Electronic Supplementary Information

"Nano-Filter"- Integrated AIMS with Machine Learning: Direct Exhaled Breath Analysis for Lung Cancer Screening

Weiqing Wang, [a]+ Yue Tang, [a]+ Zhenqiang Zhang, [c] Wenxiao Wu, [b] Yuanzhu Jiang, [d] Wenjun Wang, [b] Junzheng Meng, [d] Zhenzhen Chen, [a] Weifeng Li, [c] Yanmei Yang, *[a] Yuguo Chen, *[b] and Bo Tang*[a,e]

[a] Dr. W. Q. Wang, Prof. Z. Z. Chen, Prof. Y. M. Yang, Prof. B. Tang
College of Chemistry, Chemical Engineering and Materials Science, Key Laboratory of Molecular and Nano Probes, Ministry of Education
Shandong Normal University
Jinan, 250014, P.R. China.
E-mail: tangb@sdnu.edu.cn; yym@sdnu.edu.cn

[b] Dr. Y. Tang, W. X. Wu, W. J. Wang, Prof. Y. G. Chen*
Department of Emergency Medicine, Shandong Provincial Clinical Research Center for Emergency and Critical Care Medicine
Qilu Hospital of Shandong University
Jinan, 250014, P.R. China.
E-mail: chen919085@sdu.edu.cn

[c] Z. Q. Zhang, Prof. W. F. Li
School of Physics
Shandong University
Jinan, 250100, P.R. China.

[d] Dr. Y. Z. Jiang, J. Z. Meng,
Provincial Hospital Affiliated to Shandong First Medical University
Shandong Provincial Hospital
Jinan, 250021, P. R. China.

[e] Prof. B. Tang
Laoshan Laboratory
Qingdao 266237, P. R. China.

[+] These authors contributed equally to this work.

Table of Contents

Experimental Procedures

Materials and Reagents

All chemicals used in this work were of analytical grade and used without further purification. Water used in all experiments was doubly distilled and purified by a Milli-Q system. Ethyl acetate (≥ 96.0%), sodium sulfate anhydrous (≥ 98.0%, Na_2SO_4), sodium chloride (≥ 98.0%, NaCl), ethylene glycol (≥ 99.8%, $C_2H_6O_2$), silver nitrate (≥ 99.8%, $AgNO_3$) and chloroauric acid (≥ 99.8%, HAuCl4) were purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai). p-bromobenzoyl hydrazine (≥ 99.9%, $C_7H_7BrN_2O$) was purchased from Macklin (Shanghai). Selenourea (99.9%, CH_4N_2Se) was purchased from Alfa Aesar (Shanghai). Polyvinylpyrrolidone (Average molecular weight: 58000, $(C_6H_9NO)_n$, PVP), trisodium citrate (99.6%, $Na_3C_6H_5O_7$) and 4-aminothiophenol (≥ 99.0%, C_6H_7NS) were purchased from Aladdin Biochemical Technology Co., Ltd. (Shanghai). Absolute alcohol (≥ 99.5%, $C_4H_8O_2$) was obtained from Yuandong Fine Chemical Co., Ltd. (Shandong Province, China).

Acetonitrile (≥ 99.7%), dichloromethane (≥ 99.9%), n-hexane (≥ 97.0%), isopropyl alcohol (≥ 99.9%), methanol (≥ 99.9%, MeOH), ethanol (≥ 99.9%), ethyl acetate (≥ 99.9%), enanthol (≥ 99.7%) and acetone (≥ 99.9%) used in MS detection were HPLC grade and brought from Shandong Yuwang and Tianxia New Material Co., Ltd. (Shandong Province, China).

Valeraldehyde (98%, $C_5H_{10}O$, Val), glutathione (98%, GSH) and cysteine (≥ 99.9%, Cys) were brought from Energy Chemical Co., Ltd. (Anhui Province, China). Benzaldehyde (99%, C_7H_6O, Ben) was brought from HEOWNS Biochem Technologies LLC (Tianjin, China). 4-hydroxy-2-nonenal (98%, $C_9H_{16}O_2$, 4-HNE) was purchased from Absin Biological Technology Co., Ltd. (Shanghai, China). N-Benzylidenebenzylamine (98%, internal standard, IS) were obtained from Macklin Technology Co., Ltd (Shanghai, China).

The fiber paper was obtained from KIMTECHTM. 1.00 mL, 2.00 mL and 20.0 mL syringes were bought from Conley Medical Devices (Hunan Province, China). Aluminum foil gas sampling bags (hereinafter referred to as sampling bags) were bought from BKMANLAB (Changde, Hunan Province). Confined-cavity (SN01030559850-00), syringe (500 μL) and capillary tube were obtained from Thermo Scientific (San Jose, CA).

Instruments

Q-Exactive Orbitrap (QE, Thermo Scientific, San Jose, CA) in positive ion mode was used to record signals. MS/MS ion scans were generated through collision-induced dissociation (CID). The voltage of the capillary and tube lens was set to + 30 V. The temperature of the capillary was 320°C, and the sheath gas pressure was set to zero. MS/MS ion scans were generated through collision-induced dissociation (CID). Normalized collision energy (NCE) of 35 - 70 eV (arbitrary manufacturer's unit on the QE was applied to achieve sufficient fragmentation of the samples. QE was operated in a selected ion monitoring (SIM) mode. CID MS/MS mode with the acquisition time of 1 min. Absorption and fluorescence spectra were recorded using a UV-2600 visible spectrophotometer (Shimadzu, Japan). X-ray photoelectron spectroscopy (XPS) spectra were obtained using a Phobios 100 electron analyzer (SPECS GmbH) equipped with 5 channeltrons using an unmonochro-mated Mg Ka X-ray source (1253.6 eV). All pH measurements were performed with a FE28 pH Meter (METTLER TOLEDO, Shanghai, China) with a combined glasscalomel electrode. The scanning electron microscopy (SEM) images were recorded by JEOL JSM-6700F SEM. Transmission electron microscopy (TEM) images were recorded using a JEM-2100 microscope (JEOL, Japan). ^{1}H NMR and ^{13}C NMR spectra were recorded by Bruker Advance 400 MHz spectrometers (Bruker, Germany).

Synthesis of *p*-selenophenylhydrazide (HSe probe)

The synthesis process of *p*-selenophenylhydrazide (HSe probe) was improved based on the previous steps of synthesizing 4-aminobenzeneseleno[1]: 1.20 mmol of *p*-bromobenzoyl hydrazine was added to selenourea solution (1.32 mmol in 3.00 mL absolute ethanol) and refluxed for 2 h. The product HSe probe was collected by filtration, washed with water (50.0 mL) and used directly with not further purification (The yield was determined to be 70.12%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.86 (*s*, 1H), 7.80 - 7.72 (*m*, 2H), 7.72 - 7.62 (*m*, 2H), 4.52 (*s*, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 164.90, 132.42, 131.37, 129.08, 124.79. HRMS (ESI): calculated for $C_7H_8N_2OSe$ ($[M+H]^+$) 216.9802, found 216.9804.

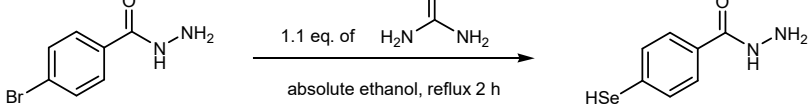


Fig. S1 Synthesis process of HSe probe.

