), 70.066 (8), 116.070 (8)	
			[M+Na]⁺	156.0995	/	/	/	
			[M+K]⁺	/	/	/	/	
N-Acetyl-histidine	HMDB0032055	197.0800	[M+H]⁺	198.0851	Measured	45	56.050 (74), 69.034 (14), 81.070 (27), 83.061 (100), 93.045 (68), 108.081 (17), 110.072 (49)	/
					Database	40	56.049 (13), 69.045 (31), 81.045 (25), 83.060 (100), 93.045 (86), 108.055 (22), 110.071 (96)	
			[M+Na]⁺	220.0667	/	/	/	
			[M+K]⁺	236.0407	/	/	/	

Leucylproline	HMDB0011175	228.1473	[M+H]+	229.1547	Measured	45	68.050 (4), 69.058 (1), 70.066 (100), 98.061 (1)	/
					Database	40	68.049 (10), 69.069 (8), 70.065 (100), 98.061 (18)	
			[M+Na]+	251.1365	/	/	/	
			[M+K]+	/	/	/	/	
LysoPC(16:0/0:0)	HMDB0010382	495.3324	[M+H]+	496.3397	Measured	45	60.082 (8), 86.097 (56), 104.107 (60), 125.000 (24), 184.073 (100)	Lysophosphatidylcholine (LysoPC) is an important intermediate in glycerophospholipid metabolism which is susceptible to interference in cancer [24]. The uptake and utilization of LysoPC(16:0/0:0) and LysoPC(18:2(9Z,12Z)/0:0) are a hallmark in Ras-driven cancer cells [25]. Glycerophospholipid metabolism can produce more matrix to promote cancer cell proliferation and regulate cancer cell migration by regulating cell signaling [24]. Previous studies indicated that LysoPC(16:0/0:0) [26] and Lyso PC(18:2(9Z,12Z)/0:0) are down-regulated metabolites in lung cancer [27]. Determined by UPLC-QTOF/MS analysis.
					Database	40	60.080 (30), 86.100 (92), 104.110 (43), 124.999 (100), 184.073 (35)	
			[M+Na]+	518.3231	/	/	/	
			[M+K]+	534.2957	/	/	/	
LysoPC(18:2(9Z, 12Z)/0:0)	HMDB0010386	519.3324	[M+H]+	520.3407	Measured	45	86.097 (54), 104.107 (77), 125.000 (20), 184.073 (100), 258.110 (1)	
					Database	40	86.096 (16), 104.107 (66), 124.998 (16), 184.073 (100), 258.110 (16)	
			[M+Na]+	542.3219	/	/	/	
			[M+K]+	558.2955	/	/	/	

Urine Metabolites	HMDB ID	Monoisotopic Molecular Weight (Da)	Adduct Ions	Detected m/z	Data Source	Collision Energy (eV)	Fragment Ions: m/z (Relative Intensity)	Biological Significance
Creatinine	HMDB0000562	113.0589	[M+H]+	114.0665	Measured	45	54.035 (19), 55.055 (100), 56.050 (18), 57.071 (2), 58.030 (39), 69.034 (8), 72.045 (4), 77.039 (13), 80.050 (28), 86.061 (4)	Creatinine can be a biomarker for lung cancer diagnosis. [28] Determined by NMR.
					Standard	40	54.056 (11), 55.036 (100), 56.030 (44), 57.039 (57), 58.010 (3), 69.055 (7), 72.086 (12), 77.020 (3), 80.098 (1), 86.061 (12)	
			[M+Na]+	136.0482	/	/	/	
			[M+K]+	152.0221	/	/	/	
1,3-Dihydro-(2H)-indol-2-one	HMDB0061918	133.0527	[M+H]+	134.0601	Measured	45	77.039 (79), 79.055 (61), 89.039 (54), 104.058 (39), 105.045 (100), 106.048 (61), 116.058 (79), 117.058 (64)	The exact isomer of 1,3-Dihydro-(2H)-indol-2-one is Indoxyl. The biospecimen locations of 1,3-Dihydro-(2H)-indol-2-one is blood, feces, saliva and urine, while that of Indoxyl is not specified in HMDB.
					Standard	40	77.011 (100), 79.017 (39), 89.030 (14), 104.024 (15), 105.085 (23), 106.030 (27), 116.030 (27), 117.078 (23)	
			[M+Na]+	156.0421	/	/	/	
			[M+K]+	172.0159	/	/	/	
L-Methionine	HMDB0000696	149.0510	[M+H]+	150.0584	Measured	30	56.050 (62), 61.011 (100), 74.024 (5), 87.027 (2), 133.051 (33)	Lung cancer can be examined by testing for radioactive methionine, as the accumulation of methionine appears to represent an increase in organ protein metabolism and tumor viability. [29] Determined by positron emission tomography scan.
					Standard	25	56.085 (80), 61.028 (100), 74.029 (10), 87.059 (7), 133.037 (1)	
			[M+Na]+	172.0403	/	/	/	
			[M+K]+	188.0143	/	/	/	

