Environmental Science & Technology

pubs.acs.org/est

Article

Advancing Breathomics through Accurate Discrimination of Endogenous from Exogenous Volatiles in Breath

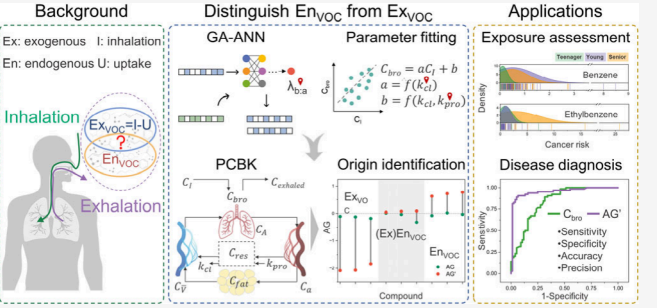
Zhengnan Cen,[⊥] Yuerun Huang,[⊥] Shangzhewen Li, Shanshan Dong, Wenshan Wang, and Xiang Li*

Cite This: https://doi.org/10.1021/acs.est.4c04575

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Breathomics, a growing field in exposure monitoring and clinical diagnostics, has faced accuracy challenges due to unclear contributing factors. This study aims to enhance the potential of breathomics in various frontiers by categorizing exhaled volatile organic compounds (VOCs) as endogenous or exogenous. Analyzing ambient air and breath samples from 271 volunteers via TD-GC × GC-TOF MS/FID, we identify and quantify 50 common VOCs in exhaled breath. Advanced quantitative structure−property relationships and compartment models are employed to obtain VOCs kinetic parameters. This in-depth approach allows us to accurately determine the alveolar concentration of VOCs and further discern their origins, facilitating personalized application of breathomics in exposure assessment and disease diagnosis. Our findings demonstrate that prolonged external exposure turns humans into secondary pollutant sources. Analysis of endogenous VOCs reveals that internal exposure poses more significant health risks than external. Moreover, by correcting environmental backgrounds, we improve the accuracy of gastrointestinal disease diagnostic models by 15−25%. This advancement in identifying VOC origins via compartmental models promises to elevate the clinical relevance of breathomics, marking a leap forward in exposure assessment and precision medicine.



KEYWORDS: *volatile organic compound, breath, origin identification, environmental exposure, disease diagnostics*

1. INTRODUCTION

Breathomics, an emerging interdisciplinary field integrating physiology, medicine, chemistry, and engineering, holds great potential for noninvasive diagnostics and therapy monitoring.[1,2] This field initially focused on discovering disease-specific biomarkers by analyzing trace gases in human breath.[3] However, despite the detection of over 1,000 volatile organic compounds (VOCs) in the past decade, none have achieved significant clinical breakthroughs in cancer detection.[4] This shortfall stems from the failure to discriminate endogenous VOCs with truly diagnostic potential from exogenous VOCs in exhaled breath.[5]

The high sensitivity of current methods in exhaled breath analysis often leads to the inadvertent detection of non-endogenous analytes.[6] They need to be regarded as exogenous VOCs, which act as confounders in disease diagnosis.[7] Therefore, exhaled breath analysis needs to consider the transport dynamics and the potential interference caused by exogenous VOCs.[8] Exogenous VOCs are transferred across the blood-air interface in alveoli during inhalation and subsequently permeate bodily tissues via the circulatory system.[9] Both these external compounds and endogenous VOCs diffuse from the blood into the breath and are excreted during exhalation.[10] This process risks misclassifying exogenous compounds as endogenous, particularly in comparative studies across different diseases or geographical locations, leading to potentially erroneous associations with specific diseases.[11] Furthermore, the chronic inhalation of environmental pollutants, and their subsequent absorption, metabolism, and accumulation in the body, complicates the interpretative reliability of breathomics in assessing human health.[12] Therefore, a key challenge lies in determining whether a VOC detected in both exhaled breath and room air is of internal origin or is merely a result of external contamination.

Researchers have responded to the challenge with three different strategies.[13] Some studies have overlooked environmental factors, reporting solely on observed exhaled VOC concentrations, which may yield false-positive results.[8] Others have required subjects to inhale filtered air, a method that is not only cumbersome but also probably introduces new contaminants.[14,15] Recent innovative approaches, such as the alveolar gradient (AG) and breath tracker algorithm, aim to tackle background variability.[16−18] However, the AG method,

Received: May 9, 2024
Revised: September 19, 2024
Accepted: September 20, 2024

reliant on subtracting inhaled concentrations, often lacks precision and requires environmental concentration adjustments.[19] The breath tracker algorithm, while effective in processing real-time data from proton transfer reaction-time-of-flight-mass spectrometry (PTR-TOF MS), falls short with offline samples.[20] Consequently, the development of more precise and robust methods for identifying endogenous and exogenous VOCs is essential to enhancing the practicality and accuracy of breathomics.

The aim of this study is to precisely differentiate between exogenous and endogenous VOCs using developed compartment models and further evaluate their distinct roles in environmental exposure and disease diagnostics. To achieve this aim, we collected 271 exhaled breath samples along with 101 ambient air samples using standardized methods and conducted nontarget breath analysis via two-dimensional gas chromatography coupled with time-of-flight mass spectrometry and flame ionization detector (GC × GC-TOF MS/FID). Through the development of compartment models based on blood-air partition coefficients ($\lambda_{b:a}$) predicted by a quantitative structure−property relationship (QSPR) model, we accurately determined the actual alveolar concentrations and tissue distribution of VOCs. This approach allowed us to compile a comprehensive catalog of endogenous and exogenous VOCs in exhaled breath, exploring their implications in both environmental exposure assessments and disease diagnostics. Our study revealed that while exogenous VOCs presented challenges in clinical diagnosis, they offered valuable insights in environmental exposure studies. Identifying endogenous VOCs significantly improved the clinical utility and accuracy of breathomics. Therefore, our research lays a foundational theoretical framework for advancing breathomics and opens new avenues for its broader application across various disciplines.

2. MATERIALS AND METHODS

2.1. Demographics. The study was performed in agreement with the Helsinki Declaration and approved by the Ethics Committee of Fudan University (FE22139I). Informed consent was obtained from all of the study participants.

For this study, we recruited a total of 271 human subjects to ensure a robust sample size capable of producing statistically significant results and facilitating a detailed analysis of the influencing factors. The demographic details of the participants were summarized in Table S1. We specifically excluded individuals with active respiratory disease symptoms, such as a productive cough or shortness of breath, as these could make sampling uncomfortable and potentially skew the results. Additionally, pregnant or breastfeeding individuals and those who consumed alcohol daily were excluded from the study. To maintain consistency in experimental conditions across all participants, we controlled various factors. All subjects were instructed to abstain from eating, drinking (except water), using cosmetics, and exercising for at least 12 h prior to breath sampling.

2.2. Sample Collection and Analysis. Exhaled breath samples in this study were collected using a ReCIVA sampler (Owlstone Medical, Cambridge, UK). The detailed procedure for sample collection has been described in our previous publications.[21] Briefly, sample collection occurred between 8:00 a.m. and 10:00 a.m. following an overnight fast. For each participant, we collected 3 L of breath gas at a flow rate of 200 mL·min^{-1}. The exhaled VOCs were captured trapped on multibed thermal desorption tubes containing Tenax/TA with a Carbograph 5TD (Markes Biomonitoring Tubes, Markes International Ltd., UK). Concurrently, environmental samples were obtained by passing 3 L of room air through a thermal desorption tube connected to the inlet of a hand-held air sampling pump (EDKORS Ltd., China). Analysis of both exhaled breath and environmental air samples was performed using GC × GC-TOF MS/FID. The details of analytical method have been summarized in Supporting Information.

2.3. Applicability Domain of QSPR Model. Applicability domain (AD) of the QSPR model was delineated by using a graphical approach, specifically through a Williams plot. This plot is a representation of standardized residuals against the leverage values of the compounds.[22] Leverage, which is the diagonal element of the Hat matrix in this context, was calculated as follows

$$h = diag[X_i^T (X_i^T X_{train})^{-1} X_i] \quad (1)$$

where, X_{train} represents the training set; $i = train$, validation and test set. AD warning limits were determined as three multiples of standard deviation (of standardized residuals), and the critical leverage value h^*:

$$h^* = 3 \times \frac{K+1}{N} \quad (2)$$

where, K represents the number of molecular descriptors, and N represents the number of observations.

2.4. Compartment Model. To accurately model the distribution and dynamics of VOCs in humans, we developed various compartment models tailored to $\lambda_{b:a}$. For VOCs with $\lambda_{b:a}$ less than 10, our model incorporated three compartments: the lung, fat, and rest of the body. Notably, upper airways have a significant impact on hydrophilic VOCs.[3] Consequently, for compounds with $\lambda_{b:a}$ exceeding 10, we introduced an additional bronchus compartment to our model. The compartment models assumed that VOC production and clearance mainly occur in the remaining body compartments, excluding the fat compartment. Detailed mass balance equations for each compartment were provided in the Supporting Information.

