Chemical Science

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: W. Wang, Y. Tang, Z. Zhang, W. Wang, J. Meng, W. Wu, Y. Jiang, Z. Chen, W. Li, Y. Yang, Y. Chen and B. Tang, *Chem. Sci.*, 2026, DOI: 10.1039/D6SC02074G.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemical-science

View Article Online
DOI: 10.1039/D6SC02074G

ARTICLE

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

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Weiqing Wang,[a, †] Yue Tang,[b, †] 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]

Exhaled breath (EB) harbors rich molecular information, providing important insights into multiple metabolism processes of the living body. Thus, EB analysis is believed to be a promising diagnostic method for fast and non-invasive disease detection in the future. In this work, we developed a cost-effective "nano-filter" integrated with ambient ionization mass spectrometry (AIMS) for the direct detection of EB aldehyde metabolites. The "nano-filter" features *p*-selenophenylhydrazide-functionalized silver nanoparticles (HSe-Ag NPs) immobilized on fiber paper, selectively capturing EB aldehydes while filtering interferences. Upon application of high voltage to induce cleavage of Ag-Se bonds, the Se-tagged aldehyde derivatives (Se-aldehydes) are liberated for AIMS detection. We demonstrated the high performance of this "nano-filter" AIMS strategy by analysing 152 clinical EB samples, including 91 healthy individuals and 61 lung cancer (LCa, non-small cell lung cancer) patients. Over 88 aldehydes were detected, most reported for the first time. Based on a machine learning (ML) model, the strategy achieved 95.8% accuracy in identifying LCa using these EB aldehydes. We believe that this novel nano-filter AIMS strategy, combined with ML technique, can provide a robust and effective tool for high-throughput LCa screening for clinical diagnosis and biomedical research.

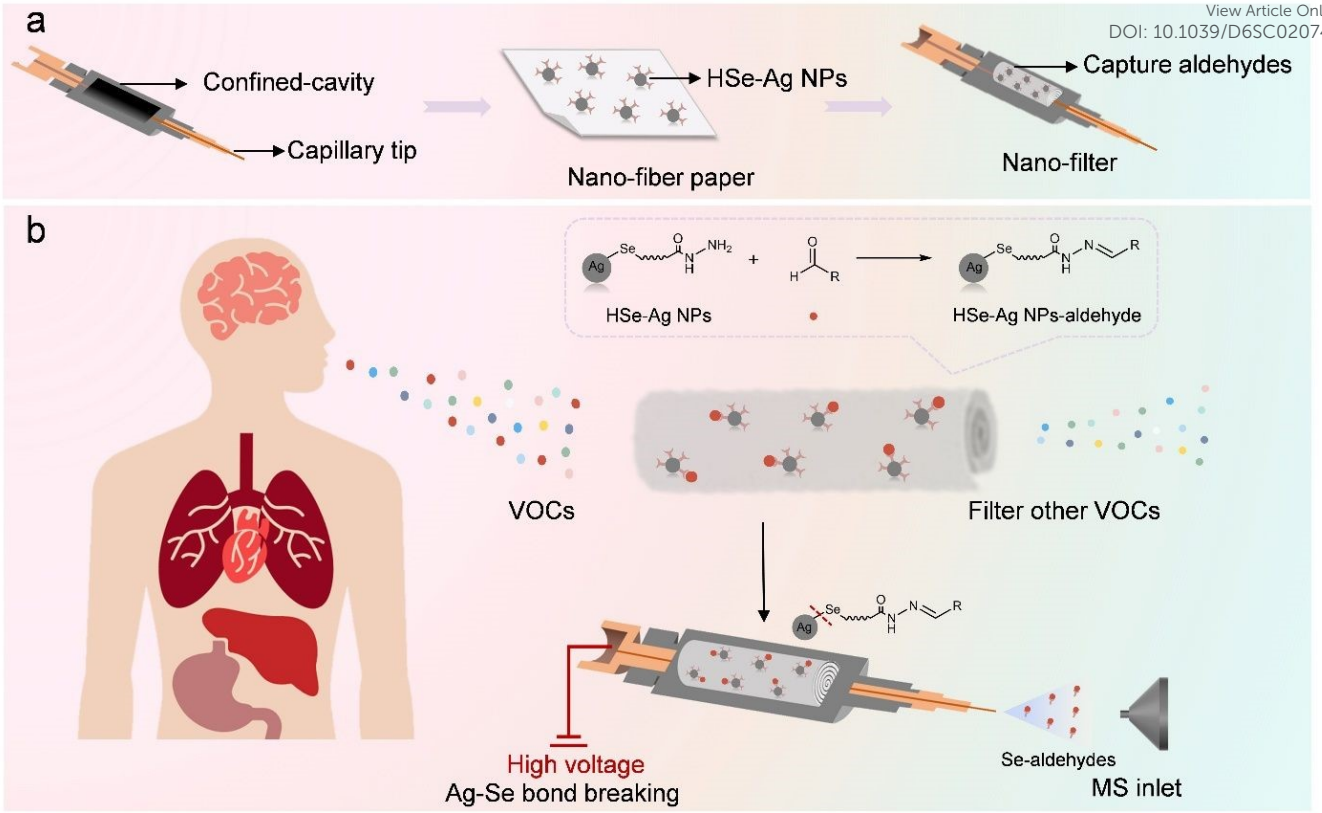
Introduction

Non-invasive but rapid disease diagnostics that have sensitivity and specificity comparable to the current gold standard techniques are crucial to the practical Point-of-Care Testing (POCT) applications.[1-6] Among various promising solutions, EB diagnosis has received great attention in recent years.[7] This is because, compared to traditional liquid biopsy (blood and urine, for instance), EB samples collection is non-invasive and convenient, making it more suitable for practical diagnostic tests.[3, 8] More importantly, EB own rich molecular fingerprint information which can reflect the metabolism processes of our body.[9-11] Among these molecules, aldehyde is an important type of diagnostic object.[12-16] For example, hexanal (Hex), benzaldehyde (Ben), heptaldehyde (Hep), octanaldehyde (Oct), 4-hydroxy-2-hexenal (4-HHE) and 4-hydroxy-2-nonenal (4-HNE) have been reported for identifying LCa.[17, 18] Some studies have demonstrated that several exhaled aldehydes (C2 - C10, straight chain aldehydes) are significantly increased in EB for the patients with lung diseases (such as COVID-19).[15, 19] Furthermore, decanaldehyde (Dec) exhibits significant abnormalities in EB of the esophageal adenocarcinoma patients.[20] In addition, endogenously generated aldehydes have been employed to differentiate healthy individuals from those with disease.[21, 22]

On the basis of the high potentialities of EB in clinical application, several methods have been developed for EB detection in laboratory, such as transistor, nanosensor, Raman scattering sensor, electrochemical biosensors and mass spectrometry (MS).[23-27] Among these approaches, MS might be the predominant technique for gas analysis. Several MS platforms, including gas chromatography MS (GC-MS),[26] proton transfer-reaction MS (PTR-MS) [28] and secondary electrospray ionization-high resolution MS (SESI-HRMS)[29] have been demonstrated to identify different types of VOCs from EB. However, these MS-based methods face two inherent weaknesses that complicating preprocessing steps of samples were required and only one or few analytes were detected, which do not satisfy the clinical demand of quick, high-throughput and easy detection.[8] Considering the massive types of VOCs in EB, the significance of EB analysis for medical purpose are clearly underestimated. To solve these problems, the development of quick and accurate detecting method for EB samples is highly desirable. Moreover, based on such method, capable of constructing fingerprint EB spectra that correspond with an EB-related diseases, reaching the ultimate functions of EB for medically disease diagnosis or screening is extremely important.