Uric acid	HMDB0000289	168.0283	[M+H]⁺	169.0356	Measured	30	87.019 (1), 126.030 (73), 141.041 (100), 152.009 (63)
					Standard	30	87.057 (5), 126.044 (42), 141.070 (100), 151.993 (48)
			[M+Na]⁺	191.0176	/	/	/
			[M+K]⁺	/	/	/	/
L-Histidine	HMDB0000177	155.0694	[M+H]⁺	156.0769	Measured	30	56.050 (38), 81.045 (20), 82.053 (37), 83.061 (100), 93.045 (55), 110.072 (47)
					Standard	30	56.088 (50), 81.032 (48), 82.034 (46), 83.019 (100), 93.012 (71), 110.030 (28)
			[M+Na]⁺	178.0587	/	/	/
			[M+K]⁺	194.0328	/	/	/
1,6,7-Trimethylnaphthalene	HMDB0059701	170.1095	[M+H]⁺	171.1128	Measured	45	77.039 (4), 91.055 (24), 115.054 (100), 128.062 (86), 139.054 (3), 141.070 (84), 145.065 (7), 153.070 (11), 155.085 (16)
					Standard	40	77.067 (8), 91.084 (7), 115.044 (100), 128.074 (31), 139.073 (6), 141.050 (81), 145.025 (6), 153.053 (22), 155.056 (18)
			[M+Na]⁺	193.0945	/	/	/
			[M+K]⁺	209.0687	/	/	/

Notes column (Uric acid): uric acid (UA) affects tumor growth and invasion. Small-cell lung cancer patients with elevated serum levels of UA have a poor prognosis. [30] Determined by clinical biochemical assays.

Notes column (L-Histidine): Histidine can be a biomarker common to many types of cancer like lung cancer, breast cancer and so on. [20] Determined by UHPLC-Q-Orbitrap-HRMS analysis.