Physiological quantities cardiac output ($\dot{Q}_c$), fractional flow bronchioles (q_{bro}), fractional flow fat (q_{fat}) for each subject were estimated by PK-Sim software.[23] Alveolar ventilation ($\dot{V}_A$) was estimated from the empirical formula reported in elsewhere.[24] The blood-fat partition coefficient ($\lambda_{b:f}$) of VOCs was estimated from the linear free energy relationship reported by Abraham and Ibrahim.[25] The derivation of the ratio of mucus-air and mucus-blood partition coefficients ($\lambda_{muc:a}/\lambda_{muc:b}$) was complex, and the detailed process was summarized in the Supporting Information.

2.5. Health Risk Assessment. Health risk assessments were performed by referring to the human health assessment manual issued by the United States Environmental Protection Agency (US EPA). The external exposure assessment estimates the lifetime average daily dose (LADD) as follows:

$$LADD = \frac{C_I \times IR \times EF \times ED}{BW \times LT} \quad (3)$$

where, C_I is the environment concentration of VOCs, ng·L^{-1}; IR is the inhalation rate (L·h^{-1}); EF is the exposure frequency (h·day^{-1}); ED is the exposure duration in a lifetime (year); BW is the body weight (kg); LT is the averaged lifetime (year).

Based on $LADD$, the noncancer risk of a chemical, hazard quotient (HQ), can be expressed as

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

$$HQ_{ext} = \frac{LADD}{RfC \times RfV \div RfBW} \quad (4)$$

where, RfC is the inhalation reference concentration provided by US EPA ($ng \cdot L^{-1}$); RfV is the reference daily inhalation rate ($L \cdot day^{-1}$); $RfBW$ is the reference body weight (kg). The calculation parameters are listed in Table S3−S4.

Given the lack of methods for assessing health risks from internal exposure, we adopted an approach similar to that for external exposure. By summing HQs across different age stages, we calculated the lifetime noncancer risk of internal exposure. This method allows for a comparison between internal and external exposures.

$$LADD_i = \frac{AG'_i \times IR_i \times EF_i \times ED_i}{BW_i \times LT_i} \quad (5)$$

where, i represents an age stage (i = teenager, young adult or senior); AG'_i is the corrected alveolar gradient of VOC ($ng \cdot L^{-1}$).

The noncancer risk of a VOC for internal exposure can be calculated by the following equation:

$$HQ_{int} = \sum \frac{LADD_i}{RfC_i \times RfV_i \div RfBW_i} \quad (6)$$

The associated parameters used for calculating HQ from internal exposure were summarized in Table S3−S4.

The cancer-associated risks from external and internal exposure were calculated as follows:

$$CR_{ext} = C_I \times IUR \quad (7)$$

$$CR_{int} = AG' \times IUR \quad (8)$$

where, IUR is the inhalation unit risk $(ng \cdot L^{-1})^{-1}$ (Table S5).

2.6. Fat Bioavailability. This study expressed the bioavailability of exogenous VOCs using the ratio of VOCs entering the fat to those inhaled.[26]

$$Fatbioavailability = \frac{VOCenteringthefat}{VOCinhaled} = \frac{C_a \times \dot{Q}_c \times \frac{1}{1+\lambda_{b:f}}}{C_I \times \dot{V}_A} \quad (9)$$

where, C_a is the arterial concentration ($ng \cdot L^{-1}$); $\dot{Q}_c$ is the cardiac output ($L \cdot min^{-1}$); $\dot{V}_A$ is the alveolar ventilation ($L \cdot min^{-1}$); $\lambda_{b:f}$ is the blood-fat partition coefficient.

2.7. Statistical Analysis. The raw VOC data were quantitatively analyzed by using an internal standard normalization method. The k-means cluster was used to expand the application of the breath tracker algorithm in the offline samples. The number of clusters was optimized by the elbow method. The experimentally measured $\lambda_{b:a}$ from literatures and molecular descriptors of 195 compounds were used to construct the QSPR model based on genetic algorithm-artificial neural network (GA-ANN). GA was implemented for feature selection in the Chemdes descriptors set. ANN was employed to describe the complex nonlinear relationships between the $\lambda_{b:a}$ and chemical structures. The predictive efficacy of the GA-ANN model was assessed using the coefficient of determination (R^2), mean absolute error (MAE), and root-mean-square error (RMSE). Random forest (RF) algorithm was used to develop the disease diagnostic models. The diagnostic model employed 5-fold cross-validation to prevent overfitting. Diagnostic performance was assessed by using receiver operating characteristic (ROC) curves. The predictive power of the model was evaluated via the sensitivity, specificity, accuracy, precision, recall, F1, and area under curves (AUC).

All of the analyses were performed using R Studio (version 4.1.1, RStudio Inc., Boston, MA, USA).

3. RESULTS AND DISCUSSION

3.1. Inadequacy of Current Methods in Exhaled VOCs Origin Determination. Breathomics research, which plays a vital role in disease diagnostics, currently faces significant challenges. A major hurdle is the difficulty in distinguishing between endogenous and exogenous VOCs, casting doubts on the reliability of biomarkers.[7] To understand the state of breathomics research, we performed a thorough search in the Web of Science database from its inception until February 14, 2024, using "volatile organic compounds" and "breath" as keywords. This search yielded 3,570 relevant articles covering disease diagnosis, health monitoring, environmental exposure, and other frontiers (Figure 1A). However, nearly all of these studies (approximately 3,550) overlooked potential confounding factors arising from the mixture of environmental compounds with exhaled air. Less than 30 studies attempted to differentiate between endogenous and exogenous VOCs,[11,27] highlighting a gap due to complex human-environment interactions and the absence of effective methods for tracing VOC origins.[17] This methodological shortfall increases the risk of misattribution, leading to breath analysis results that are challenging to replicate or apply clinically.[8] Acknowledging these issues is essential for advancing breath analysis techniques. Consequently, we collected extensive breath samples to evaluate and enhance the precision of methods distinguishing endogenous from exogenous VOCs.

We recruited 271 volunteers from three research centers between 2021 and 2023 to provide exhaled breath samples (Figure 1B). These volunteers were categorized into teenager (≤18 years), young adult (19−49 years), and senior (≥50 years) groups based on age.[24] This stratification enabled us to explore the impact of varying environmental exposure durations and age-related metabolic functions on endogenous and exogenous VOCs identification. Using GC × GC-TOF MS/FID, we conducted a nontargeted analysis of the breath and ambient air samples, identifying a wide range of VOCs. We found 50 VOCs with an occurrence rate above 80% and categorized them into chemical groups (Figure 1C). The results indicated that alkanes (24%), aromatics (18%), and ketones (18%) dominated the composition of exhaled breath, followed by aldehydes (10%), alkenes (8%), acids (8%), esters (6%), heterocycles (6%), and nitriles (2%). This diversity in VOCs facilitated a comprehensive evaluation of origin discrimination methodologies, with subsequent focus directed toward elucidating the origins of these 50 VOCs.

The AG method is widely used to distinguish between endogenous and exogenous VOCs due to its rapid identification of environmental influences.[28] However, its magnitude and polarity can be affected by individual and environmental variations.[29] To improve reliability in identifying endogenous compounds, we analyzed a large sample set with the AG method, calculating peak intensity differences of target VOCs in exhaled versus environmental samples.[30] We established benchmarks for differentiating endogenous VOCs from exogenous VOCs through average AG and endogenous