[a] 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.
[b] 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.
[c] School of Physics Shandong University, Jinan, 250100, P.R. China.
[d] Provincial Hospital Affiliated to Shandong First Medical University, Shandong Provincial Hospital, Jinan, 250021, P. R. China.
[e] Laoshan Laboratory, Qingdao 266237, P. R. China.
† These authors contributed equally to this work.
Email: yym@sdnu.edu.cn, chen919085@sdu.edu.cn, tangb@sdnu.edu.cn
Supplementary Information available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/D6SC02074G



Scheme 1. Schematic diagram of the "nano-filter" AIMS for EB detection and testing procedure. (a) Design of the "nano-filter" with a confined cavity and capillary tip. (b) Workflow of EB sample processing using "nano-filter" AIMS platform.

In this work, we report the design of a novel "nano-filter" AIMS for accurate detection of multiple aldehyde metabolites in EB. As shown in Scheme 1a, the filter containing a confined-cavity (220 μL) for EB storage, and a capillary at the tip for generating a stable Taylor cone under high voltage (Video S1). Moreover, a nano-fiber paper which is coated with *p*-selenophenylhydrazide functionalized Ag nanoparticles (abbreviated as HSe-Ag NPs) is placed in the filter. After injected into the "nano-filter", exhaled aldehydes in EB will be captured by the hydrazine group on HSe-Ag NPs, while the interfering matrix components in EB were filtered out. Then, followed by high voltage being applied at the "nano-filter", the Ag-Se bond cleaved and the ionized aldehyde derivatives (Se-aldehydes) are detected by MS (Scheme 1b). Without any pre-treatment for EB, accurate, quick, high-throughput and real-time quantitative detection of exhaled aldehydes is realized.

Results and discussion

Preparation of Ag-Se bond based "nano-filter"

The preparation steps of this novel Ag-Se bond based "nano-filter" are illustrated in Fig. 1a. Specifically, to accurately capture aldehydes in EB, a small molecule *p*-selenophenylhydrazide (HSe probe) is designed (Fig. S1) which contains a hydrazide group for reacting with aldehyde group, and a seleno group (-SeH) serves as mass-tag with its characteristic isotope peak distribution.[30] To enhance the reaction efficiency with aldehydes, HSe probe is linked to Ag NPs surface through Ag-Se bond to form HSe-Ag NPs. It is worth to mentioning that, in the preliminary experiments, we tested the fracture ability of four kinds of bonds (Ag-Se, Au-Se, Ag-S, and Au-S), which are formed by Ag NPs and Au NPs linked with -SeH or -SH group, while monitoring the MS signals of the linked small moleculars under different voltage. The results indicated that Ag-Se bond showed the highest performance under 3200 V (Fig. 1b and Fig. S2a). Thus, HSe-Ag NPs are chosen in the subsequent experiments. Transmission Electron Microscope (TEM) results showed that the average sizes of both Ag NPs and HSe-Ag NPs were 15.0 ± 1.7 nm (Fig. 1c, Fig. S2b and S2c). In addition, X-ray photoelectron spectroscopy (XPS) (Fig. 1d) and ultraviolet-visible spectrum (UV-vis) (Fig. S2d) results also confirmed the successful loading of HSe probe onto Ag NPs surface.

The prepared HSe-Ag NPs are coated onto a fiber paper (1.20 × 1.50 cm) and characterized by scanning electron microscope (SEM) (Fig. 1e and Fig. S3). Subsequently, as depicted in Fig. 1a, the formed nano-fiber paper is placed into the cavity of the filter. In detail, a rectangular fiber paper (1.20 × 1.50 cm) is cylindrically folded and placed in the reaction cavity of the filter. 10 μL of HSe-Ag NPs in methanol solution was deposited onto the fiber paper, after which the lid of the filter is tightened. Then the EB sample is slowly injected into the filter via a 20.0 mL disposable medical syringe. After 1 min for EB adsorption and