Notes column (1,6,7-Trimethylnaphthalene): /

Name	HMDB ID	Mass	Adduct	m/z	Type	CE	Fragments (intensity)	Notes
4-Pyridoxic acid	HMDB0000017	183.0531	[M+H]+	184.0606	Measured	30	65.039 (34), 92.050 (22), 120.045 (19), 148.039 (100), 166.050 (35)	PAr index (the ratio of 4-pyridoxic acid over the sum of pyridoxal and pyridoxal -5'-phosphate) reflects increased vitamin B6 catabolism was associated with an increased risk of lung cancer. [31] Determined by LC-MS/MS.
					Standard	25	65.028 (12), 92.058 (10), 120.085 (10), 148.065 (100), 166.060 (15)	
			[M+Na]+	206.0423	/	/	/	
			[M+K]+	222.0158	/	/	/	
Diethyl decanedioate	HMDB0040429	258.1831	[M+H]+	259.1906	Measured	30	55.055 (49), 67.055 (50), 69.071 (100), 79.055 (48), 81.070 (98), 91.055 (38), 93.070 (91), 95.086 (83), 101.060 (7), 121.101 (50), 129.070 (12), 143.086 (21)	/
					Standard	25	55.077 (24), 67.040 (19), 69.053 (82), 79.019 (51), 81.036 (48), 91.089 (16), 93.012 (100), 95.031 (32), 101.093 (4), 121.117 (86), 129.080 (2), 143.058 (1)	
			[M+Na]+	281.1723	/	/	/	
			[M+K]+	297.1446	/	/	/	
3-Amino-isobutanoic acid	HMDB0003911	103.0633	[M+H]+	104.0709	Measured	30	56.050 (63), 57.058 (28), 58.066 (100), 59.050 (33), 60.009 (5), 86.061 (27)	/
					Database	40	56.050 (94), 57.034 (6), 58.066 (100), 59.050 (6), 60.082 (21), 86.060 (21)	
			[M+Na]+	126.0527	/	/	/	
			[M+K]+	142.0265	/	/	/	

2-Aminobenzoic acid	HMDB0001123	137.0476	[M+H]⁺	138.0550	Measured	45	51.024 (21), 53.039 (10), 65.039 (100), 78.034 (1), 92.058 (17), 94.066 (10), 120.045 (2)	The exact isomers of 2-Aminobenzoic acid are m-Aminobenzoic acid and p-Aminobenzoic acid. The 2-Aminobenzoic acid has been determined with colorectal cancer, while only few researches about m-Aminobenzoic acid and p-Aminobenzoic acid are recorded in HMDB.

Compound	HMDB ID	Mass	Adduct	m/z	Source	CE	Fragments	Note
2-Aminobenzoic acid	HMDB0001123	137.0476	[M+H]⁺	138.0550	Measured	45	51.024 (21), 53.039 (10), 65.039 (100), 78.034 (1), 92.058 (17), 94.066 (10), 120.045 (2)	The exact isomers of 2-Aminobenzoic acid are m-Aminobenzoic acid and p-Aminobenzoic acid. The 2-Aminobenzoic acid has been determined with colorectal cancer, while only few researches about m-Aminobenzoic acid and p-Aminobenzoic acid are recorded in HMDB.
					Database	40	51.023 (53), 53.039 (10), 65.039 (100), 78.034 (10), 92.049 (56), 94.065 (23), 120.044 (19)	
			[M+Na]⁺	160.0370	/	/	/	
			[M+K]⁺	176.0109	/	/	/	
5-Hydroxy-hexanoic acid	HMDB0000525	132.0786	[M+H]⁺	133.0860	Measured	45	53.039 (100), 55.055 (32), 59.050 (8), 67.055 (22), 70.996 (16), 72.999 (10), 115.054 (95)	/
					Database	40	53.039 (81), 55.055 (100), 59.050 (4), 67.055 (19), 71.036 (39), 73.029 (21), 115.076 (19)	
			[M+Na]⁺	155.0679	/	/	/	
			[M+K]⁺	171.0418	/	/	/	
(E)-3-(2-Hydroxyphenyl)-2-propenal	HMDB0031725	148.0524	[M+H]⁺	149.0599	Measured	30	77.016 (50), 93.009 (100), 98.962 (15), 117.057 (5), 121.057 (29)	/
					Database	40	77.039 (64), 93.034 (100), 99.023 (15), 117.034 (53), 121.029 (12)	
			[M+Na]⁺	171.0418	/	/	/	
			[M+K]⁺	/	/	/	/	