Synthesis of Ag NPs

The alcohol reduction method is used to synthesize Ag NPs with slight modification[2], 0.500 g PVP was added to 40.0 mL ethylene glycol at 47°C. After PVP was totally dissolved, 10.0 mg $AgNO_3$ was added and meanwhile the reaction flask was covered with aluminum foil. After $AgNO_3$ was dissolved completely, the solution was heated to 130°C and stirred for 1 h. The reaction mixture was stirred and cooled down to room temperature. The obtained Ag NPs was washed by acetone for three times and stored in MeOH (40.0 mL, 1.47×10^{-3} mM).

Synthesis of Au NPs

Au NPs were synthesized according to previous report[3]. In detail, 94.0 mL aqueous solution of $HAuCl_4$ (2.00 mM) was stirred and heated until vigorous boiling, then 19.0 mL trisodium citrate solution (38.8 mM) was added. The mixture was kept boiling until the color of the solution changed from yellow to wine red. This reaction was terminated when the suspension cooled down to room temperature. The synthesized AuNPs was stored in aqueous solution (6.00×10^{-5} mM) and stored at 4°C, and concentrated to 1.47×10^{-3} mM before use.

Synthesis of HSe-Ag NPs

200 μL MeOH solution containing Ag NPs (1.47×10^{-3} mM) was stored in a brown glass bottle. Then 0.210 mg HSe probe in 1.00 mL MeOH was added and stirred for 1 h at room temperature. Then, the mixture was centrifuged at 15000 rpm for 15 min at 4°C. The precipitate was collected and washed by MeOH for three times to remove the excess HSe Probe.

Preparation of HS-Ag NPs, HSe-Au NPs and HS-Au NPs nanomaterials

For HS-Ag NPs, 200 μL MeOH solution containing Ag NPs (1.47×10^{-3} mM) was stored in a brown glass bottle. Then 0.130 mg 4-aminothiophenol in 1.00 mL MeOH was added and stirred for 1 h at room temperature. The subsequent processing steps were the same as the preparation of HSe-Ag NPs. The synthesis of HSe-Au NPs is the same as HSe-Ag NPs. The preparation of HS-Au NPs is the same as HS-Ag NPs.

Preparation of gaseous aldehyde

Take Ben as an example, a 100 mL sampling bag was first pre-filled with a specific volume of air (about one-third of the bag volume) using a peristaltic pump. Then, 100 μL of a Ben/MeOH solution (9.4×10^{-2} nM) was injected into the bag using a 1.00 mL disposable syringe. The sampling bag was then stored at 37°C for 3 h to allow equilibration of Ben vapor.

Preparation of nano-fiber paper

A rectangular fiber paper (1.20 × 1.50 cm) was rolled up and placed into the reaction cavity. Then HSe-Ag NPs in MeOH (10.0 μL, 1.00 mM) were added to the fiber paper. The reaction cavity was fixed on a glass rod. During detection, 20.0 mL gas (gaseous Ben or multiple aldehydes or EB samples) from sampling bag was injected into the cavity by a syringe. The operations of extracting 20.0 mL gas can be seen in Video S2.

EB samples collection

Totally 120 EB samples from volunteers were collected, including 60 healthy samples and 60 lung cancer (LCa). Before collection, drink or eat were not allowed for the volunteers for up to 3 h. The volunteers had to take three deep breaths followed by immediate sampling collection using aluminum sampling bags (300 mL). This study was approved by the Scientific Research Ethics Committee of Qilu Hospital of Shandong University (Project Number: KYLL-202312-015)

Mass spectrometry analysis

All the MS analysis were carried out in full mass scan mode or SIM mode with QE, under positive ion mode. The optimal NCE for collision-induced dissociation was 45 - 60 eV. Raw data were exported from *Xcalibur* software (*Thermo Scientific*, San Jose, CA) for further analysis.

Quantification of aldehydes

Gaseous aldehydes (Val, Ben and 4-HNE) with different concentrations (5.00 ppt, 10.0 ppt, 20.0 ppt, 50.0 ppt, 100 ppt, 200 ppt) were prepared in sampling bags. All the gaseous aldehydes were injected into the confined-cavity, followed by supplying with spay solvent MeOH and isopropanol alcohol (V_{MeOH}: $V_{Isopropanol\ alcohol}$=3:1) which contained IS (the concentration is 100 ppt). The average MS peak area ratios of aldehyde derivatization products to IS were obtained from 6 rounds of tests for each concentration.

Recovery and interference experiments

Standard gaseous Ben with known concentrations (8.00 ppt, 15.0 ppt, or 25.0 ppt) were prepared in sampling bags, respectively. The Ben concentrations were determined by "nano-filter" AIMS with six parallel measurements. The recovery (%) was calculated by the formula below: [3]

$$\text{Recovery (\%)} = (C_{tested} - C_{blank})/\ C_{spiked} \times 100\%$$

Here, C_{tested} is the concentration of Ben tested by "nano-filter" AIMS, C_{blank} is the concentration of Ben in sampling bag before injection with standard Ben solution. C_{spiked} is the final concentration after injection.

The interference experiments were performed as follows: firstly, standard gaseous Ben (the concentration was 100 ppt) was prepared in a sampling bag. Then sulfhydryl small molecule metabolites (SMMs), including Cys (1.00 ppm), GSH (1.00 ppm), acetone (1.00 ppm), were added separately. The mixtures were detected by "nano-filter" AIMS.

High-fidelity signal reconstruction *via* machine learning

To address instrumental drift and background noise in non-targeted metabolomics, we developed an end-to-end computational pipeline that combines dynamic signal processing with ML (Fig. S13). During data preprocessing, rather than relying on generic denoising approaches, we applied a resolution-locked dynamic peak picking algorithm (R = 70000) designed around the physical properties of the QE. Early scans affected by the solvent front were excluded using a temporal cutoff, removing non–steady-state spectra acquired at retention times shorter than 0.5 min. This step reduced ion suppression and improved signal stability. On the basis of the filtered scans, we implemented a Global Multiplicative Calibration strategy to correct non-linear drift along the mass axis. An IS (*m/z* 196.11) was used as a spectral reference to derive a scan-specific scaling factor, which was then applied to adjust m/z values for each retained scan. Compared with conventional linear offset correction, this approach provides finer correction at the individual-scan level. Mass calibration was followed by IS-guided quality control, in which spectra exhibiting internal standard intensities below 1000 were excluded. Together, these procedures eliminated approximately 58.5% of low-quality spectra and yielded a curated set of 3447 high-fidelity scan frames for downstream feature extraction, improving measurement consistency and reproducibility from the outset.

To reduce feature redundancy caused by isotopic envelopes in high-resolution mass spectrometry, we implemented a greedy blockout strategy. The algorithm first selects primary peaks with high signal-to-noise ratios and then searches for associated isotopic clusters (M–2 to M+1) within a 0.03 Da mass tolerance. Once an isotopic pattern is identified, the corresponding spectral region is masked to prevent repeated feature extraction from the same metabolite. Feature intensities are subsequently normalized to the internal standard to account for systematic variation, including differences in injection volume. Following feature extraction, we applied a frequency-based aggregation strategy to merge multi-frame data into a single representative profile for each sample. An occupancy filter (>5%) was then applied during this aggregation process, retaining only those features that appeared in at least 5% of the scans within a sample to effectively eliminate transient noise and sporadic artifacts. These steps effectively compress spectra containing tens of thousands of raw signals into a compact, high-confidence feature matrix. Finally, using this refined dataset, we applied a Random Forest-based feature importance screening strategy to select the 500 most informative metabolic features, which were then used to train a robust diagnostic classifier based on the XGBoost[4] architecture.