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

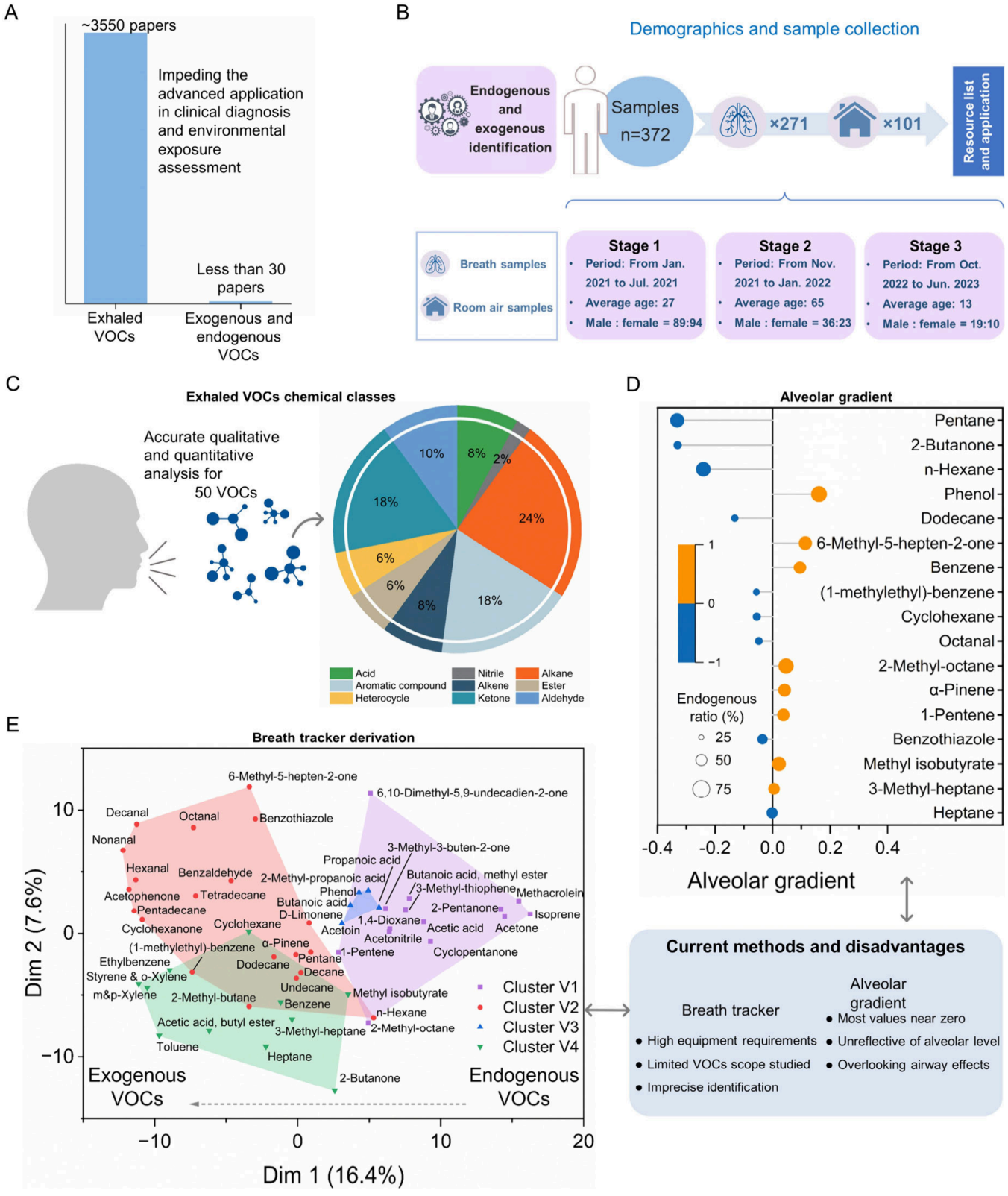


Figure 1. Current state and methods for exogenous and endogenous VOCs identification in breath researches. (A) Publication statistics on differentiating exhaled VOCs origins. (B) Demographics and sample overview in this study. (C) Chemical classes of 50 exhaled VOCs accurately identified and quantified by TD-GC × GC-TOF MS/FID. (D) Alveolar gradients and endogenous rates of 17 exhaled VOCs. (E) Breath tracker-based cluster analysis for exhaled VOCs.

ratios. Typically, VOCs with an average AG > 0 and an endogenous ratio >70% are regarded as endogenous. Our analysis identified 16 VOCs meeting these criteria (Figure S1). Conversely, VOCs showing a negative AG typically indicate an environmental origin, as they are inhaled at higher levels than they are exhaled.[31] We found 17 VOCs with an average AG <

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

0 and an endogenous ratio <30%, classifying them as exogenous (Figure S1). However, 17 VOCs had ambiguous classifications, with average AGs near zero or endogenous ratios between 30% and 70%, making their origins uncertain (Figure 1D). This uncertainty arises from the similarity in concentrations of these VOCs in exhaled breath and environmental air, compounded by individual and environmental variations. Moreover, the current AG represents VOC concentration after passing through the upper airways, not directly from the alveoli, thus overlooking the effect of bronchus on exhaled VOCs. Studies have shown that VOCs with a high $\lambda_{b:a}$ significantly interact with the water-like mucus membrane lining the conductive airways, reducing the concentration of hydrophilic volatiles as they move from the alveoli to the airway opening.[32] This discrepancy between alveolar air and measured end-tidal air depends on factors like airway temperature, perfusion, breathing patterns, and primarily $\lambda_{b:a}$.[33] Therefore, it is necessary to take into account the physical and chemical characteristics of various VOCs and convert exhaled concentrations to true alveolar concentrations.

Breath tracker algorithm offers an alternative method for distinguishing endogenous from exogenous VOCs.[20] However, its use is limited to online PTR-TOF MS data. To broaden its applicability, we developed a clustering approach inspired by the breath tracker concept, using acetone and isoprene as benchmarks for endogenous VOCs.[18] This method involves classifying compounds that cluster closely with these benchmarks as endogenous, while those further away are deemed exogenous. We applied k-means clustering on samples and VOCs, optimizing the number of clusters through the elbow function (Figure S2A). This process divided the samples into five clusters, reflecting the chemical heterogeneity in exhaled breath and the environmental air (Figure S2B). Our analysis identified notable similarities between the two, indicating the exhalation of compounds into the ambient air. We then applied k-means clustering to differentiate between endogenous and exogenous VOCs, categorizing 50 VOCs into four clusters based on their chemical behaviors (Figure 1E). Acetone and isoprene, indicative of endogenous VOCs, predominantly formed one cluster, while unequivocally exogenous compounds such as toluene, xylene, and ethylbenzene formed another. A third cluster, containing three acids, overlapped with the endogenous group, suggesting a link to microbiome emissions.[34] However, the algorithm faced difficulties with compounds in middle clusters of mixed origins, highlighting the need for more accurate methods to pinpoint VOC origins. A comprehensive understanding of the processes that govern VOC distribution and transport across different biological matrices is essential to accurately identifying their origins.

3.2. Accurately Estimating VOC Kinetics for Compartment Model Development. To improve the distinction between exogenous and endogenous VOCs in exhaled breath, we developed physiological-compartment-based kinetic (PCBK) models. These models integrated environmental, exhaled air, bronchial, lung, arterial, venous, fat, and remaining body compartments through physiological parameters, distribution processes, and mass balance (Figure S3). Detailed descriptions of the PCBK model were summarized in the Supporting Information. By estimation of the actual alveolar concentrations (C_A), our models offer robust theoretical support for identifying endogenous VOCs. Simulation results showed that the $\lambda_{b:a}$, production (k_{pro}) and clearance rates (k_{cl}) are crucial physicochemical parameters for C_A calculation. However, a comprehensive database of these parameters in exhaled VOCs is currently lacking. Given the high cost, time investment, and significant sample size requirements for experimental determination, theoretical and computational methods present a viable alternative for parameter estimation.

We developed a QSPR model using GA-ANN to predict $\lambda_{b:a}$. This parameter significantly influences the uptake of exogenous vapors and the exhalation of endogenous compounds. The variability of $\lambda_{b:a}$ in the human volatilome can span over 12 orders of magnitude,[35] indicating that compounds with similar exhaled levels may have vastly different concentrations in the alveoli and blood. Accurate $\lambda_{b:a}$ values are essential for understanding VOCs behavior within the body, identifying associated biochemical pathways, and assessing their diagnostic and therapeutic potential. Our model was trained on a data set comprising $\lambda_{b:a}$ values for 195 VOCs gathered from the literature.[35−41] We obtained chemical structures from PubChem, optimized geometric structures, and calculated molecular descriptors using ChemDes.[42] The GA-ANN model was designed to efficiently map nonlinear relationships between these descriptors and $\lambda_{b:a}$.[40,43] The workflow of this model is illustrated in Figure 2A. After running the model 20 times with varying initial populations (Figure S4), we evaluated the frequency of descriptor selection (Figure S5A), incrementally incorporating the most relevant variables into the ANN. Six variables proved ideal for our model (Figures S5B-D and S6), involving two-dimensional topological and charge indices (MLFER_S, Qmin, XLogP, ATSC0c) and three-dimensional shape and functionality (RDF10s, MoRSEU15), suggesting that polarizability, charge density, and changes in Gibbs energy of solvation predominantly determine the distribution mechanism at the blood-gas interface.[42] This result led to a network structure with an input layer with 6 neurons, a hidden layer with 11 neurons, and an output layer with a single neuron for $\lambda_{b:a}$ prediction (Figure 2B). The model underwent extensive training to adjust weights and biases, demonstrating high reliability and excellent fitting across validation and test data sets (Training-R^2 = 0.95625, Validation-R^2 = 0.90362, Test-R^2 = 0.90073) (Figure 2C). The RMSE for the test set was 0.34230, comparable to the RMSE of the training set (0.23598), indicating consistent accuracy and generalizability. Although the R^2 and RMSE of nonlinear model in this study are comparable to those reported by Luan et al.,[41] our model covers a broader range of compounds, thus expanding its applicability and improving stability. The residual plot further validated the accuracy of the model, showing a uniform distribution of residuals without systematic errors in the QSPR models (Figure 2D).