View Article Online
DOI: 10.1039/D6SC02074G

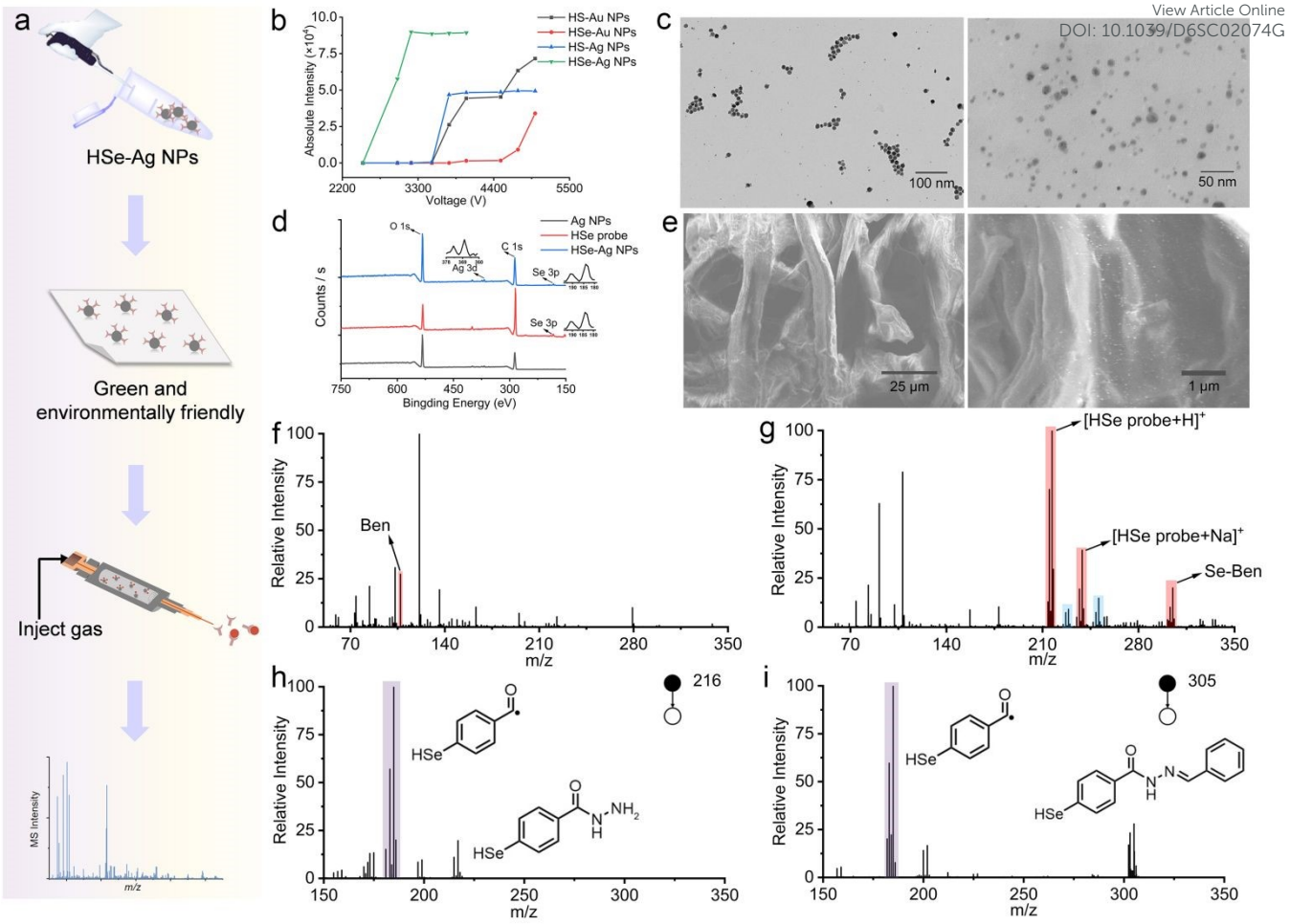


Fig. 1 Characterization of HSe-Ag NPs and verification of the capture gaseous aldehydes performance of HSe-Ag NPs. (a) Schematic diagram of the HSe-Ag NPs used for preparing nano-fiber paper and "nano-filter" AIMS. (b) MS signal intensity of the linked small molecules from the four kinds of bonds under different voltage: Ag-Se, Au-Se, Ag-S, and Au-S. (c) TEM images of Ag NPs (left) and HSe-Ag NPs (right). (d) XPS for Ag NPs, HSe probe and HSe-Ag NPs. (e) SEM images of nano-fiber paper (left) and the enlarged image (right). Mass spectra of (f) pure gaseous Ben w/o nano-fiber paper and (g) pure gaseous Ben w/ nano-fiber paper (blue color: [HSe probe- Methanal + H]$^+$ and [HSe probe- Methanal + Na]$^+$; (formaldehyde: Methanal)). CID MS/MS of (h) HSe probe and (i) Se-Ben (purple colour: CID MS/MS structure of *m/z* 184).

reaction with the HSe-Ag NPs, the spray solvent (methanol/isopropyl alcohol mixture with varying ratios) is introduced into the filter through a 500 μL syringe pump at a flow rate of 3-9 μL·min^{-1}. Meanwhile, a high voltage was simultaneously applied to the needle tip of the syringe for AIMS detection.

To address the performance of this "nano-filter", gaseous Ben is chosen as a prototype model and tested by using the filter w/o and w/ the nano-fiber paper. It is seen that for Ben w/o fiber paper treatment, there is a relatively lower MS signal peak of Ben (*m/z* 107, Fig. 1f). In contrast, with the existence of nano-fiber paper in the filter, a clear MS signal with typical Se isotopic peeks distribution (*m/z* 305, Fig. 1g) for Ben derivative is detected (HSe probe-Ben, abbreviated as Se-Ben). Meanwhile, another peak of *m/z* 216 is also detected, revealing the excess HSe probe ([HSe probe + H]$^+$), which further confirmed by CID MS/MS mode where the same ionic fragment (*m/z* 184) is observed from *m/z* 216 (Fig. 1h) and *m/z* 305 (Fig. 1i), due to the cleavage of C-N bond in CID mode. This result indicated that this "nano-filter" own the ability to capture gaseous aldehydes. Thus, the detection conditions are optimized. As summarized in Fig. S5, the optimal supplied high voltage is 3150 V, the degree between the capillary of the filter and MS inlet is 17° and the distance is 7 mm. To improve ionization efficiency, methanol is added to isopropyl alcohol with volumetric ratio of 3:1, which results in the best MS signal. The optimized flow rate of the spray solution is determined to be 8 μL·min^{-1}.

Simultaneous identification of aldehyde mixture by "nano-filter" AIMS

Before testing real EB samples, the detecting performance of this "nano-filter" AIMS platform was further evaluated by aldehydes mixture sample, based on the optimized conditions. As the representatives, valeraldehyde (Val), Hep, 2-furaldehyde (2-Fur), 10-undecenaldehyde (10-Und) and Ben were mixed in methanol solution and stored in a sampling bag (Fig. 2a). This sampling bag was kept at 37°C for 3 h to form gaseous sample (the final concentration for each aldehyde was 50.0 ppt). N-