Results and Discussion

1. Characterization of nanomaterials

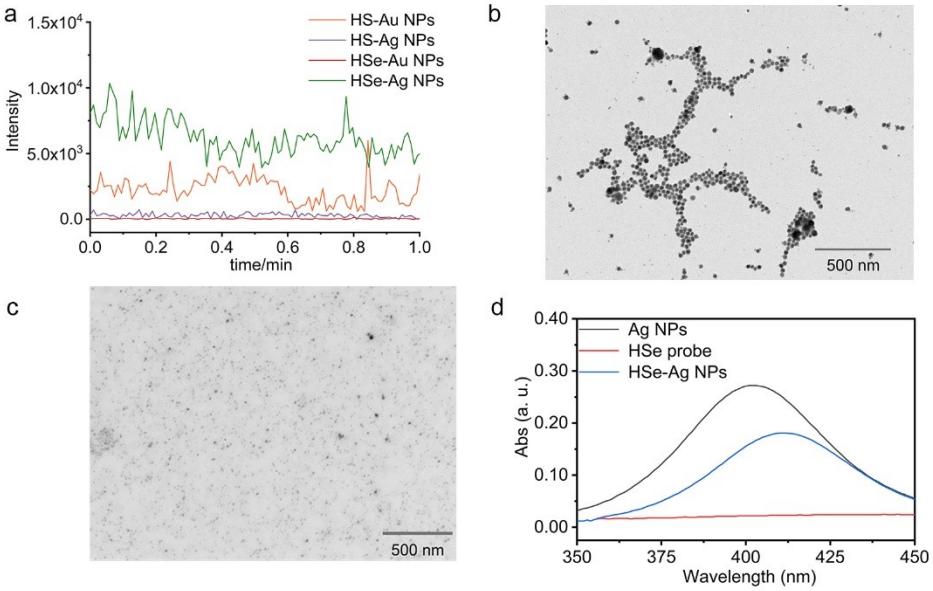


Fig. S2. (a) TIC of HSe-Ag NPs, HSe-Au NPs, HS-Ag NPs and HS-Au NPs at 3200 V. TEM images of (b) Ag NPs and (c) HSe-Ag NPs. (d) UV spectra for Ag NPs, HSe probe and HSe-Ag NPs.

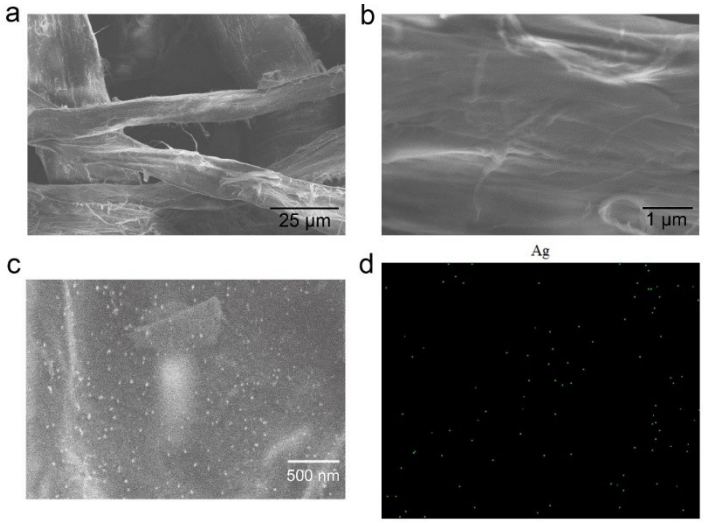


Fig. S3. SEM images of (a) fiber paper, (b) enlargement of the fiber paper, (c) HSe-Ag NPs on the fiber paper and (d) the distribution of Ag elements.

2. Setup of the “nano-filter” AIMS device

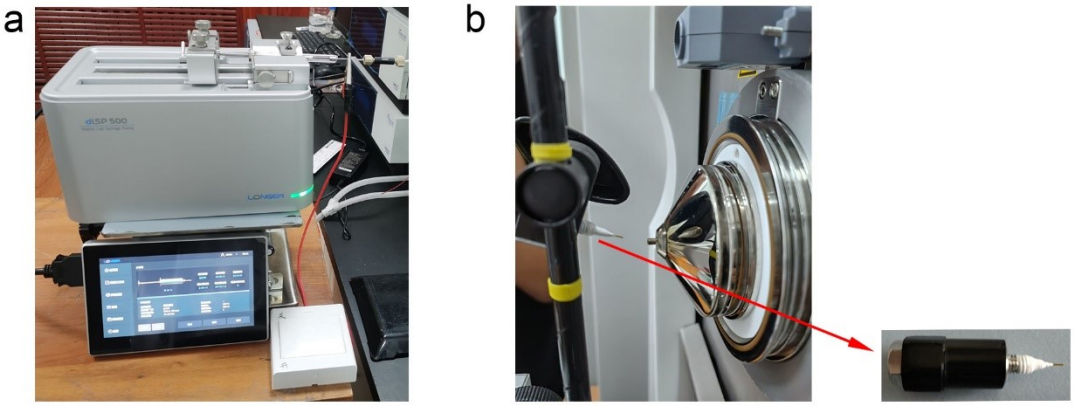


Fig. S4. Photos of (a) part of the AIMS platform (a high-voltage switch, an injection pump and an injector connecting confined-cavity part) and (b) the confined-cavity and mass spectrometry inlet.

3. Optimization of detecting conditions

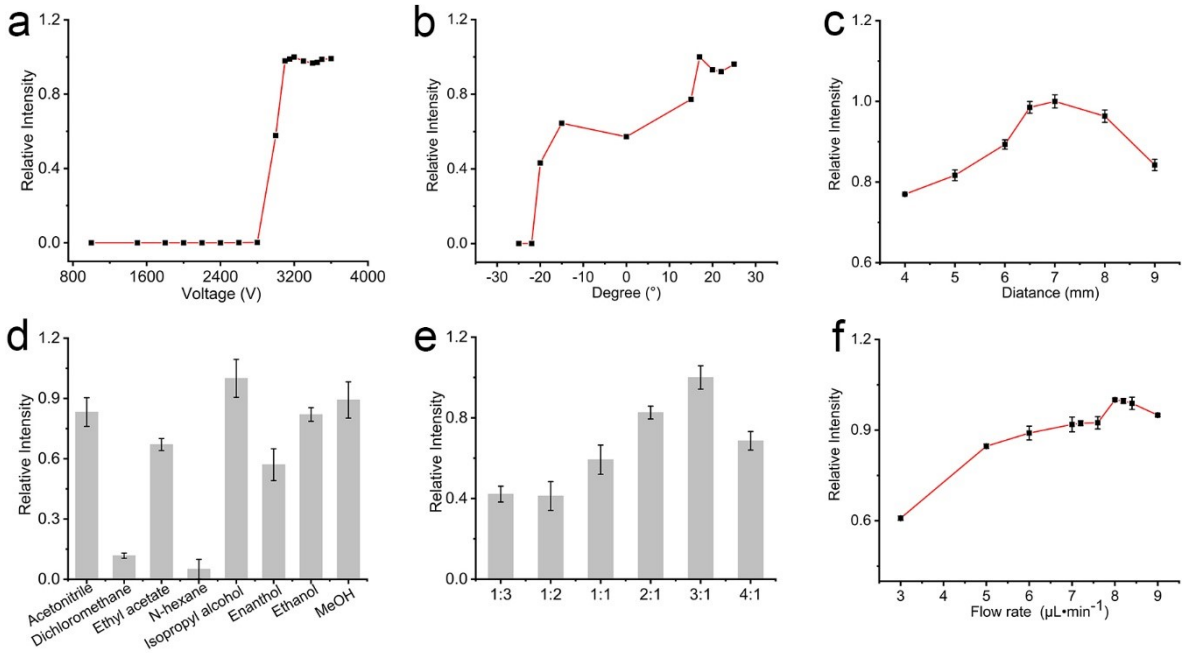


Fig. S5. Optimization of experimental conditions. The MS relative intensity with respect to (a) the voltage, (b) the degree between the capillary in the confined-cavity and MS inlet, (c) the distance between the capillary tip and MS-inlet, (d) different spray solvent, (e) the ratio of methanol to isopropyl alcohol and (f) the flow rate of the spray solvent.