We further assessed the AD of our model, which is a critical step in any QSAR research. The AD of a QSAR model refers to the physicochemical, structural or biological space, knowledge or information on which the training set is developed and is applicable for predicting new unknown compounds.[44] In this study, the AD was determined using a leverage approach (eq 1, Figure 2E). The Williams plot identified two response outliers with standardized residuals over 3 and four structural outliers with leverage values exceeding h* for our model (eq 2). Structural outliers were considered as good leverage points since the information that these four compounds encode made the QSAR model more precise.[45] Overall, the majority of VOCs in our study set fall within the AD, with the model reliably predicting the $\lambda_{b:a}$ for 96.92% (189/195) of the compounds. We also assessed the AD for 33 exhaled VOCs

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

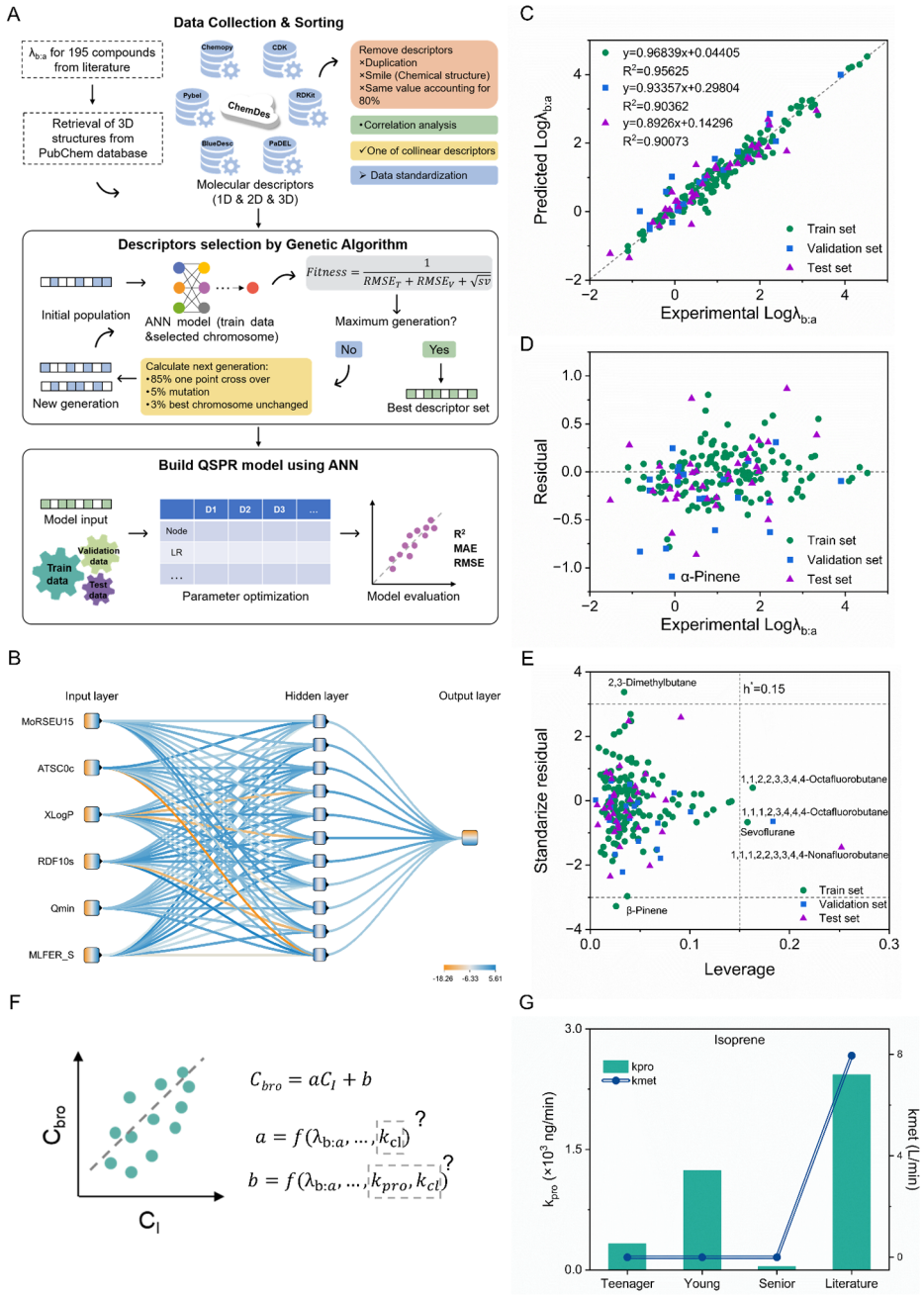


Figure 2. Partition coefficient and VOCs kinetics estimation for compartment model development. (A) Flowchart of the GA-ANN algorithm-based QSPR model. (B) The architecture of ANN shows the nonlinear relationship between chemical structures and $\lambda_{b:a}$. (C) Comparison of experimental log $\lambda_{b:a}$ versus predicted log $\lambda_{b:a}$ by the QSPR model. (D) Plot of residuals versus experimental log $\lambda_{b:a}$. (E) Applicability domain of developed QSPR model. (F) Estimating VOCs production and clearance rates using the affine function of the exhaled and inhaled concentrations. (G) Comparison of estimated isoprene production and clearance rates with reported values.

with unknown $\lambda_{b:a}$, finding most within the AD of the QSPR model, suggesting their predicted values were likely reliable and close to experimental measurements (Table S2). Among the 50 VOCs analyzed, we found diverse $\lambda_{b:a}$ values, with 11 below 10, 25 exceeding 100, and the remainder between 10 and 100. Furthermore, we observed that compounds with the same substituents showed increased $\lambda_{b:a}$ with higher carbon numbers, while chemical complexity and the addition of polar functional groups also raised $\lambda_{b:a}$. This observation confirmed that $\lambda_{b:a}$ increased with structural complexity and molecular polarity.[46]

After the gaps in $\lambda_{b:a}$, the focus shifted to determining k_{pro} and k_{cl}. Our PCBK model can estimate the C_A devoid of environmental impacts, termed as corrected alveolar gradients (AG'), thereby allowing for quantifying the endogenous fraction of exhaled VOCs. The equation for AG' resembles a straight line of the form

$$AG' = C_A(0) = C_A(C_I) - \alpha C_I \quad (10)$$

Inhaled concentration C_I is the variable here. Different constants α were applied to VOCs with varying $\lambda_{b:a}$. The complete equations were available in the Supporting Information (eqs S41, S48 and S54). Our analysis indicated