View Article Online
DOI: 10.1039/D6SC02074G

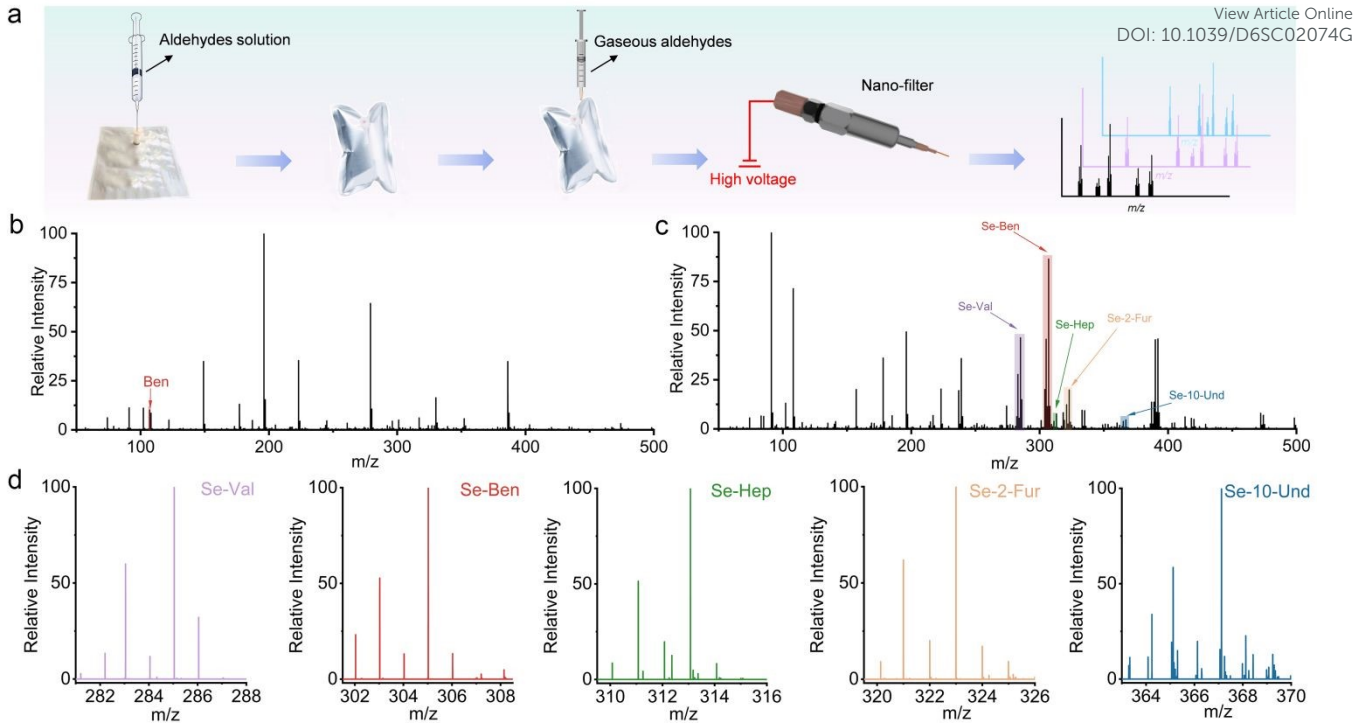


Fig. 2 Use "nano-filter" AIMS to simultaneous detect multiple gaseous aldehydes. (a) The workflow of preparing gaseous aldehyde mixture. (b) Mass spectrum of gaseous aldehyde mixture by using the filter w/o nano-fiber paper. (c) Mass spectrum of gaseous aldehyde mixture by using the filter w/ nano-fiber paper. (d) MS signals of the five Se-aldehydes.

Benzylidenebenzylamine was selected as internal standard (IS) because its MS/MS fracture mode was similar to Se-aldehyde (Fig. S6). The aldehydes mixture was also detected through the "nano-filter" w/o and w/ nano-fiber paper. Without nano-fiber paper treatment, only Ben from the five aldehydes can be detected (*m/z* 107, Fig. 2b). On the contrary, in the presence of nano-fiber paper, all the five aldehydes' derivatives are successfully detected simultaneously (*m/z* 285 for Se-Val, *m/z* 305 for Se-Ben, *m/z* 313 for Se-Hep, *m/z* 323 for Se-2-Fur and *m/z* 367 for Se-10-Und, Fig. 2c and 2d). Further, quantitative analyses of the area under the peaks (A) enable us to calculate the reaction efficiency of Se-aldehydes, which all reach more than 85.0% for these five aldehydes (Reaction efficiency= $A_{Se\text{-}aldehyde}$/ [$A_{Se\text{-}aldehyde}$ + $A_{aldehyde}$]). The total time for simultaneous detection of the five aldehydes is only 5 min. Moreover, the MS signal is considerably stable during the whole detection process, revealing the high performance of this designed "nano-filter" AIMS (Fig. S7).

Screening of the fingerprint aldehydes in LCa EB samples by "nano-filter" AIMS

The ultimate aim for developing this "nano-filter" AIMS strategy is to screen exhaled aldehydes from real EB samples, especially for the samples from LCa. Generally, EB samples from volunteers were collected by using sampling bags. Totally 120 EB samples were collected, including 60 LCa samples (45 early-stage and 15 mid-stage, non-small cell lung cancer) and 60 healthy volunteers. Overall, LCa EB samples exhibited more complex MS signals (Fig. S9) than those from healthy volunteers (Fig. S10). In order to accurately compare the MS data, *Compound Discoverer* (*CD*) software[31] was utilized to screen the specific aldehydes and the corresponding intensity of the selected Se MS signals (the workflow of *CD* software can be found in Fig. S11).

Based on *CD* analysis, totally 64 aldehydes were identified in LCa EB samples and 39 in healthy samples (≥50% occurrence rate), the full name and abbreviation of these aldehydes are summarized in Table S1. Numerous aldehydes are mutual species, including Methanal, Ethanal, Propanal, Ben and Hep and etc. However, a large number of aldehydes are peculiar to LCa, like 4-HNE, 4-HHE and more. Fig. 3a summarizes the 64 aldehydes in LCa samples and their occurrences in healthy people, with the size of the chord link representing the abundance of aldehyde species in EB. The aldehydes are ordered by the molecular weight (Supplementary Table 1). Further, from differential analyses, 47 important aldehydes have been identified which showed significantly higher difference in LCa patients than in healthy individuals (30% higher), or peculiar to LCa, as illustrated in Fig. 3b. For clarity, the full names of the first ten aldehydes are given.

In Fig. 3c, we list the aldehydes that are ordered by their occurrence in female and male LCa samples. Three aldehydes, Methanal, Ethanal, Propanal, were consistently found in all LCa and healthy samples. Furthermore, Ben and Hep were also present in all the LCa samples, and in 91.7% and 83.3% of healthy samples, respectively. Combing the aldehydes found in healthy samples, totally 88 aldehydes were identified with high