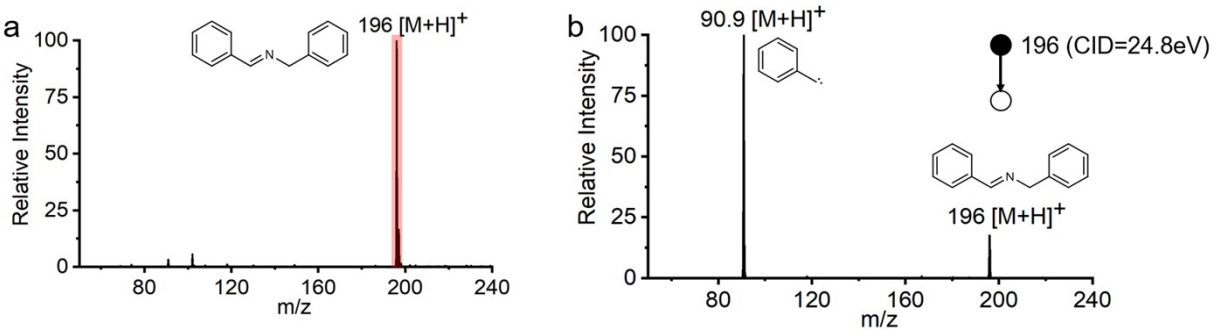


Fig. S6. (a) Mass spectrum of IS; (b) CID MS/MS mass spectrum of IS.

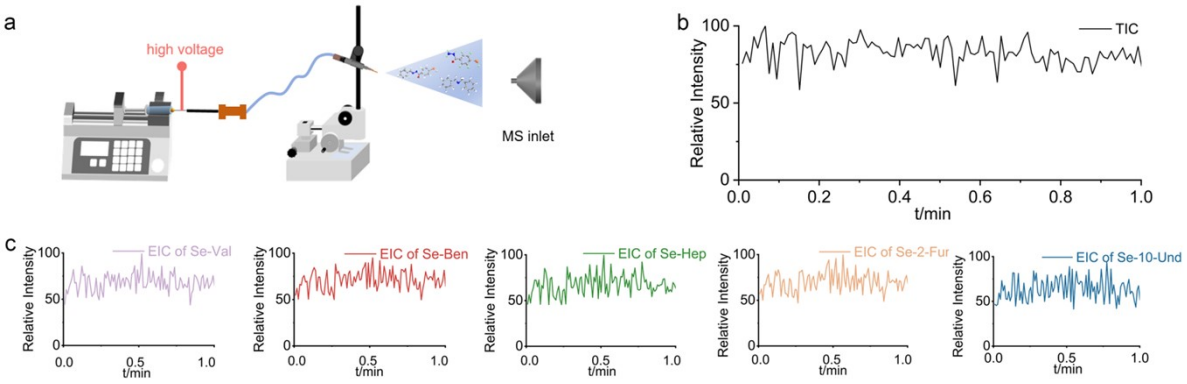


Fig. S7. (a) The operation process of "nano-filter" based AIMS. (b) TIC of full MS and (c) EIC of five Se-aldehydes during 1.0 minute detection.

4. Detection of real EB samples

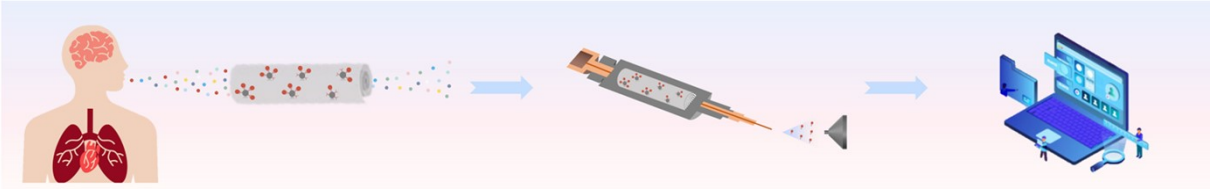


Fig. S8 Schematic diagram of using "nano-filter" AIMS to detect real EB samples.

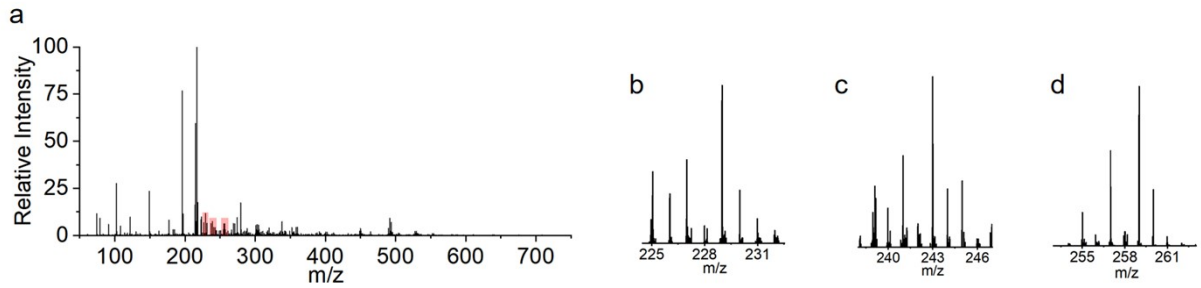


Fig. S9 (a) Full MS spectrum of one EB sample from healthy volunteer. Enlargement of (b) Se-Methanal, (c) Se-Ethanal and (d) Se-Propanal, respectively.

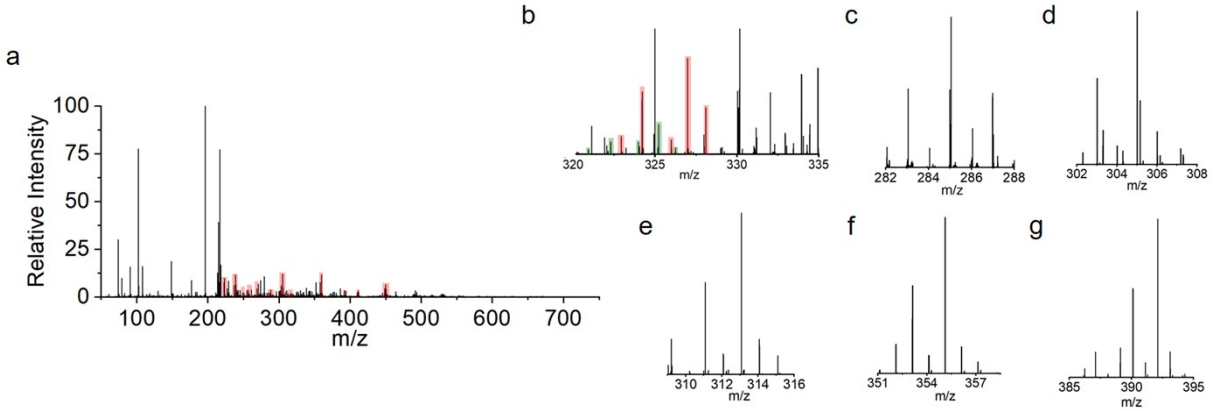


Fig. S10 (a) Full MS spectra of one EB sample from LCa volunteer. Enlargement of (b) Se-Oct (red) and Se-Octe (green), (c) Se-Val, (d) Se-Ben, (e) Se-Hep, (f) Se-4-HNE and (g) Se-Pent.