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

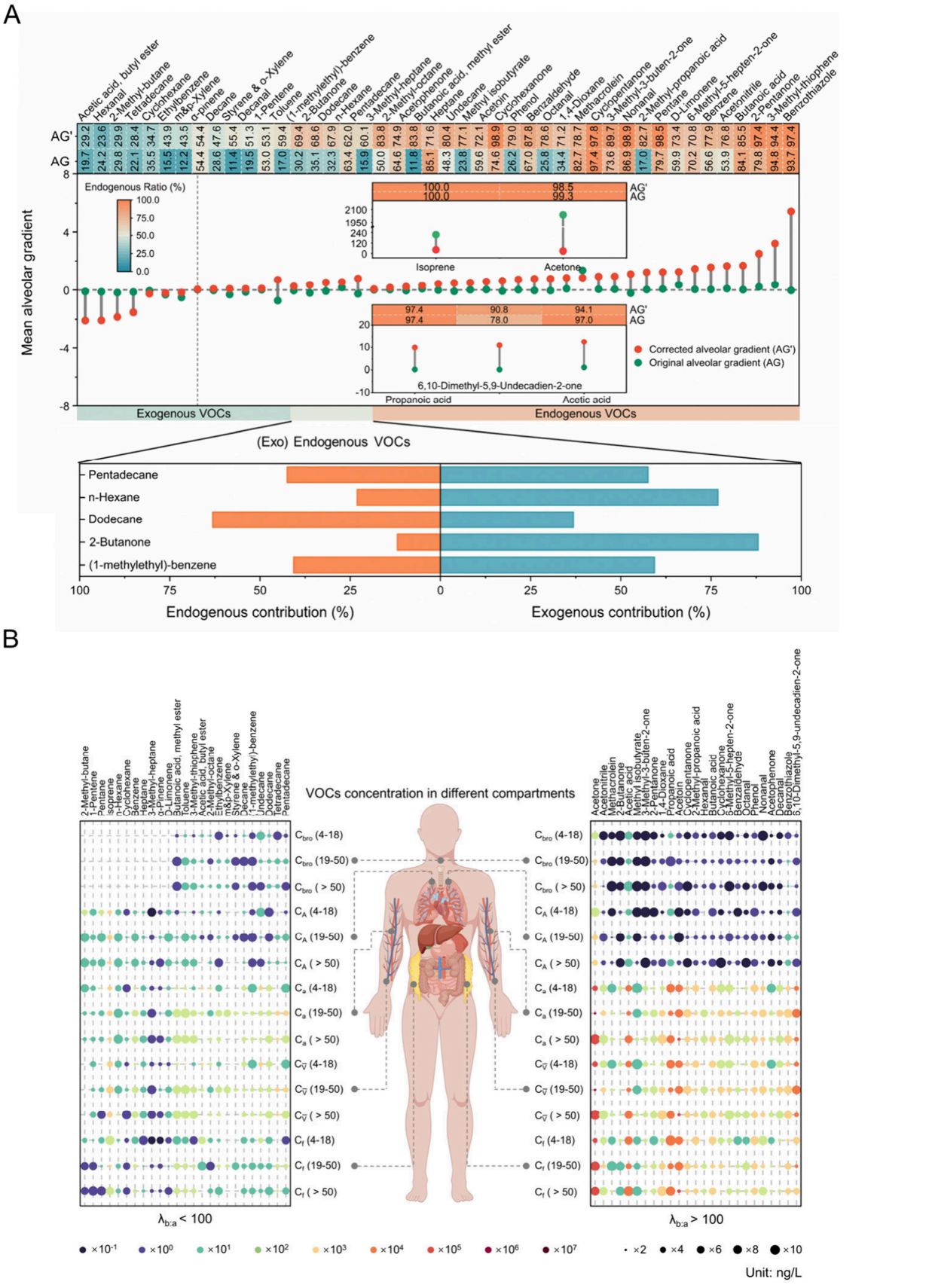


Figure 3. Compilation of exogenous and endogenous VOCs in exhaled breath and their distribution characteristics in human body. (A) *AG* and *AG′* of 50 exhaled VOCs. The color bar in the top subplot corresponds to the endogenous ratio. The bottom subplot shows the endogenous contributions of VOCs that may have both exogenous and endogenous origins. (B) VOCs concentration distributions in different compartments across different age groups. The circle color represents the orders of magnitude, and the circle size indicates the quantity.

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

that the constant α was completely determined by physiological quantities including $\dot{V}_A$, $\dot{Q}_c$, conductance parameter (D), q_{bro}, q_{fat}, $\lambda_{muc:a}/\lambda_{muc:b}$, $\lambda_{b:a}$ and k_{cl}. Under steady-state conditions, we could measure or estimate all parameters except k_{cl}. The k_{cl} and k_{pro} are derived from the affine function of exhaled concentrations (C_{bro}) and C_I (Figure 2F, eqs S13−14 and S30−31). By matching with experimental data, we were able to fit the intercepts and slopes of the functions, thereby calculating k_{cl} and k_{pro}. Given the impact of age on these rates,[24] we conducted distinct k_{cl} and k_{pro} estimations for the 50 VOCs across different age groups (Figure S7). Notably, there existed significant interindividual variations in both k_{cl} and k_{pro}, hence the estimated k_{cl} and k_{pro} here reflected population-wide averages.[19] Our findings on k_{pro} and k_{cl} aligned with the limited existing research (Figure 2G). For instance, our estimated k_{pro} for isoprene in young adults (1.24×10^3 ng·min^{-1}) closely matched previously reported values (2.43×10^3 ng·min^{-1}), but we observed a zero k_{cl} across age groups, significantly deviating from the reported 10 L·min^{-1} in previous research.[19] It is important to note that the small numerators in the derivative formulas for k_{cl} and k_{pro} may lead to significant errors if the sample size is not sufficiently large. Given that previous studies involved significantly fewer participants ($n < 10$) than our own (n = 271), the estimated k_{cl} and k_{pro} in this study should be more reliable. Our estimated k_{cl} and k_{pro} for most VOCs were within the same order of magnitude across different age group and aligned with existing literature.[33,47] Moreover, employing k_{cl} and k_{pro} from our study, the predicted venous concentration of acetone closely matched the typical blood measurement of 1 mg·L^{-1}, further validating the accuracy of our model.[33]

3.3. Inventory and Tissue Distribution of Endogenous and Exogenous VOCs. By applying $\lambda_{b:a}$, k_{pro}, k_{cl} and other parameters (Table S2) to the PCBK model, we determined the gradient α and the AG'. Figure 3A illustrated the comparison of average alveolar gradients and endogenous ratios for 50 exhaled VOCs before and after correction, highlighting more distinct differences postcorrection, thus improving the differentiation between endogenous and exogenous VOCs. Specifically, 43 VOCs exhibited an average AG' above 0, yet 11 of these had an endogenous ratio under 70%. Accounting for individual variability in alveolar gradients, 32 compounds with average $AG' > 0$ and endogenous ratios >70% were classified as endogenous. Meanwhile, 13 VOCs were identified as exogenous due to their average $AG' < 0$ or endogenous ratio <60%. The remaining 5 compounds showed slightly positive average AG' values with endogenous ratios between 60% and 70%, indicating a potential mixed origin from human metabolism and ambient air. Further analysis revealed their average endogenous contributions (eq S56), leading to a breakdown of these compounds into endogenous and exogenous parts for closer scrutiny. We determined the average endogenous contributions of these 5 VOCs, such as 2-butanone at 11.95%, *n*-hexane at 23.10%, (1-methylethyl)-benzene at 40.69%, pentadecane at 42.46%, and dodecane at 63.13% (Figure 3A). Thus, our study offered a detailed catalog of the endogenous and exogenous sources of 50 commonly exhaled VOCs. This comprehensive understanding not only reveals the interaction between humans and the environment but also refines the application of exhaled VOCs across various fields based on their origins.

We additionally employed the PCBK model to analyze the distribution of VOCs across various tissues (Figure 3B). Notably, acetone was the most prevalent VOC in all compartments, with concentrations three to 4 orders of magnitude higher than other VOCs. Isoprene also exhibited higher concentrations in all compartments compared to VOCs with $\lambda_{b:a} < 100$. This likely arose from acetone and isoprene having higher k_{pro} values than other VOCs. Beyond these two compounds, the distribution of other VOCs varied, with those having a $\lambda_{b:a}$ less than 100 concentrating mostly in the blood, fat, and alveoli and least in the bronchus. Conversely, highly soluble VOCs ($\lambda_{b:a} > 100$) were found at lower concentrations in the alveolar and bronchial compartments. At the blood-gas exchange interface, the concentration variation of highly soluble VOCs spanned two to 3 orders of magnitude, highlighting the crucial influence of $\lambda_{b:a}$ on the transmembrane transport efficiency of polar VOCs.[32] Unlike VOCs with $\lambda_{b:a} < 100$, fat appeared to be a key reservoir for highly soluble VOCs, where their concentrations were similar to those in the blood. Our analysis also indicated that the distribution patterns of certain endogenous VOCs might change with age, with compounds like pentane and undecane showing a negative age correlation, suggesting their potential as aging biomarkers.[48] The compartmental concentrations of endogenous *n*-hexane, benzene, heptane, butanoic acid, methyl ester, 3-methylthiophene, methacrolein, 1,4-dioxane, acetoin, benzaldehyde, and acetophenone increased with age. However, these substances could not be directly regarded as age biomarkers since their elevated levels in exhaled breath might also result from widespread exposure to environmental stressors. This necessitated further research to isolate age as a determinant. The distribution of VOCs across age groups showed that age alone did not dictate concentration differences; intrinsic production and metabolic rates were also crucial factors. Similar to acetone and isoprene, 20 VOCs peaked in young adults, suggesting a link to more active metabolism and behavioral activities. In contrast, compounds such as 1-pentene, decane, dodecane, 2-methyl-octane, 2-butanone, acetic acid, methyl isobutyrate, propanoic acid, 2-methyl-propanoic acid, butanoic acid, and phenol exhibited the lowest concentrations in the young adults, possibly due to enzyme-mediated metabolism or chemical degradation.[49−51] The detailed quantification of VOCs across compartments enhances our understanding of their absorption, distribution, metabolism, and excretion mechanism in the body, which is vital for assessing the health implications of VOC exposure in different populations.

3.4. Applications of Distinguished Endogenous and Exogenous VOCs. By accurately distinguishing between exogenous and endogenous VOCs in exhaled breath, we enhanced the utility of breathomics in environmental exposure assessment and diagnosing diseases.[52] Breath acts as a pivotal medium for human-environment interaction, with its chemical components providing insights into environmental exposure levels.[53] While conventional methods have predominantly concentrated on external sources of VOC exposure to humans, the mixture of exogenous and endogenous VOCs in exhaled breath poses a challenge for effectively assessing external exposure levels.[54] Discerning between these two components enables a more holistic exposure assessment and a comparison between external and internal exposures.