					Measured	45	65.039 (30), 77.039 (6), 91.055 (100), 115.054 (56), 117.070 (21), 127.054 (3), 128.062 (25)	
N-Methyl-tryptamine	HMDB0004370	174.1156	[M+H]+	175.1190	Database	40	65.038 (14), 77.038 (27), 91.053 (87), 115.054 (100), 117.057 (73), 127.054 (25), 128.050 (12)	/
			[M+Na]+	197.1009	/	/	/	
			[M+K]+	213.0735	/	/	/	
2-Methylhippuric acid	HMDB0011723	193.0738	[M+H]+	194.0789	Measured	45	65.039 (78), 79.055 (74), 91.055 (100), 132.081 (3)	/
					Database	40	65.038 (16), 79.054 (3), 91.054 (100), 132.081 (12)	
			[M+Na]+	216.0609	/	/	/	
			[M+K]+	232.0348	/	/	/	

a When encountering the situation of exact isomers, the biological relevance of the metabolites to cancer diseases was also adopted for the consideration of metabolite identification.

References

(1) Nah, H.Y.; Kim, Y.; Kim, T.; Lee, K.Y.; Parale, V.G.; Lim, C.H.; Seo, J.Y.; Park, H.H. *J. Sol-Gel Sci. Tech*. **2020**, *96*, 346-359.

(2) Su, Y.; Lai, X.; Guo, K.; Wang, X.; Chen, S.; Liang, K.; Pu, K.; Wang, Y.; Hu, J.; Wei, X.; Chen, Y.; Wang, H.; Lin, W.; Ni, W.; Lin, Y.; Zhu, J.; Ng, K.M. *Anal. Chem*. **2022**, *94*, 16910-16918.

(3) Escorihuela, J.; Bañuls, M.J.; Puchades, R.; Maquieira, Á. *J. Mater. Chem. B*. **2014**, *2*, 8510-8517.

(4) Du, R.; Hu, Y.; Hübner, R.; Joswig, J.O.; Fan, X. Schneider, K.; Eychmüller, A. *Sci. Adv*. **2019**, *5,* eaaw4590.

(5) Johnson, P.B.; Christy, R.W. *Phys. rev. B* **1972**, *6,* 4370.

(6) Plimpton, S. *J. Comput. Phys.* **1995**, *117,* 1-19.

(7) Pedretti, A.; Villa, L.; Vistoli, G. *J. Mol. Graph. Model.* **2002**, *21*, 47-49.

(8) Lee, B.J.; Baskes, M.I. *Phys. Revi. B* **2000**, *62*, 8564.

(9) Stukowski, A. *Model Simul Mater SC*. **2010**, *18*, 015012.

(10) Ghaderinezhad, F.; Ceylan Koydemir, H.; Tseng, D.; Karinca, D.; Liang, K.; Ozcan, A.; Tasoglu, S. *Sci. Rep*. **2020**, *10*, 13620.

(11) Lepoittevin, M.; Blancart-Remaury, Q.; Kerforne, T.; Pellerin, Luc.; Hauet, T.; Thuillier, R. *Cell Mol. Biol. Lett.* **2023**, *28*, 43.

(12) Barri, T.; Dragsted, L. O. *Anal. Chim. Acta* **2013**, *768*, 118-128.

(13) Chadha, V.; Garg, U.; Alon, U.S. *Pediatr. Nephrol*. **2001**, *16*, 374–382.

(14) Meister, I.; Zhang, P.; Sinha, A.; Sköld, C.M.; Wheelock, Å.M.; Izumi, T.; Chaleckis, R,; Wheelock, C.E. *Anal. Chem.* **2021**, *93*, 5248-5258.

(15) Gibb, S.; Strimmer, K. *Bioinformatics*. **2012**, *28*, 2270-2271.

(16) Wu, Y.; Li, L. *J. Chromatogr. A*. **2016**, *1430*, 80−95.

(17) Pang, Z.; Zhou, G.; Ewald, J.; Chang, L.; Hacariz, O.; Basu, N.; Xia, J. *Nat. Protoc*. **2022**, *17*, 1735-1761.

(18) Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.A. *PLoS Comput. Biol*. **2017**, *13*, e1005752.

(19) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. *J. Mach. Learn Res*. **2011**, *12*, 2825-2830.