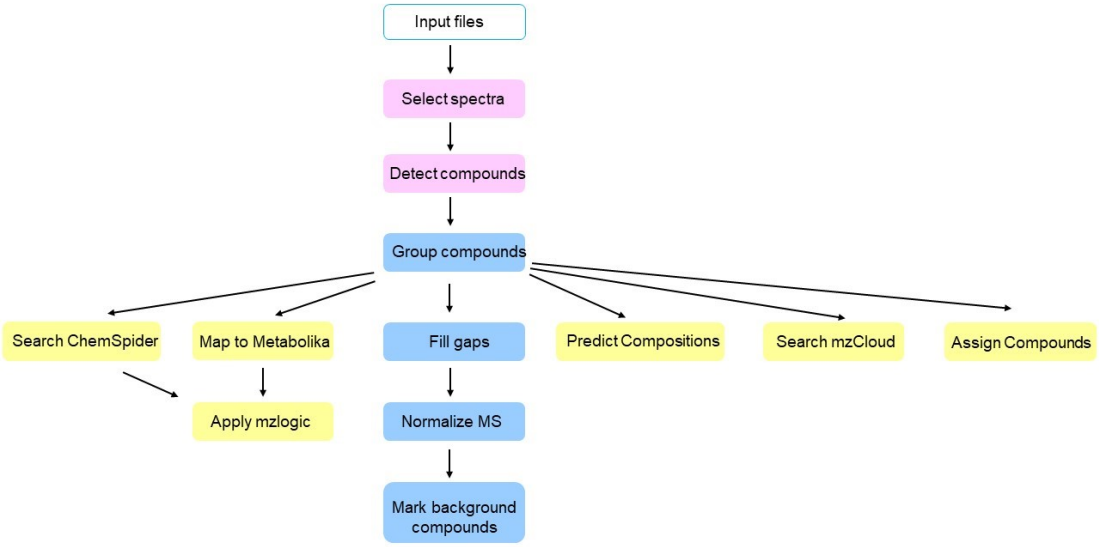
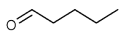
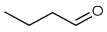
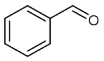
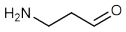
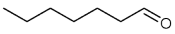
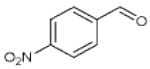
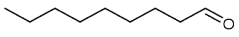
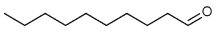
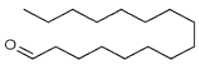
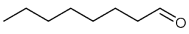
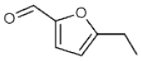
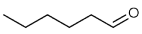
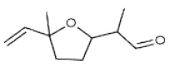
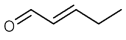
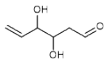
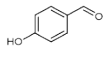
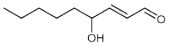
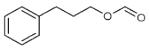
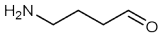
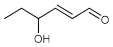
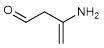
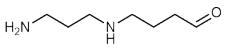
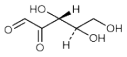
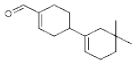
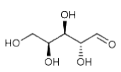
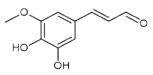
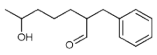
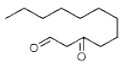
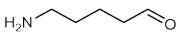
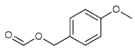
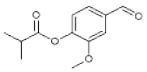
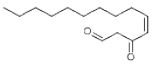
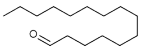
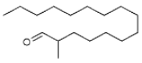
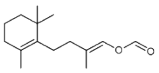
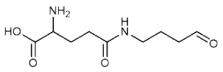
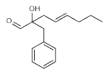
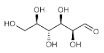
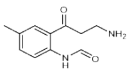
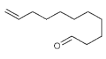
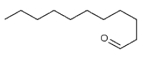
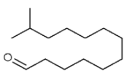
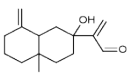


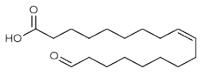
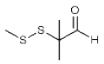
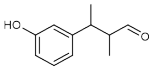
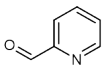
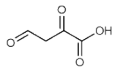
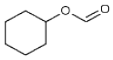
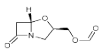
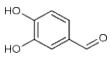
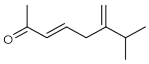
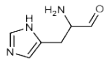
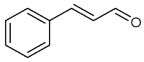
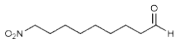
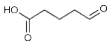
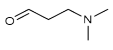
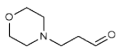
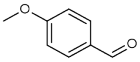
Fig. S11 The workflow of Compound Discoverer software.

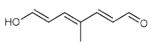
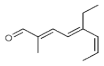
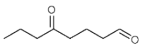
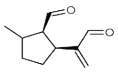
Table S1. The abbreviation and ID of 88 aldehydes

Number (ID)	Name	Abbreviation	Structure	Percentage occupied of Healthy (%)	Percentage occupied of LCa (%)
1	formaldehyde	Methanal		100	100
2	acetaldehyde	Ethanal		100	100
3	propionic aldehyde	Propanal		100	100
4	valeraldehyde	Val		98.3	86.7
5	butyraldehyde	Buty		93.3	86.7
6	benzaldehyde	Ben		91.7	100
7	3-aminopropanal	3-Npro		83.3	53.3
8	heptaldehyde	Hep		83.3	100
9	4-nitrobenzaldehyde	p-nitroben		81.7	63.3
10	nonanaldehyde	Nona		75.0	60.0
11	decanaldehyde	Dec		76.7	66.7
12	palmitaldehyde	Pal		63.3	66.7
13	octanaldehyde	Oct		60.0	85.0
14	5-ethylfuran-2-carbaldehyde	5-ethyl		55.0	61.7
15	hexanal	Hex		50.0	93.3
16	2-(5-methyl-5-vinyltetrahydrofuran-2-yl)propanal	2-Fur-propan		NA	51.7
17	(E)-pent-2-enal	Pente		NA	50.0
18	3,4-dihydroxyhex-5-enal	3,4-DH-hexe		NA	75.0
19	(E)-but-2-enal	Bute		NA	73.3
20	(Z)-3-(tetrahydrofuran-2-yl)acrylaldehyde	3-Fur-acryl		NA	81.7
21	2-hydroxybenzaldehyde	p-Hben		NA	80.0