Inhaling exogenous VOCs from the environment is a crucial route to external exposure. We determined the inhalation rates of exogenous VOCs, including the exogenous fractions of (exo) endogenous VOCs, using C_I and corresponding $\dot{V}_A$

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

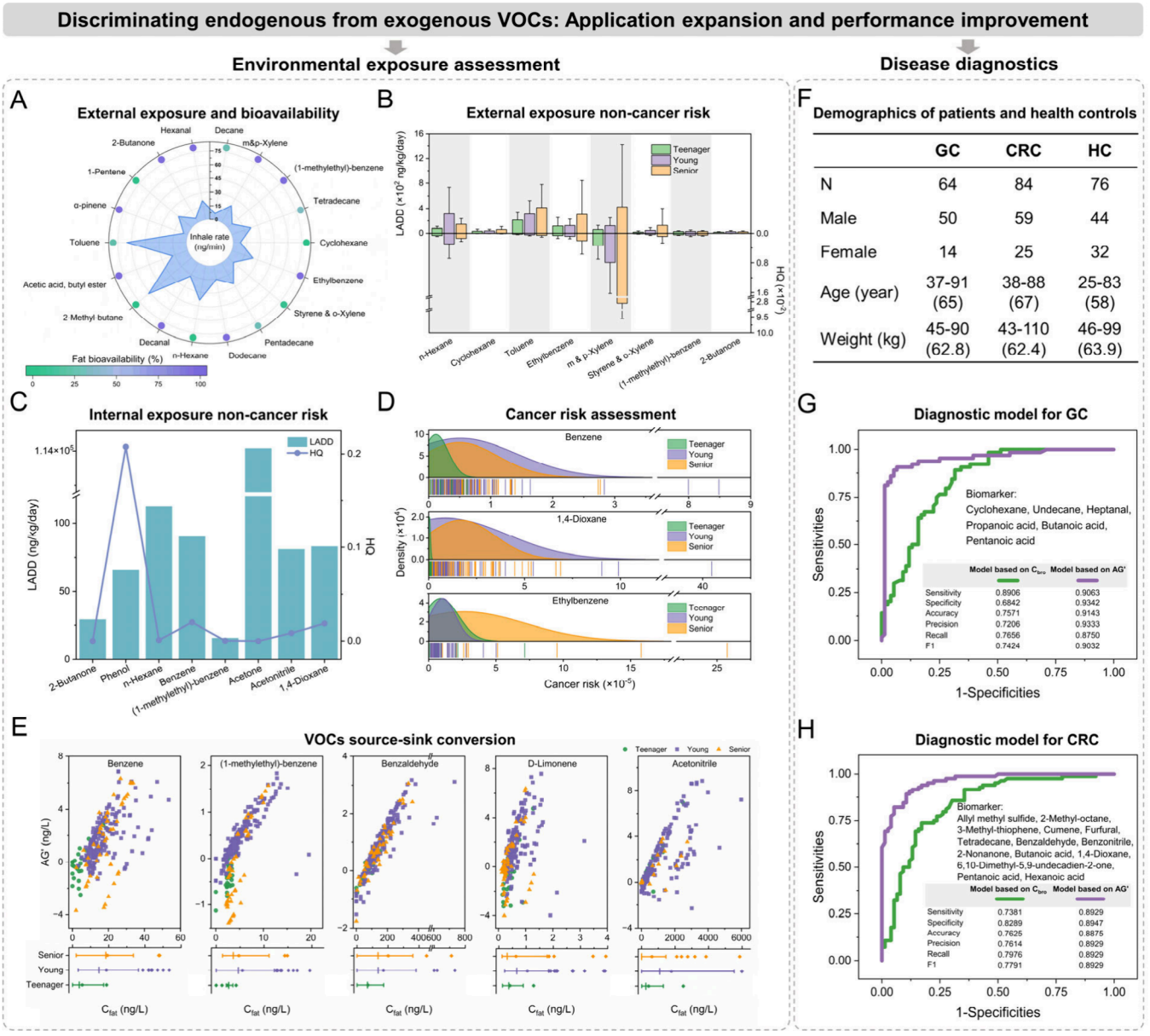


Figure 4. Personalized application of exogenous and endogenous VOCs in environmental exposure assessment and disease diagnostics. (A) The inhaled rate and fat bioavailability of exogenous VOCs. (B) LADD and *HQ* illustrate the noncancer risk assessment of external exposure across different age groups. (C) Noncancer risk assessment of internal exposure based on LADD and *HQ*. (D) Density curves display cancer risk assessment of benzene, ethylbenzene and 1,4-dioxane across different age groups. (E) The relationship between C_{fat} and AG' shows the source-sink conversion for five endogenous VOCs. (F) Demographics of gastrointestinal tumor patients and healthy controls. Age and weight are displayed as min-max (mean). (G) Performance comparison of GC diagnostic models using $C_{bro}$and AG'. (H) Performance comparison of CRC diagnostic models using C_{bro} and AG'.

(Figure 4A), thus quantifying actual exposure doses. The analysis showed that inhalation rates for 18 exogenous VOCs were all below 70 ng·min^{-1}, with toluene exhibiting the highest rate at 65.08 ng·min^{-1}. To assess health risks from exogenous VOC exposure, we calculated the LADD and compared it with reference data from the EPA for eight compounds (eq 3, Figure 4B, Tables S3−4). The LADD values ranged from 17.70 ng·kg^{-1}·day^{-1} to 422.61 ng·kg^{-1}·day^{-1}. The *HQ* for these VOCs varied from 1.95×10^{-5} to 2.84×10^{-2} (eq 4, Figure 4B), with toluene, ethylbenzene, m and *p*-xylene, and *n*-hexane presenting higher risks, highlighting their potential as hazardous pollutants at the sampling sites. Nonetheless, all *HQ* values were below the safety threshold of 0.2, indicating that inhalation exposure alone might not significantly risk health.[55] Adjusting for age-related changes in physiological parameters, we found seniors were at greater risk from inhalation exposure than young adults and teenagers (Figure 4B), underscoring the need for cleaner environments for the elderly.

Internal exposure represents the body's response to environmental stressors. We combined exposure data from teenagers, young adults, and seniors to model lifetime internal exposure to endogenous VOCs. Given the limited reference values for many VOCs, we estimated the LADD for eight endogenous VOCs using AG' (eq 5, Table S3−4). Figure 4C illustrated the noncancer risks from endogenous VOC exposure (eq 6). The lifetime average *HQ* for these VOCs

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

ranged from 2.56×10^{-5} to 0.208, with most values below the safety threshold of 0.2, indicating minimal noncancer risks. Despite its high internal exposure dose (1.14×10^5 ng·kg^{-1}·day^{-1}), acetone presented the lowest estimated HQ because of its low toxicity.[56] Conversely, phenol exhibited the highest HQ value among the eight VOCs; followed by benzene (0.0197) and 1,4-dioxane (0.0184), suggesting potential nervous and respiratory system impacts. We further assessed the comprehensive noncancer risks posed by all eight endogenous VOCs (HQ_{total} = 0.256), which indicated higher risks from internal exposure compared to external exposure (HQ_{total} = 0.00543−0.0339). Therefore, regarding human exposure, the external and internal exposures should be coupled to accurately evaluate the health risks. Although the noncancer risks did not exceed the overall safe level of 1,[57] cancer-related risks remained a concern. Figure 4D presented a cancer risk assessment for three carcinogenic VOCs across different age groups (eqs 7-8), including two endogenous (benzene and 1,4-dioxane) and one exogenous (ethylbenzene) VOCs (Table S5). The estimated lifetime cancer risks for these VOCs fell between 10^{-5} and 10^{-4}, pointing to potential carcinogenic risks from both internal and external exposures.[57] Age-specific analysis showed young adults were most at risk from benzene and 1,4-dioxane, while seniors faced higher risks from ethylbenzene, with these carcinogens being the least threatening to teenagers. This suggested that cancer risks from external exposure increased with age or exposure duration, whereas internal risks were linked to metabolic and physiological functions. Therefore, distinguishing exhaled VOCs origins not only enables accurate quantification of external environmental exposure doses but also facilitates assessments of internal health risks. These insights highlight the different cancer risk mechanisms from internal versus external exposure, thus calling for stratified strategies in health risk management.

Quantifying the fate of exogenous VOCs in the body helps to clarify the long-term effects of these inhaled compounds. Fat tissues are typically recognized as primary storage sites for VOCs.[25] We evaluated the fat bioavailability of 18 exogenous VOCs in fat (eq 9), finding a wide range from 0.38% to 100% (Figure 4A). There was a trend where fat bioavailability increased with the molecular weight and the presence of polar functional groups, suggesting compounds with higher molecular weights or polarity were more prone to fat storage. Notably, nine VOCs showed a fat bioavailability of 100% or higher, indicating pre-existing accumulation in fat. Further analysis of the correlation between exogenous VOC fat concentrations (C_{fat}) and body fat content revealed that most C_{fat} plateaued at fat contents above 10% (Figure S8), suggesting a dilution effect at higher fat levels. However, aromatic hydrocarbons such as toluene, ethylbenzene, m- and p-xylene, styrene, xylene, and (1-methylethyl)-benzene exhibited slight increases with fat content, likely due to their higher LADD, which could counteract the dilution effect of increased fat content. Our study suggests that aromatic hydrocarbons represent the most significant long-term health risk among all exogenous VOCs.