View Article Online
DOI: 10.1039/D6SC02074G

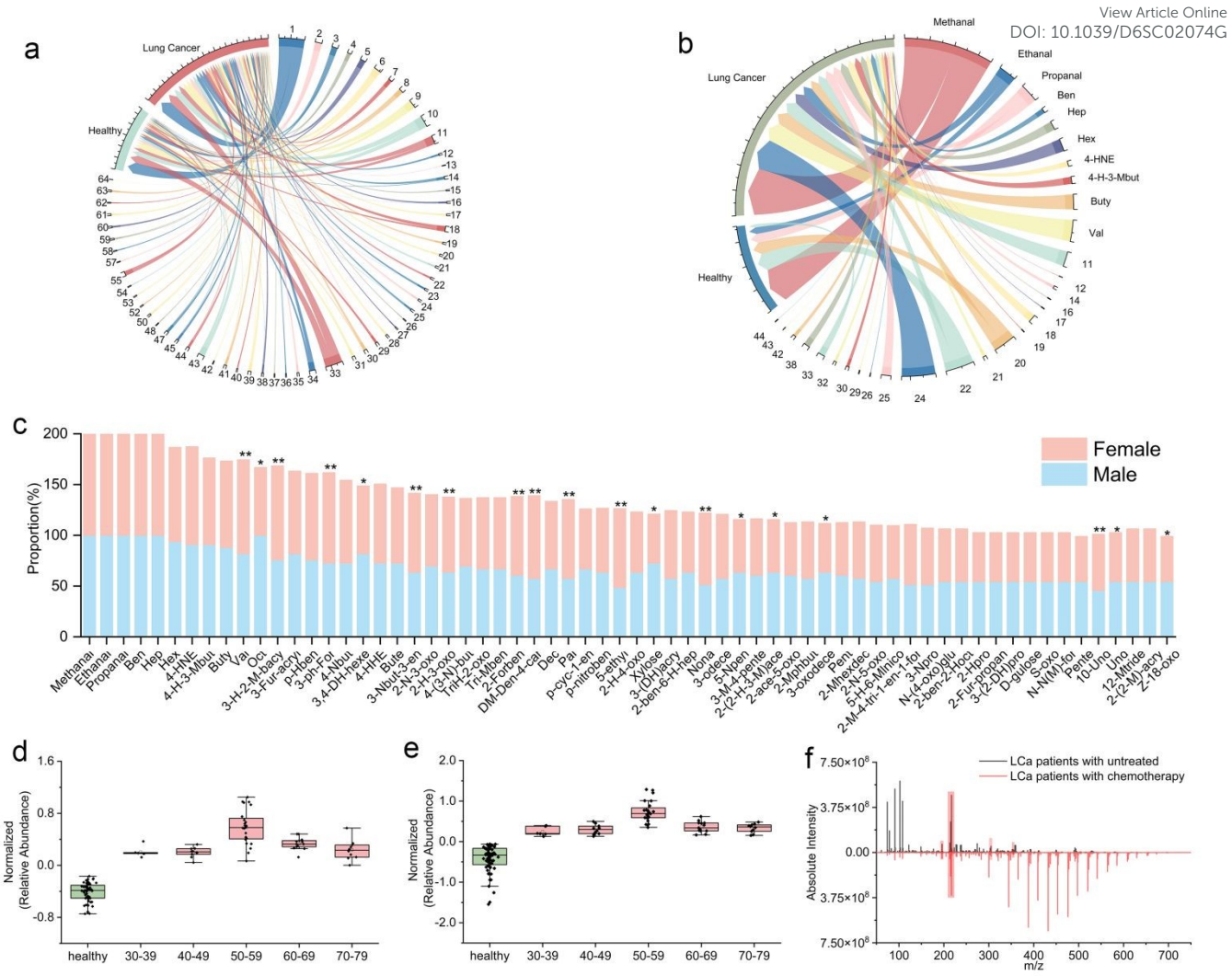


Fig. 3 Multidimensional analysis of exhaled aldehydes in real EB samples. Chord diagram of (a) EB aldehyde species in healthy or LCa and (b) important aldehyde species from differential analysis which are peculiar to LCa, or mutual aldehydes but with large difference (> 30%) in healthy and LCa people respectively. (c) Proportion of certain aldehydes in 60 LCa patients (gender dependence exceeding 10.0%, *: male>female, **: female>male). Age dependence of (d) Ben and (e) Hep. Mass spectra of (f) LCa patients before and after chemotherapy (The red annotations from left to right are IS, HSe probe, Se-Ben and Se-4-HNE).

occurrence. Because there exists a historical discussion about the gender disparities of LCa incidence,[32] we especially compared the levels of fingerprint aldehydes in all samples which are summarized in Fig. 3c. It is found that Val, 3-H-2-M-bacy, 3-ph-For, 3-Nbut-3-en, 2-Forben, DM-Den-4-car, Pal, 5-ethyl and Nona show clearly higher occurrence in female than male LCa patients. In sharp contrast, the occurrence of Oct, 3,4-DH-hexe, Xylose and 5-Npen in male samples is higher than that of female samples. Especially that, Oct is found in all male LCa samples while in female samples the occurrence is only 66.7%. We further noticed that the relative average MS intensity of Oct in male is significantly higher than that of healthy samples, may be due to some lifestyle habits of the male (*e.g.* smoking).[33] As shown in Fig. 3c and 3d, the change trend of Ben and Hep at different age stages is similar, which increase from 30 years old and reach maximum at 50-59 age group. Then a clear decrease happens at 60-79 stage. This further supports our conclusion that these fingerprint aldehydes can serve as important indicators to interpret the age-related metabolite condition of the body.

In addition, one special case of the mid-stage LCa patient has received chemotherapy treatment. We have collected the EB samples from this patient before and after treatment. Our MS analysis indicated that a significant decrease in the MS peak area ratios of Ben to IS (*m/z* 305/196) and 4-HNE to IS (*m/z* 355/196) after chemotherapy. Meanwhile, the MS peak area ratio of Hep to IS (*m/z* 313/196) was still kept high (Fig. 3f). We speculate that different metabolites (including aldehydes) have shown varying sensitivity to chemotherapy drug.[34] However, this is a one-case study and further efforts are surely desired accounting this observation.