22	4-(prop-1-en-2-yl)cyclohex-1-ene-1-carbaldehyde	*p*-cyc-1-en		NA	63.3
23	(E)-3-methyl-4-oxopent-2-enal	3-M-4-pente		NA	58.3
24	2-amino-5-oxopentanoic acid	2-N-5-oxo		NA	55.0
25	2-(hydroxymethyl)-4-oxobutanoic acid	2-H-4-oxo		NA	61.7
26	4-hydroxy-2-nonenal	4-HNE		NA	93.3
27	(Z)-4-hydroxy-3-methylbut-2-enal	4-H-3-Mbut		NA	88.3
28	(Z)-3-(3-hydroxyphenyl)-2-methylacrylaldehyde	3-H-2-M-bacy		NA	83.3
29	3-phenylpropyl formate	3-ph-For		NA	80.0
30	4-aminobutanal	4-Nbut		NA	76.7
31	4-hydroxy-2-hexenal	4-HHE		NA	75.0
32	3-aminobut-3-enal	3-Nbut-3-en		NA	70.0
33	(S)-2-amino-3-oxopropanoic acid	2-N-3-oxo		NA	70.0
34	2-hydroxy-3-oxopropanoic acid	2-H-3-oxo		NA	68.3
35	4-((3-aminopropyl)amino)butanal	4-(3-N)-but		NA	68.3
36	3,4,5-trihydroxy-2-oxopentanal	TriH-2-oxo		NA	68.3
37	3,4,5-trimethylbenzaldehyde	Tri-Mben		NA	68.3
38	2-formamidobenzoic acid	2-Forben		NA	68.3
39	5',5'-dimethyl-[1,1'-bi(cyclohexane)]-1',3-diene-4-carbaldehyde	DM-Den-4-car		NA	68.3
40	(2R,3R,4S)-2,3,4,5-tetrahydroxypentanal	Xylose		NA	61.7
41	(E)-3-(3,4-dihydroxy-5-methoxyphenyl)acrylaldehyde	3-(DH)acry		NA	61.7
42	2-benzyl-6-hydroxyheptanal	2-ben-6-H-hep		NA	61.7

43	3-oxododecanal	3-odece		NA	60.0
44	5-aminopentanal	5-Npen		NA	58.3
45	2-(4-hydroxy-3-methoxyphenyl)acetaldehyde	2-(2-H-3-M)ace		NA	58.3
46	2-acetamido-5-oxopentanoic acid	2-ace-5-oxo		NA	56.7
47	4-formyl-2-methoxyphenyl isobutyrate	2-Mphbut		NA	56.7
48	3-oxotetradecenal	3-oxodece		NA	56.7
49	pentadecanal	Pent		NA	56.7
50	2-methylhexadecanal	2-Mhexdec		NA	56.7
51	5-hydroxy-4-(hydroxymethyl)-6-methylnicotinaldehyde	5-H-6-Mlnico		NA	55.0
52	(E)-2-methyl-4-(2,6,6-trimethylcyclohex-1-en-1-yl)but-1-en-1-yl formate	2-M-4-tri-1-en-1-for		NA	55.0
53	N5-(4-oxobutyl)glutamine	N-(4-oxo)glu		NA	53.3
54	2-benzyl-2-hydroxyoctanal	2-ben-2-Hoct		NA	53.3
55	(R)-2-hydroxypropanal	2-Hpro		NA	51.7
56	3-hydroxy-3-(2-hydroxycyclohexyl)propanal	3-(2-DH)pro		NA	51.7
57	(2S,3R,4R,5R)-2,3,4,5,6-pentahydroxyhexanal	D-gulose		NA	51.7
58	(S)-2-acetamido-6-oxohexanoic acid	S-oxo		NA	51.7
59	N-(2-(3-aminopropanoyl)-4-methylphenyl)formamide	N-N(M)-for		NA	51.7
60	10-undecylenal	10-Und		NA	50.0
61	undecanal	Und		NA	50.0
62	12-methyltridecanal	12-Mtride		NA	50.0
63	2-(2-hydroxy-4a-methyl-8-methylenedecahydronaphthalen-2-yl)acrylaldehyde	2-(2-M)-acry		NA	50.0

64	(Z)-18-oxooctadec-9-enoic acid	Z-18-oxo		NA	50.0
65	2-methyl-2-(methyldisulfaneyl) propanal	2-M-pro		88.3	NA
66	3-(3-hydroxyphenyl)-2-methylbutanal	3-2-Mbuty		83.3	NA
67	picolinaldehyde	Pico		81.7	NA
68	piperidine-1-carbaldehyde	Formylpiperidine		73.3	NA
69	2,4-dioxobutanoic acid	2,4-Dbuty		71.7	NA
70	phthalaldehyde	*o*-Phthal		70.0	NA
71	(E)-oct-2-enal7-aminooctanal	7-N-octe		68.3	NA
72	N-methyl-N-(3-oxopropyl) nitrous amide	N-N-nit		63.3	NA
73	cyclohexyl formate	Cyc-for		63.3	NA
74	((3R,5S)-7-oxo-4-oxa-1-azabicyclo [3.2.0] heptan-3-yl) methyl formate	7-oxo-4-1-Mfor		63.3	NA
75	3,4-dihydroxybenzaldehyde	D-H-ben		61.7	NA
76	(E)-7-methyl-6-methyleneoct-3-en-2-one	6-Moct-3-en		61.7	NA
77	2-amino-3-(1H-imidazol-5-yl) propanal	2-N-5-pro		58.3	NA
78	cinnamaldehyde	Cin		55.0	NA
79	9-nitrononanal	9-nitnona		55.0	NA
80	5-oxopentanoic acid	5-oxo		53.3	NA
81	3-(dimethylamino) propanal	3-nbut		53.3	NA
82	3-morpholinopropanal	3-pro		51.7	NA
83	4-nitrooxepan-2-one	4-pan-2-one		51.7	NA
84	4-methoxybenzaldehyde	*p*-Anisaldehyde		51.7	NA

85	(2E,4E,6E)-7-hydroxy-4-methylhepta-2,4,6-trienal	7-hy-4-M-trienal		50.0	NA
86	((2E,4E,6Z)-5-ethyl-2-methylocta-2,4,6-trienal	5-Et-2-M-trienal		50.0	NA
87	5-oxooctanal	5-tanal		50.0	NA
88	(1R,5S)-2-methyl-5-(3-oxoprop-1-en-2-yl)cyclopentane-1-carbaldehyde	2-M-5-1-carb		50.0	NA

*NA: The occurrence of individual aldehyde is < 50.0%.