Differentiating between endogenous and exogenous VOCs reveals potential flaws in previous assessments of the inhalation exposure risks. While benzene is typically regarded as exogenous,[24] our findings indicate it may also have endogenous characteristics, as shown by its AG'. Similar observations were made for compounds like (1-methylethyl)-benzene, benzaldehyde, D-limonene, and acetonitrile, potentially due to sink-source conversion effects. To explore the determinants behind the endogenous output of these five VOCs, we investigated the relationship between their AG' and C_{fat} (Figure 4E). Our analysis found a significant positive impact of C_{fat} on the AG' of these VOCs. This suggested that the role of human breath as a VOC source or sink was highly dependent on C_{fat}. Specifically, AG' values turned predominantly positive when C_{fat} surpassed certain thresholds, implying that exogenous VOCs could undergo reverse release after accumulating to a certain extent in fat. We also observed that the thresholds of C_{fat} required for this source-sink conversion varied with the $\lambda_{b:a}$ of each compound. The highest C_{fat} threshold required for this conversion was for acetonitrile, approximately 200 ng·L^{-1}. As $\lambda_{b:a}$ decreased, C_{fat} threshold levels followed with benzaldehyde (∼70 ng·L^{-1}), benzene (∼10 ng·L^{-1}), (1-methylethyl)-benzene (∼3 ng·L^{-1}), and D-limonene (∼0.4 ng·L^{-1}) in sequence. These results indicate that compounds with lower $\lambda_{b:a}$ values are more likely to turn the human body into a secondary source. Furthermore, due to prolonged exposure and greater accumulation of environmental substances in fat, adults are more likely to become secondary sources of these pollutants (Figure 4E). The similar C_{fat} levels in young adults and seniors might be due to dilution effects, indicating that the fat tissue could inhibit the release of these pollutants. Our findings reveal that under high environmental concentrations some pollutants migrate from the alveoli to the capillary and pulmonary veins due to the partial pressure gradient, making humans act as pollutant sinks. These compounds can then be distributed and potentially stored within the body, to be exhaled or otherwise excreted when external levels decrease.[24] Additionally, their low k_{cl} suggested that these compounds were eliminated slowly from the body, facilitating their accumulation and gradual release. This challenges traditional views on the origins of VOCs and refines our understanding of the health risks they pose.

The majority of breath research hitherto undertaken has focused on exploring endogenous biomarkers for diseases.[1,3] Although these breathomics-based diagnostic models were statistically promising, their clinical applicability often fell short.[4] This limitation is mainly due to their failure to distinguish between exhaled VOC concentration changes caused by environmental exposure and metabolic changes. We enhanced disease diagnostic models by incorporating endogenous VOCs from exhaled breath, using gastrointestinal tumor patients and healthy controls as examples (Figure 4F). Recent study has shown that ALDH1A3 gene deletion and pyruvate metabolism abnormalities in gastric cancer (GC) patients significantly alter six exhaled VOCs.[58] Fourteen exhaled VOC biomarkers in colorectal (CRC) patients are closely linked to the gut microbiome.[59] We developed models based on the C_{bro} and AG' of these VOC biomarkers. Figures 4G-H illustrates ROC curves for the diagnosis models of GC and CRC cancers. Analysis using the RF classifier on C_{bro} of six GC biomarkers yielded an accuracy of 0.7571, sensitivity of 0.8906, specificity of 0.6842, and AUC of 0.8376, validating the efficacy of these potential biomarkers. Models developed with AG' showed superior performance, with accuracy increasing to 0.9143, sensitivity to 0.9063, specificity to 0.9342, and AUC to 0.9512. Adjusting for environmental influences also significantly improved the CRC model's accuracy by 16.39%, sensitivity by 20.97%, specificity by 7.94%, and AUC by

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

14.44%. These results demonstrated that previous disease diagnostic models, which failed to account for confounding effects of environmental VOCs, might have erroneously interpreted variations in environmental concentrations as changes in biomarkers. A novel approach recently suggested simulating environmental VOCs at the time of sampling to improve model accuracy.[60] However, our study employed a different strategy, utilizing AG' to refine diagnostic models. This method directly mitigated the impact of environmental VOCs, bypassing the biases of algorithmic environmental concentration predictions. Our findings reveal that AG'-based models not only enhance prediction accuracy but also improve the interpretability of disease-specific metabolic changes against environmental backgrounds. Therefore, our research identifies a critical gap in the statistical robustness and clinical applicability of diagnostic models by emphasizing the importance of considering environmental VOCs.

4. IMPLICATIONS

Several limitations might affect the results obtained in this study. First, additional factors influencing the solubility of compound (e.g., blood proteins binding, differences in blood composition) were not taken into consideration.[61] This omission means that the $\lambda_{b:a}$ predictions should be viewed as approximations rather than precise values, potentially introducing bias and uncertainty in estimating VOC tissue concentrations. Second, the endogenous compounds identified include VOCs from gut microbiota and undigested food residues, which are conceptually different from those produced by cellular metabolism. Despite collecting exhaled breath samples after an overnight fast, the influence of food on exhaled VOCs could not be completely eliminated. Moreover, due to the unclear mechanism of gut-released VOCs being exhaled, our developed PCBK model does not explicitly classify a gut compartment, preventing a clear distinction between VOCs produced by the host and microbiota. Future improvements to the PCBK model can provide better technical support for exploring the exhaled VOC origins.

Overall, we demonstrate that exhaled VOCs are affected by alveoli, mucus, and external environment conditions, limiting the effectiveness of directly using exhaled concentrations for environmental exposure assessments or as disease biomarkers. We developed the PCBK model that estimates the distribution of VOCs across various tissues and quantifies the interaction between environmental exposures and endogenous production. By accurately calculating alveolar concentrations without external effects, our model accurately differentiates between endogenous and exogenous VOCs, essential for personalized applications based on their origins. The PCBK model elucidates the complex processes of VOCs absorption, distribution, metabolism, and excretion within the body. These advancements enable a more comprehensive assessment of health impacts from both internal and external exposures, aiding in the refinement of exposure evaluation systems and stratified management strategies. Furthermore, accurately identifying the origin of exhaled VOCs significantly enhances the accuracy of the breath biopsy, leading to more targeted and effective healthcare interventions. Therefore, our work significantly contributes to the fields of public health, environmental exposure assessment, and precision medicine by propelling breathomics forward.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c04575.

Supplementary Methods, Supplementary Figures (Figure S1-S8), Supplementary Tables (Table S1-S5) (PDF)

AUTHOR INFORMATION

Corresponding Author

Xiang Li − *Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P. R. China; Institute of Eco-Chongming (IEC), Shanghai 200062, P. R. China;* orcid.org/0000-0002-0434-3057; Email: lixiang@fudan.edu.cn

Authors

Zhengnan Cen − *Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P. R. China*

Yuerun Huang − *Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P. R. China*

Shangzhewen Li − *Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P. R. China*

Shanshan Dong − *Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P. R. China*

Wenshan Wang − *Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P. R. China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c04575

Author Contributions

[⊥]The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Z.C. and Y.H. contributed equally to this work.

Funding

National Natural Science Foundation of China (No. 22276038 and 22476023) Agilent Research Gift (No. 4956)

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 22276038 and 22476023) and Agilent Research Gift (No. 4956).

REFERENCES

(1) Djago, F.; Lange, J.; Poinot, P. Induced volatolomics of pathologies. *Nature Reviews Chemistry* **2021**, *5* (3), 183−196.

(2) Brinkman, P.; Ahmed, W. M.; Gomez, C.; Knobel, H. H.; Weda, H.; Vink, T. J.; Nijsen, T. M.; Wheelock, C. E.; Dahlen, S.-E.; Montuschi, P.; Knowles, R. G.; Vijverberg, S. J.; Maitland-van der Zee, A. H.; Sterk, P. J.; Fowler, S. J. Exhaled volatile organic compounds as markers for medication use in asthma. *Eur. Respir. J.* **2020**, *55* (2), 2019.

(3) Hakim, M.; Broza, Y. Y.; Barash, O.; Peled, N.; Phillips, M.; Amann, A.; Haick, H. Volatile Organic Compounds of Lung Cancer and Possible Biochemical Pathways. *Chem. Rev.* **2012**, *112* (11), 5949−5966.