View Article Online
DOI: 10.1039/D5SC02074G

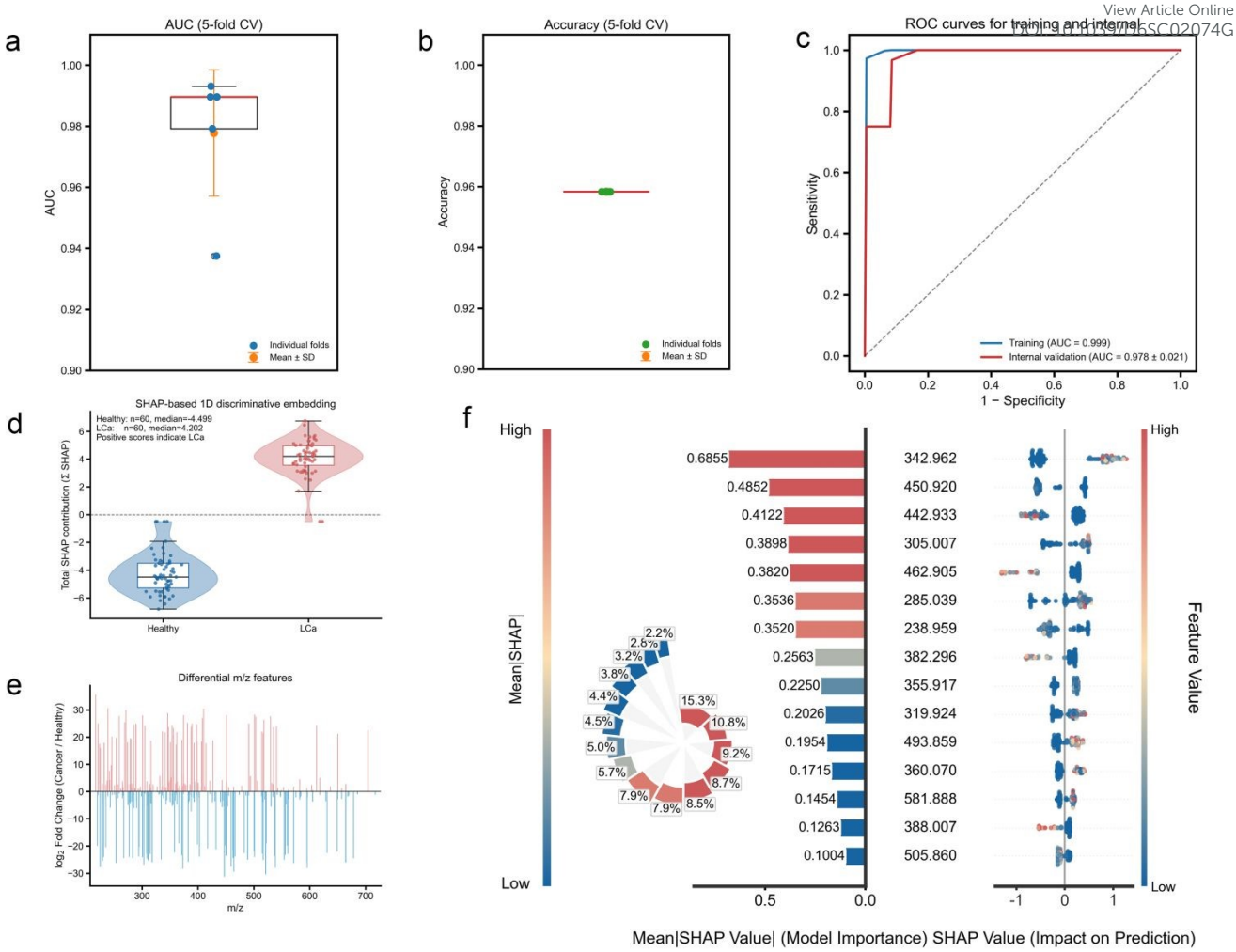


Fig. 4 Performance evaluation and model interpretability of the LCa diagnostic framework. (a) Boxplots of the area under the curve (AUC) obtained from 5-fold stratified cross-validation. The center line indicates the median value. (b) Boxplots of classification accuracy across cross-validation folds. (c) Receiver operating characteristic (ROC) curves for the training set (blue, AUC = 0.999), internal validation set (red, AUC = 0.978). (d) One-dimensional SHAP-based embedding showing the distribution of healthy controls and LCa samples in the learned feature space. (e) Differential *m/z* features exhibiting large abundance changes between groups (>10-fold). (f) SHAP summary plot of the top 15 features contributing to model prediction (Each point represents a sample, with color indicating feature abundance).

Machine learning model for LCa screening

The performance of the ML model[35] was evaluated using on a dataset of 152 medical samples, comprising 61 LCa patients and 91 healthy controls, following the inclusion of 1 additional LCa and 31 additional healthy controls (Fig. S12). The dataset was partitioned into a training set (n=120, with a positive-to-negative sample ratio of 1:1) for model construction and cross-validation, and an independent external test set (n=32, containing 1 LCa and 31 control samples) for final evaluation. The distributions of the Area Under the Curve (AUC) and accuracy from the cross-validation are shown in Fig. 4a and 4b. The comparative analysis of Receiver Operating Characteristic (ROC) curves presented in Fig. 4c indicates that the model performs consistently well across the training (AUC = 0.999), internal validation (AUC = 0.978). As illustrated in Fig. S13, the model successfully identified the single LCa case with very high confidence (0.994). The decision threshold of 0.636 is determined as the median of the optimal thresholds derived from 5-fold cross-validation tasks. For the vast majority of healthy controls, the predicted probabilities are well below the threshold. There exist two false positive cases with relatively high probability of 0.773 and 0.649, respectively, exceeding the threshold. These results visually demonstrate that the model can effectively control the false positive rate and maintain high specificity even in a highly imbalanced, real-world screening scenario.

The SHAP-based one-dimensional embedding (Fig. 4d) shows a separation between healthy controls and LCa samples based on their metabolic profiles. To elucidate the biological basis of the model, we analyzed the abundance changes (Fold Change, LCa/healthy) of all 500 features (Fig. 4e) and identified the top 15 contributing features through a SHAP summary plot (Fig. 5f). Methanal, Ethanal and Propanal were excluded from the features panel because they exhibited no statistically significant