5. Exploring LCa potential features by using ML

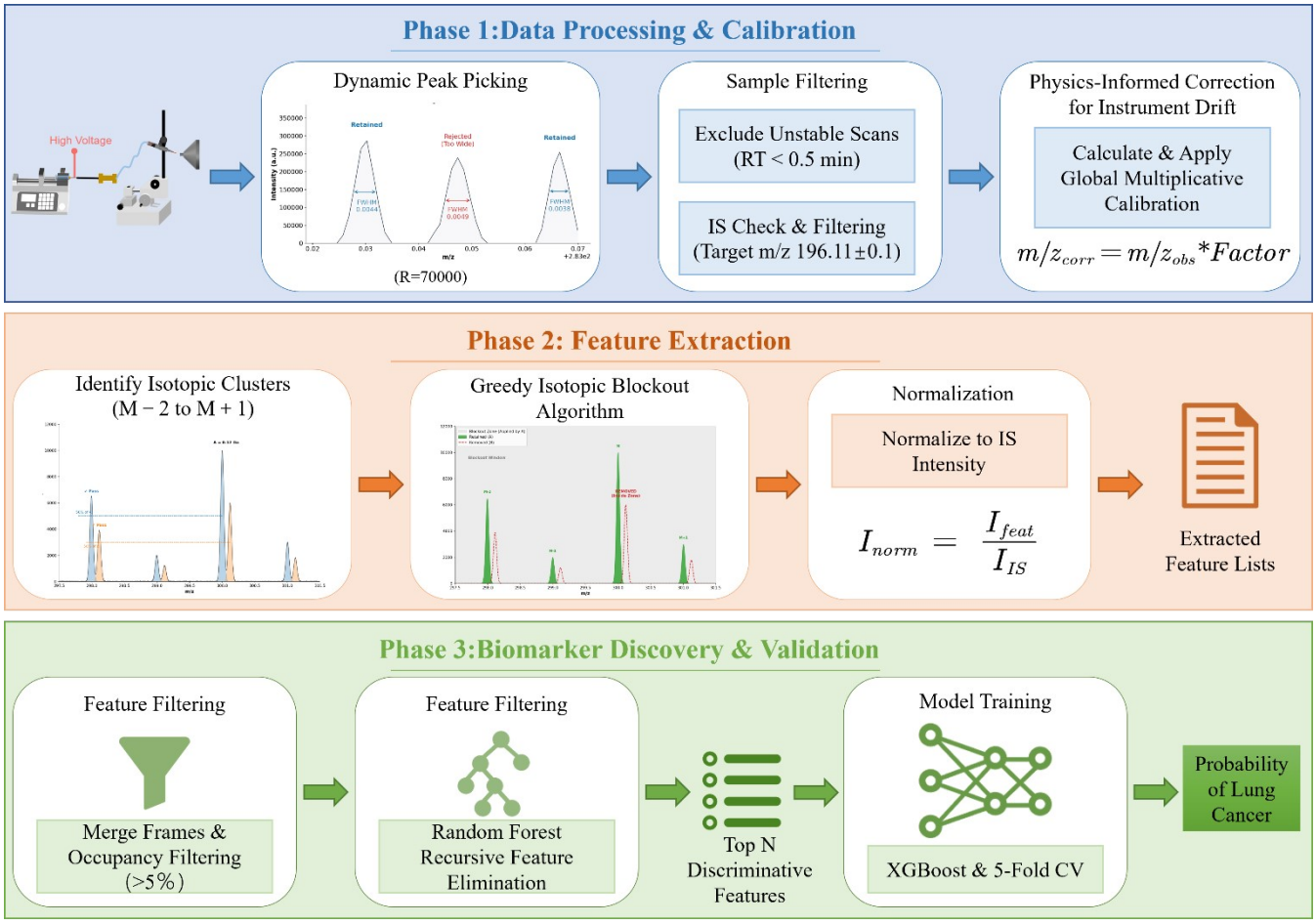


Fig. S12 Overview of the physics-informed metabolomics analysis workflow. The pipeline consists of three main stages designed for high-precision biomarker discovery. Phase I: Data processing & Calibration. Raw mass spectrometry data undergo resolution-locked dynamic peak picking (R = 70,000), followed by global multiplicative calibration. An internal standard (*m/z* 196.11) is used as a spectral reference to correct scan-level mass drift by deriving frame-specific scaling factors. Phase II: Feature extraction. Isotopic clusters (M–2 to M+1) are identified using a greedy isotopic blockout strategy within a 0.03 Da tolerance window. Features are normalized to the internal standard and filtered by occupancy (>5%) to suppress background signals. Phase III: Biomarker discovery & validation. High-dimensional features are screened using a Random Forest model. The selected feature set is then evaluated with ML (XGBoost model) classifier under stratified cross-validation to assess diagnostic performance.

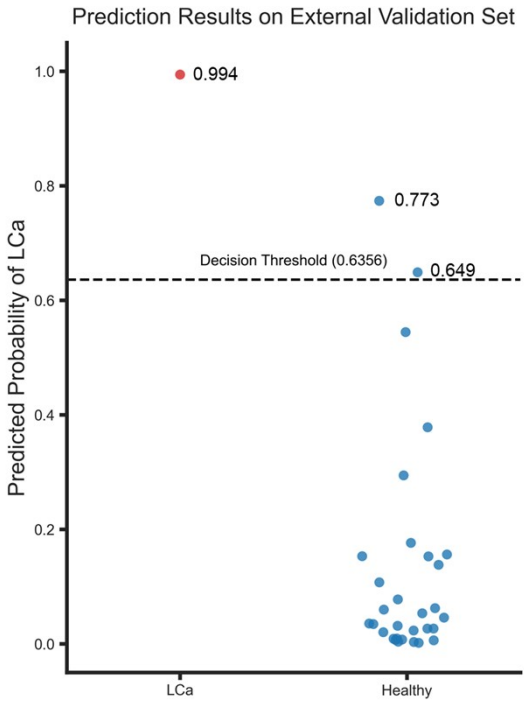


Fig. S13 Prediction probability distribution of the XGBoost model on the external validation set. The y-axis represents the predicted probability of LCa, and the x-axis indicates the true clinical labels (1 LCa and 31 Healthy Controls). The dashed black line denotes the predefined decision threshold (0.636) derived from the 5-fold cross-validation phase. The single true LCa case (red dot) was successfully identified with high confidence (0.994). Among the healthy controls (blue dots), the vast majority were correctly classified with probabilities well below the threshold, with only two cases misclassified as false positives (FP; 0.773 and 0.649) marginally exceeding the threshold.

6. Quantitation of different aldehyde

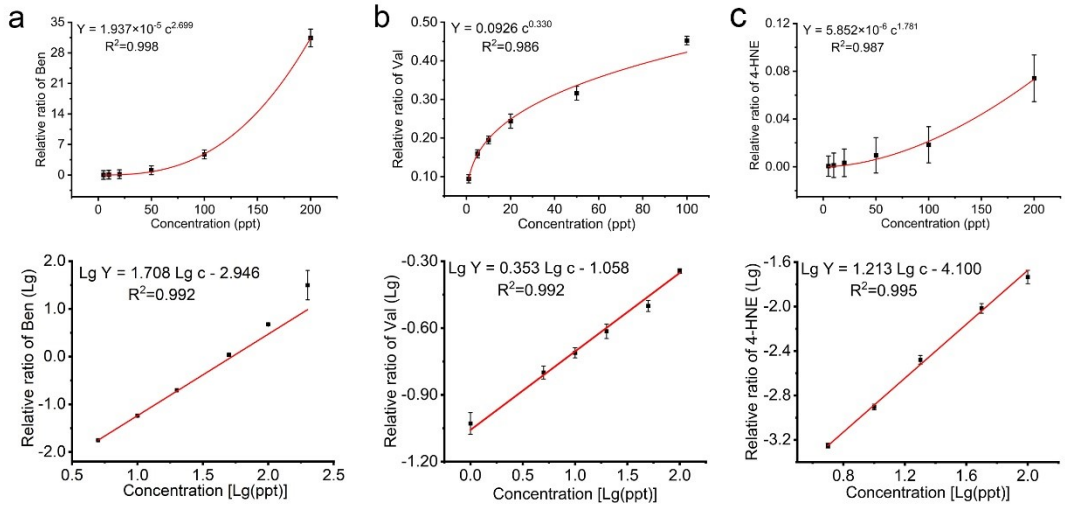


Fig. S14 Calibration curves (up) and curves (down) of different aldehydes: (a) Ben, (b) Val and (c) 4-HNE.