(20) Klupczynska, A.; Dereziński, P.; Garrett, T.J.; Rubio, V.Y.; Dyszkiewicz, W.; Kasprzyk, M.; Kokot, Z.J. *J. Cancer Res. Clin. Oncol.* **2017**, *143*, 649-659.

(21) Li, T.; He, J.; Mao, X.; Bi, Y.; Luo, Z.; Guo, C.; Tang, F.; Xu, X.; Wang, X.; Wang, M.; Chen, J.; Abliz, Z. *Sci Rep*. **2015**, *5*, 14089.

(22) An, Z.; Chen, Y.; Zhang, R.; Song, Y.; Sun, J.; He, J.; Bai, J.; Dong, L.; Zhan, Q.; Abliz, Z. *J. Proteome. Res*. **2010**, *9*, 4071–4081.

(23) Miyamoto, S.; Taylor, S.L.; Barupal, D.K.; Taguchi, A.; Wohlgemuth, G.; Wikoff, W.R.; Yoneda, K.Y.; Gandara, D.R.; Hanash, S.M.; Kim, K.; Fiehn, O. *Metabolites*. **2015**, *5*, 192-210.

(24) Glunde, K.; Bhujwalla, Z.M.; Ronen, S.M. *Nat. Rev. Cancer.* **2011**, *11*, 835-848.

(25) Kamphorst, J.J.; Cross, J.R.; Fan, J.; de Stanchina, E.; Mathew, R.; White, E.P.; Thompson, C.B.; Rabinowitz, J.D. *PNAS.* **2013**, *110*, 8882–8887.

(26) Wu, H.; Chen, Y.; Li, Q.; Gao, Y.; Zhang, X.; Tong, J.; Zhang, Z.; Hu, J.; Wang, D.; Zeng, S.; Li, Z. *J. Chromatogr B Analyt Technol Biomed Life Sci.* **2018**, *1102*, 23-33.

(27) Wu, H.; Wang, L.; Zhan, X.; Wang, B.; Wu, J.; Zhou, A.A *Phytomedicine*. **2020**, *76*, 153259.

(28) Rocha, C.M.; Carrola, J.; Barros, A.S.; Gil, A.M.; Goodfellow, B.J.; Carreira, I.M.; Bernardo,

J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I.F. *J. Proteome Res.* **2011**, *10*, 4314–4324.

(29) Kubota, K.; Matsuzawa, T.; Ito, M.; Ito, K.; Fujiwara, T.; Abe, Y.; Yoshioka, S.; Fukuda, H.; Hatazawa, J.; Iwata, R. *J. Nucl. Med.* **1985**, *26*, 37-42.

(30) Horsfall, L. J.; Nazareth, I.; Petersen, I. *Thorax* **2014**, *69*, 978-979.

(31) Zuo, H.; Ueland, P.M.; Midttun, Ø.; Tell, G.S.; Fanidi, A.; Zheng, W.; Shu, X.; Xiang, Y.; Wu, J.; Prentice, R.; Pettinger, M.; Thomson, C.A.; Giles, G.G.; Hodge, A.; Cai, Q.; Blot, W.J.; Johansson, M.; Hultdin, J.; Grankvist, K.; Stevens, V.L.; McCullough, M.L.; Weinstein, S.J.; Albanes, D.; Ziegler, R.G.; Freedman, N.D.; Caporaso, N.E.; Langhammer, A.; Hveem, K.; Næss, M.; Buring, J.E.; Lee, I.; Gaziano, J.M.; Severi, G.; Zhang, X.; Stampfer, M.J.; Han, J.; Zeleniuch-Jacquotte, A.; Marchand, L.L.; Yuan, J.; Wang, R.; Koh, W.; Gao, Y.; Ericson, U.; Visvanathan, K.; Jones, M.R.; Relton, C.; Brennan, P.; Johansson, M.; Ulvik, A. *Ann. Oncol.* **2019**, *30*, 478-485.