View Article Online
DOI: 10.1039/D6SC02074G

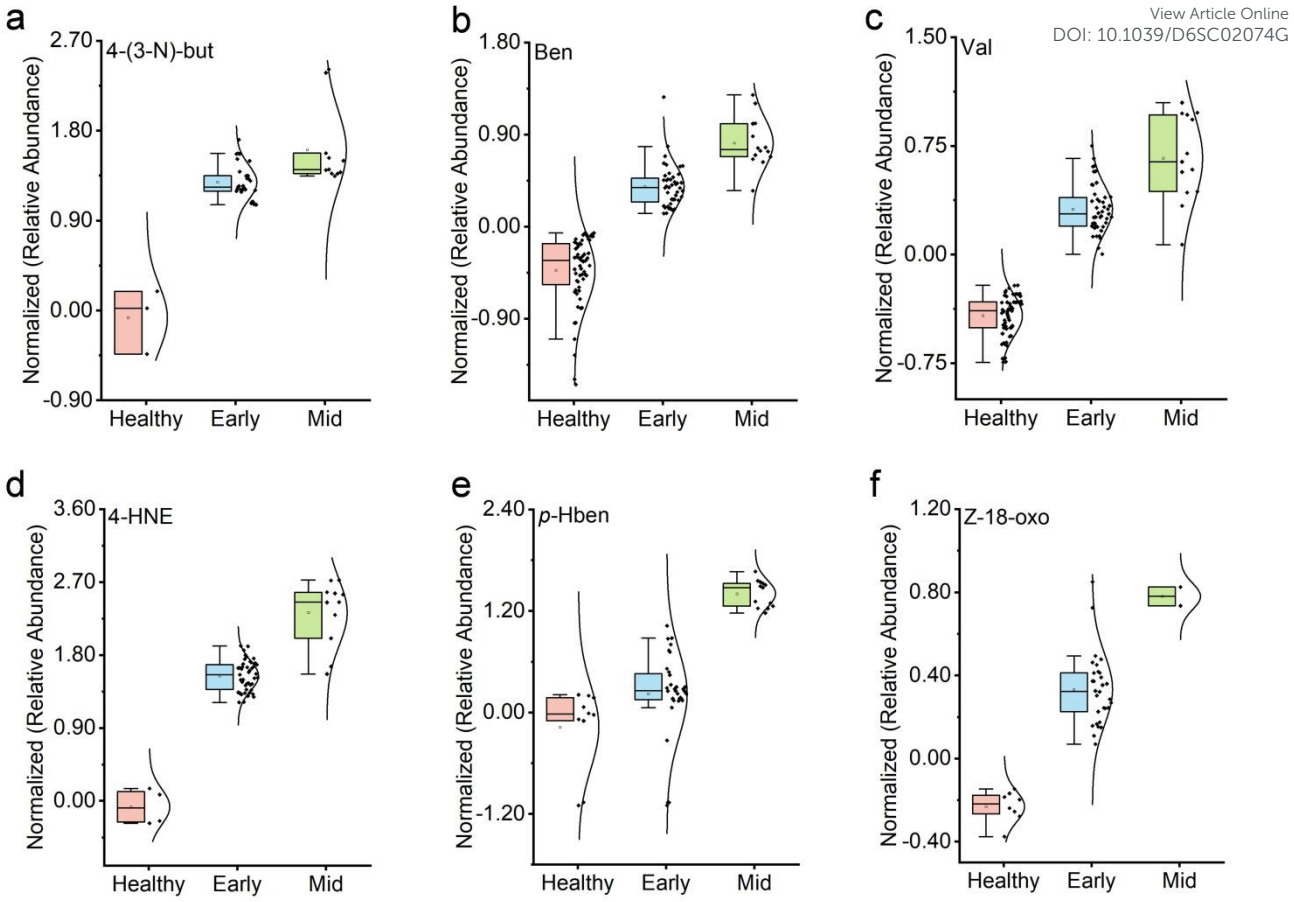


Fig. 5. Relative concentrations of six aldehydes in healthy controls and LCa patients with different disease stages (by *CD* software): (a) 4-(3-N)-but, (b) Ben, (c) Val, (d) 4-HNE, (e) *p*-Hben and (f) Z-18-oxo.

differential expression between LCa patients and healthy controls in the previous "nano-filter" AIMS detection. Thus, potential key features were pointed: for instance, 12-Mtride (*m/z* 450.920), Pal (*m/z* 442.933), Pent (*m/z* 462.905), excess HSe probe (*m/z* 238.959, $[HSe\ probe + Na]^+$), 2-M-5-1-carb (m/z 382.296) and *p*-nitroben (*m/z* 388.007) are notably elevated in healthy group, while 4-(3-N)-but (*m/z* 342.962), Ben (*m/z* 305.007), Val (*m/z* 285.029), 4-HNE (*m/z* 355.917), *p*-Hben (*m/z* 319.924) and Z-18-oxo (*m/z* 493.859) show significantly higher levels in LCa patients. These specics are well in line with the AIMS detections as well as previous experimental reports[10, 11, 36, 37] revealing the high accuracy and good explanation of our ML model.

Quantification of the deterministic aldehydes

To further investigate aldehydes characteristic of LCa, we performed relative quantitative analysis using *CD* software firstly. As shown in Fig. 5, the 6 aldehydes, 4-(3-N)-but, Ben, Val, 4-HNE, *p*-Hben and Z-18-oxo, selected by ML exhibited significantly increased abundances with disease progression. Furthermore, integrated with *CD* software and ML results, three aldehydes, including Ben, Val and 4-HNE, demonstrated the most significant differences across disease stages. Therefore, to reach the ultimate value of the aldehydes in clinically diagnostic applications, logarithm standard curves and calibration curves were constructed (Fig. S14). Quantitative analysis results (Fig. S15) revealed significantly higher concentrations of these aldehydes in LCa samples compared to the healthy group. Specifically, Val, Ben and 4-HNE exhibited concentrations exceeding 4.47 ppt, 17.84 ppt and 9.70 ppt, respectively, in LCa patients. Besides, parallel experiments (Fig. S16), recovery tests (Table S2 - Table S4), and interference experiments (Fig. S17) have also been conducted, which also suggested that the convenient, efficient and non-invasive characteristics of this "nano-filter" AIMS strategy held great promise for practical applications.

Conclusions

Unlike traditional laboratory EB tests, which are time-consuming and typically require complicated sample preparation, our innovative "nano-filter" AIMS platform demonstrates significant advantages in both operational efficiency and cost-effectiveness (~$0.73/sample), while simultaneously integrating economic feasibility with efficient LCa screening compatibility. The key component of this "nano-filter" is composed of a confined-cavity for EB storage and a

capillary at the cavity tip. A rollup nano-fiber paper coated with HSe-Ag NPs is placed within the device to facilitate aldehyde derivatization. Using this "nano-filter" AIMS, we screened 152 EB samples from healthy individuals and patients at different stage of LCa, more than 88 aldehydes were detected. In addition, by integrating a ML model, we enable fast and high-throughput screening of LCa patients using EB samples, with the candidate exhaled aldehydes. Moreover, by replacing the functionalized small molecule on Ag NPs, this "nano-filter" can be adapted for the detection of other classes of volatile metabolites.

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.

Conflicts of interest

There are no conflicts to declare.

Data availability

The machine learning model is available at GitHub repository https://github.com/sdnu-ms/AIMS. Materials, instrumentation, and experimental details characterization data (TEM, SEM, UV); construction and optimization of the "nano-filter" AIMS; mass spectrum of IS and different kind of EB samples; the kinetic curves of the critical aldehydes. All the data supporting this article have been included in the main text and the ESI. †

Ethical Statement

All human exhaled breath experiments were performed in accordance with the Declaration of Helsinki, and approved by the Medical Ethics Committee of Qilu Hospital of Shandong University (KYLL-202312-015). Informed consents were obtained from human participants of this study.

Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (12275163, 22134004), the Taishan Scholars Program of Shandong Province (tsqn202312159, tstp20240807). We thank the volunteers from Shandong Normal University, Qilu Hospital of Shandong University and Shandong Provincial Hospital for providing EB samples.