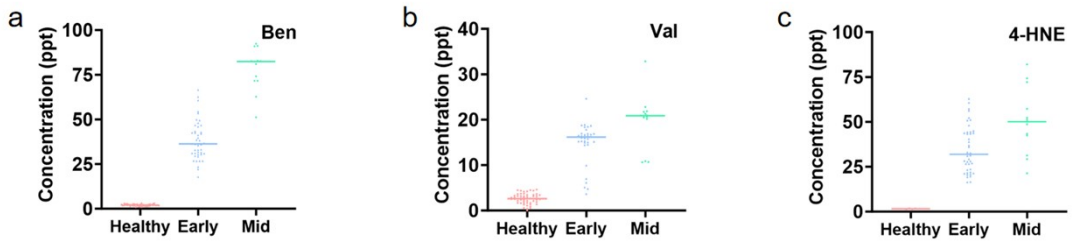


Fig. S15 The respective concentrations of different aldehydes in healthy and LCa volunteers: (a) Ben, (b) Val and (c) 4-HNE.

7. Stability of "nano-filter" AIMS

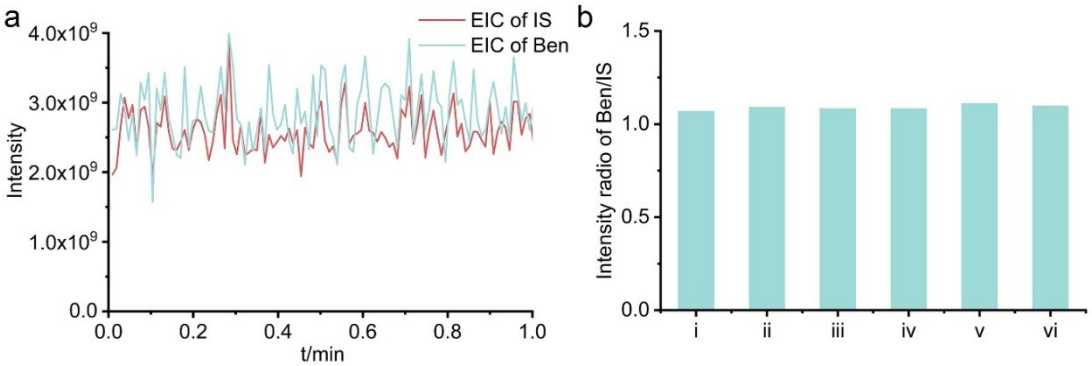


Fig. S16 (a) EIC of the characteristic ion of Ben (blue, *m/z* 305) and IS (red, *m/z* 196). (b) MS peak area ratio of Ben to IS from six parallel tests.

8. Recovery tests of different aldehydes

Recovery tests were conducted by comparing the determined concentrations with the spiked standard aldehyde concentrations. As summarized in Table S2 - Table S4, the average recovery was in the range of 91.3% to 107% (RSD ≤9.33%), indicating the considerably high reliability of this method.

Table S2. Recovery tests of Ben.

Added (μM)	Found (μM)	Recovery	Average recovery	RSD (n=6)
8.00	7.41	0.926	0.917	4.56%
	6.93	0.866		
	7.21	0.901		
	7.59	0.949		
	6.99	0.874		
	7.88	0.985		
15.0	14.9	0.993	0.944	4.91%
	13.7	0.913		
	14.8	0.987		
	14.7	0.980		
	13.8	0.920		
	13.1	0.873		
25.0	25.5	1.02	1.06	5.58%
	27.7	1.11		
	27.8	1.11		
	26.4	1.06		
	27.6	1.10		
	24.4	0.976		

Table S3. Recovery tests of Val.

Added (μM)	Found (μM)	Recovery	Average recovery	RSD (n=6)
8	7.04	0.880	0.913	5.14%
	6.77	0.846		
	7.43	0.929		
	7.32	0.915		
	7.99	0.999		
	7.25	0.906		
15	14.4	0.960	0.921	6.37%
	13.1	0.873		
	13.5	0.900		
	14.3	0.953		
	15.1	1.01		
	12.5	0.833		
25	25.3	1.01	1.065	7.43%
	27.4	1.10		
	29.8	1.19		
	24.5	0.980		
	26.6	1.06		
	26.1	1.04		

Table S4. Recovery tests of 4-HNE.

Added (μM)	Found (μM)	Recovery	Average recovery	RSD (n=6)
8.00	8.55	1.07	1.02	9.33%
	7.99	0.999		
	7.21	0.901		
	9.41	1.18		
	7.78	0.973		
	8.14	1.02		
15.0	14.7	0.980	0.942	4.06%
	14.1	0.940		
	14.6	0.973		
	14.5	0.967		
	13.8	0.920		
	13.1	0.873		
25.0	26.5	1.06	1.03	6.39%
	27.1	1.08		
	24.8	0.992		
	27.4	1.10		
	24.6	0.984		
	23.4	0.936		

9. Interference experiment and stability test

Gaseous Ben mixed with different SMMs (Cys, GSH or acetone) was detected by "nano-filter" AIMS. Compared with the gaseous Ben, these gaseous mixture solutions showed similar MS peak area ratio values (*m/z* 305/196) for Ben.

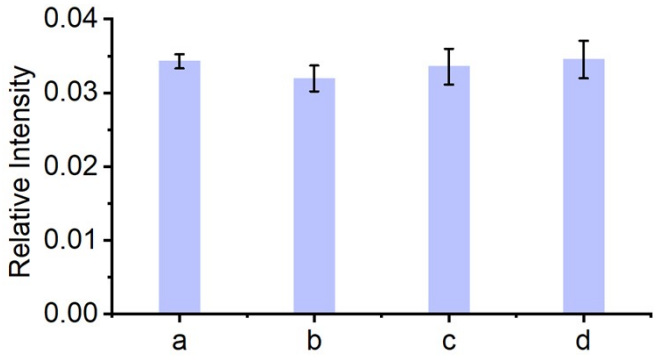


Fig. S17 MS peak area ratio values of Ben (100 ppt) mixed with different SMMs (1 ppm): (a) Ben only; (b) Ben + Cys; (c) Ben + GSH and (d) Ben + acetone.

References

[1] A. C. Ruberte, D. Plano, I. Encio, C. Aydillo, A. K. Sharma, C. Sanmartin, *Eur J Med Chem* **2018**, *157*, 14-27.
[2] Jun Dong, Hairong Zheng, Xuqiang Li, Xiaoqing Yan, Yu sun, Z. Zhang, *Applied Optics* **2011**, *50*, G123-G126.
[3] Y. Yang, W. Wang, H. Liu, L. Tong, X. Mu, Z. Chen, B. Tang, *Angew Chem Int Ed Engl* **2022**, *61*, e202113051.
[4] S. Zhang, Y. Yang, Y. Zhang, C. Chen, Z. Liang, D. Sun, X. Gao, Z. Yang, F. Zhang, *Journal of Materials Science & Technology* **2026**, *260*, 162-173.

Author Contributions

Y.Y. and B.T. designed the experiments. Y.Y. and W.Q.W. performed the experiments and analyzed the data. T.Y., W.X.W., W.J.W. and Y.C. collected EB samples from Qilu hospital. Y.J. and J.M. collected EB samples from Shandong Provincial Hospital. Z. Z. and W. L. conducted the machine learning analysis. T.Y., Z.C. and Y.C. assisted with the experimental instrument support. Y.Y., W.Q.W. and B.T. wrote the manuscript.