Notes and references

View Article Online
DOI: 10.1039/D6SC02074G

1. X. Chen, S. Zhou, Y. Wang, L. Zheng, S. Guan, D. Wang, L. Wang and X. Guan, *Trends Analyt Chem*, 2023, **162**, 117060.
2. H. Jin, Z. Zheng, Z. Cui, Y. Jiang, G. Chen, W. Li, Z. Wang, J. Wang, C. Yang, W. Song, X. Chen and Y. Zheng, *Nat Commun*, 2023, **14**, 4692.
3. T. Bruderer, T. Gaisl, M. T. Gaugg, N. Nowak, B. Streckenbach, S. Muller, A. Moeller, M. Kohler and R. Zenobi, *Chem Rev*, 2019, **119**, 10803-10828.
4. K. L. Viola, J. Sbarboro, R. Sureka, M. De, M. A. Bicca, J. Wang, S. Vasavada, S. Satpathy, S. Wu, H. Joshi, P. T. Velasco, K. MacRenaris, E. A. Waters, C. Lu, J. Phan, P. Lacor, P. Prasad, V. P. Dravid and W. L. Klein, *Nat Nanotechnol*, 2015, **10**, 91-98.
5. Y. Y. Broza, R. Vishinkin, O. Barash, M. K. Nakhleh and H. Haick, *Chem Soc Rev*, 2018, **47**, 4781-4859.
6. C. Beyer, *Nature Reviews Bioengineering*, 2025, **3**, 444.
7. X. Gu, A. Pan, L. Wu, J. Zhang, Z. Xu, T. Wen, M. Wang, X. Shi, L. Wu and Y. Qin, *Chem Sci*, 2024, **15**, 18411-18418.
8. R. J. Smyth, S. M. Toomey, A. Sartori, E. O'Hanrahan, S. D. Cuffe, O. S. Breathnach, R. K. Morgan and B. T. Hennessy, *J. Thorac Oncol*, 2018, **13**, 1213-1216.
9. G. Peng, U. Tisch, O. Adams, M. Hakim, N. Shehada, Y. Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten and H. Haick, *Nat Nanotechnol*, 2009, **4**, 669-673.
10. F. Djago, J. Lange and P. Poinot, *Nat Rev Chem*, 2021, **5**, 183-196.
11. Z. Xie, J. D. Morris, S. J. Mattingly, S. R. Sutaria, J. Huang, M. H. Nantz and X. A. Fu, *Anal Chem*, 2023, **95**, 4344-4352.
12. M. Corradi, I. Rubinstein, R. Andreoli, P. Manini, A. Caglieri, D. Poli, R. Alinovi and A. Mutti, *Am J Respir Crit Care Med*, 2003, **167**, 1380-1386.
13. M.-G. Zhou, Y. Liu, W.-W. Li, X. Yuan, X.-F. Zhan, J. Li, Y.-X. Duan, Y. Liu, Z.-H. Gao, Y. Cheng, S.-Q. Cheng, H. Li and Y. Liang, *Chinese Science Bulletin*, 2014, **59**, 1992-1998.
14. W. Lin, L. P. Conway, M. Vujasinovic, J. M. Lohr and D. Globisch, *Angew Chem Int Ed Engl*, 2021, **60**, 23232-23240.
15. S. R. Sutaria, S. S. Gori, J. D. Morris, Z. Xie, X. A. Fu and M. H. Nantz, *Metabolites*, 2022, **12**, 561.
16. L. E. McGivern, Z. H. Lim, Y. Yuan, Z. Bo, G. Wu, H. Bayley and Y. Qing, *Nature Communications*, 2025, **16**, 9409.
17. Z. Li, Y. Li, L. Zhan, L. Meng, X. Huang, T. Wang, Y. Li and Z. Nie, *Anal Chem*, 2021, **93**, 9158-9165.
18. X. A. Fu, M. Li, R. J. Knipp, M. H. Nantz and M. Bousamra, *Cancer Med-US*, 2013, **3**, 174-181.
19. S. X. Leong, Y. X. Leong, E. X. Tan, H. Y. F. Sim, C. S. L. Koh, Y. H. Lee, C. Chong, L. S. Ng, J. R. T. Chen, D. W. C. Pang, L. B. T. Nguyen, S. K. Boong, X. Han, Y. C. Kao, Y. H. Chua, G. C. Phan-Quang, I. Y. Phang, H. K. Lee, M. Y. Abdad, N. S. Tan and X. Y. Ling, *ACS Nano*, 2022, **16**, 2629-2639.
20. S. Antonowicz, Z. Bodai, T. Wiggins, S. R. Markar, P. R. Boshier, Y. M. Goh, M. E. Adam, H. Lu, H. Kudo, F. Rosini, R. Goldin, D. Moralli, C. M. Green, C. J. Peters, N. Habib, H. Gabra, R. C. Fitzgerald, Z. Takats and G. B. Hanna, *Nat Commun*, 2021, **12**, 1454.
21. J. Zhang, Y. Guo, X. Zhao, J. Pang, C. Pan, J. Wang, S. Wei, X. Yu, C. Zhang, Y. Chen, H. Yin and F. Xu, *Nat Rev Cardiol*, 2023, **20**, 495-509.
22. M. Yang, J. Fan, J. Zhang, J. Du and X. Peng, *Chem Sci*, 2018, **9**, 6758-6764.

View Article Online
DOI: 10.1039/D6SC02074G

23. O. Nakano-Baker, H. Fong, S. Shukla, R. V. Lee, L. Cai, D. Godin, T. Hennig, S. Rath, I. Novosselov, S. Dogan, M. Sarikaya and J. D. MacKenzie, *Biosens Bioelectron*, 2023, **229**, 115237.
24. G. Shang, D. Dinh, T. Mercer, S. Yan, S. Wang, B. Malaei, J. Luo, S. Lu and C. J. Zhong, *ACS Sens*, 2023, **8**, 1328-1338.
25. Y. Chen, Y. Zhang, F. Pan, J. Liu, K. Wang, C. Zhang, S. Cheng, L. Lu, W. Zhang, Z. Zhang, X. Zhi, Q. Zhang, G. Alfranca, J. M. de la Fuente, D. Chen and D. Cui, *ACS Nano*, 2016, **10**, 8169-8179.
26. L. W. Chan, M. N. Anahtar, T. H. Ong, K. E. Hern, R. R. Kunz and S. N. Bhatia, *Nat Nanotechnol*, 2020, **15**, 792-800.
27. Wenzheng Heng, Shukun Yin, Jihong Min, Canran Wang, Hong Han, Ehsan Shirzaei Sani, Jiahong Li, Yu Song, Harry B. Rossiter and W. Gao, *Science*, 2024, **385**, 954-961.
28. X. Zou, W. Zhou, Y. Lu, C. Shen, Z. Hu, H. Wang, H. Jiang and Y. Chu, *J. Gastroenterol and Hepatol*, 2016, **31**, 1837-1843.
29. D. Garcia-Gomez, L. Bregy, Y. Nussbaumer-Ochsner, T. Gaisl, M. Kohler and R. Zenobi, *Environ Sci Technol*, 2015, **49**, 12519-12524.
30. W. Wang, Y. Yang, Z. Chen, X. Wang, G. L. Zhang, T. He, L. Tong and B. Tang, *Anal Chem*, 2024, **96**, 787-793.
31. S. Qin, Y. Zhang, M. Shi, D. Miao, J. Lu, L. Wen and Y. Bai, *Nat Commun*, 2024, **15**, 4387.
32. L. May, K. Shows, P. Nana-Sinkam, H. Li and J. W. Landry, *Cancers (Basel)*, 2023, **15**, 3111.
33. P. B. Bach, *JAMA*, 2009, **301**, 539-541.
34. C. Glorieux, L. Cui, P. Zeng, X. Xia and P. Huang, *Cancer Commun (Lond)*, 2021, **41**, 432-435.
35. Y. Takefuji, *Journal of Hazardous Materials*, 2025, **488**, 137382.
36. S. Lee, M. Kim, B. J. Ahn and Y. Jang, *J Hazard Mater*, 2023, **455**, 131555.
37. D. Garcia-Gomez, P. Martinez-Lozano Sinues, C. Barrios-Collado, G. Vidal-de-Miguel, M. Gaugg and R. Zenobi, *Anal Chem*, 2015, **87**, 3087-3093.

View Article Online
DOI: 10.1039/D6SC02074G

The complete data that support the findings of this study (e.g., age, sex, etc. and MS data) are available from the corresponding author for research purpose only. The machine learning model is available at GitHub repository https://github.com/sdnu-ms/AIMS.