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(4) Cao, W. Q.; Duan, Y. X. Breath analysis: Potential for clinical diagnosis and exposure assessment. *Clin Chem.* **2006**, *52* (5), 800−811.
(5) Pleil, J. D.; Stiegel, M. A.; Risby, T. H. Clinical breath analysis: discriminating between human endogenous compounds and exogenous (environmental) chemical confounders. *J. Breath Res.* **2013**, *7* (1), 017107.
(6) Xu, M. J.; Tang, Z. T.; Duan, Y. X.; Liu, Y. GC-Based Techniques for Breath Analysis: Current Status, Challenges, and Prospects. *Critical Reviews in Analytical Chemistry* **2016**, *46* (4), 291−304.
(7) Ibrahim, W.; Carr, L.; Cordell, R.; Wilde, M. J.; Salman, D.; Monks, P. S.; Thomas, P.; Brightling, C. E.; Siddiqui, S.; Greening, N. J. Breathomics for the clinician: the use of volatile organic compounds in respiratory diseases. *Thorax* **2021**, *76* (5), 514−521.
(8) Salman, D.; Ibrahim, W.; Kanabar, A.; Joyce, A.; Zhao, B.; Singapuri, A.; Wilde, M.; Cordell, R. L; McNally, T.; Ruszkiewicz, D.; Hadjithekli, A.; Free, R.; Greening, N.; Gaillard, E. A; Beardsmore, C.; Monks, P.; Brightling, C.; Siddiqui, S.; Thomas, C L P. The variability of volatile organic compounds in the indoor air of clinical environments. *J. Breath Res.* **2022**, *16* (1), 3565.
(9) Di Gilio, A.; Palmisani, J.; Ventrella, G.; Facchini, L.; Catino, A.; Varesano, N.; Pizzutilo, P.; Galetta, D.; Borelli, M.; Barbieri, P.; Licen, S.; de Gennaro, G. Breath Analysis: Comparison among Methodological Approaches for Breath Sampling. *Molecules* **2020**, *25* (24), 5823.
(10) Pham, Y. L.; Beauchamp, J. Breath Biomarkers in Diagnostic Applications. *Molecules* **2021**, *26* (18), 5514.
(11) Westhoff, M.; Friedrich, M.; Baumbach, J. I. Simultaneous measurement of inhaled air and exhaled breath by double multi-capillary column ion-mobility spectrometry, a new method for breath analysis: results of a feasibility study. *ERJ. Open Res.* **2022**, *8* (1), 00493-2021.
(12) Gaude, E.; Nakhleh, M. K.; Patassini, S.; Boschmans, J.; Allsworth, M.; Boyle, B.; van der Schee, M. P. Targeted breath analysis: exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes. *J. Breath Res.* **2019**, *13* (3), 032001.
(13) Phillips, M.; Herrera, J.; Krishnan, S.; Zain, M.; Greenberg, J.; Cataneo, R. N. Variation in volatile organic compounds in the breath of normal humans. *J. Chromatogr B* **1999**, *729* (1−2), 75−88.
(14) Belluomo, I.; Boshier, P. R.; Myridakis, A.; Vadhwana, B.; Markar, S. R.; Spanel, P.; Hanna, G. B. Selected ion flow tube mass spectrometry for targeted analysis of volatile organic compounds in human breath. *Nat. Protoc* **2021**, *16* (7), 3419−3438.
(15) Doran, S. L. F.; Romano, A.; Hanna, G. B. Optimisation of sampling parameters for standardised exhaled breath sampling. *J. Breath Res.* **2018**, *12* (1), 016007.
(16) Phillips, M.; Sabas, M.; Greenberg, J. Increased pentane and carbon disulfide in the breath of patients with schizophrenia. *Journal of Clinical Pathology* **1993**, *46* (9), 861−864.
(17) Westhoff, M.; Kessler, M.; Baumbach, J. I. Alveolar gradients in breath analysis. A pilot study with comparison of room air and inhaled air by simultaneous measurements using ion mobility spectrometry. *J. Breath Res.* **2023**, *17* (4), 046009.
(18) Sukul, P.; Schubert, J. K.; Kamysek, S.; Trefz, P.; Miekisch, W. Applied upper-airway resistance instantly affects breath components: a unique insight into pulmonary medicine. *J. Breath Res.* **2017**, *11* (4), 047108.
(19) Unterkofler, K.; King, J.; Mochalski, P.; Jandacka, M.; Koc, H.; Teschl, S.; Amann, A.; Teschl, G. Modeling-based determination of physiological parameters of systemic VOCs by breath gas analysis: a pilot study. *J. Breath Res.* **2015**, *9* (3), 036002.
(20) Trefz, P.; Schmidt, M.; Oertel, P.; Obermeier, J.; Brock, B.; Kamysek, S.; Dunkl, J.; Zimmermann, R.; Schubert, J. K.; Miekisch, W. Continuous Real Time Breath Gas Monitoring in the Clinical Environment by Proton-Transfer-Reaction-Time-of-Flight-Mass Spectrometry. *Anal. Chem.* **2013**, *85* (21), 10321−10329.
(21) Cen, Z. N.; Lu, B. Q.; Ji, Y. Y.; Chen, J.; Liu, Y. Q.; Jiang, J. K.; Li, X.; Li, X. Virus-induced breath biomarkers: A new perspective to study the metabolic responses of COVID-19 vaccinees. *Talanta* **2023**, *260*, 124577.
(22) Zuvela, P.; Macur, K.; Jay Liu, J.; Baczek, T. Exploiting non-linear relationships between retention time and molecular structure of peptides originating from proteomes and comparing three multivariate approaches. *J. Pharm. Biomed. Anal.* **2016**, *127*, 94−100.
(23) Willmann, S.; Lippert, J.; Sevestre, M.; Solodenko, J.; Fois, F.; Schmitt, W. PK-Sim(R): A physiologically based pharmacokinetic 'whole-body' model. *Biosilico* **2003**, *1* (4), 121−124.
(24) Sun, X.; He, J. Z.; Yang, X. D. Human breath as a source of VOCs in the built environment, Part II: Concentration levels, emission rates and factor analysis. *Building and Environment* **2017**, *123*, 437−445.
(25) Abraham, M. H.; Ibrahim, A. Air to fat and blood to fat distribution of volatile organic compounds and drugs: Linear free energy analyses. *Eur. J. Med. Chem.* **2006**, *41* (12), 1430−1438.
(26) Shen, H. T.; Han, J. L.; Guan, R. F.; Cai, D. L.; Zheng, Y. B.; Meng, Z.; Chen, Q.; Li, J. G.; Wu, Y. N. Use of different endpoints to determine the bioavailability of polychlorinated dibenzo-*p*-dioxins/furans (PCDD/Fs) and polychlorinated biphenyls (PCBs) in Sprague-Dawley rats. *Sci. Rep-Uk* **2022**, *12* (1), 3.
(27) Westhoff, M.; Rickermann, M.; Litterst, P.; Baumbach, J. I. Exogenous factors of influence on exhaled breath analysis by ion-mobility spectrometry (MCC/IMS). *International Journal for Ion Mobility Spectrometry* **2019**, *22* (2), 59−69.
(28) Spanel, P.; Dryahina, K.; Smith, D. A quantitative study of the influence of inhaled compounds on their concentrations in exhaled breath. *J. Breath Res.* **2013**, *7* (1), 017106.
(29) Schubert, J. K.; Miekisch, W.; Birken, T.; Geiger, K.; Nöldge-Schomburg, G. F. E. Impact of inspired substance concentrations on the results of breath analysis in mechanically ventilated patients. *Biomarkers* **2005**, *10* (2−3), 138−152.
(30) Phillips, M.; Cataneo, R. N.; Chaturvedi, A.; Kaplan, P. D.; Libardoni, M.; Mundada, M.; Patel, U.; Zhang, X. Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *PLoS One* **2013**, *8* (9), No. e75274.
(31) Phillips, M.; Greenberg, J.; Sabas, M. Alveolar gradient of pentane in normal human breath. *Free Radical Research* **1994**, *20* (5), 333−337.
(32) Mochalski, P.; King, J.; Mayhew, C. A.; Unterkofler, K. Modelling of Breath and Various Blood Volatilomic Profiles-Implications for Breath Volatile Analysis. *Molecules* **2022**, *27* (8), 2381.
(33) Ager, C.; Unterkofler, K.; Mochalski, P.; Teschl, S.; Teschl, G.; Mayhew, C. A.; King, J. Modeling-based determination of physiological parameters of systemic VOCs by breath gas analysis, part 2. *J. Breath Res.* **2018**, *12* (3), 036011.
(34) Sanna, S.; van Zuydam, N. R.; Mahajan, A.; Kurilshikov, A.; Vich Vila, A.; Vosa, U.; Mujagic, Z.; Masclee, A. A. M.; Jonkers, D. M. A. E.; Oosting, M.; Joosten, L. A. B.; Netea, M. G.; Franke, L.; Zhernakova, A.; Fu, J.; Wijmenga, C.; McCarthy, M. I. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* **2019**, *51* (4), x.
(35) Haick, H.; Broza, Y. Y.; Mochalski, P.; Ruzsanyi, V.; Amann, A. Assessment, origin, and implementation of breath volatile cancer markers. *Chem. Soc. Rev.* **2014**, *43* (5), 1423−1449.
(36) Sprunger, L. M.; Gibbs, J.; Acree, W. E.; Abraham, M. H. Correlation of human and animal air-to-blood partition coefficients with a single linear free energy relationship model. *Qsar & Combinatorial Science* **2008**, *27* (9), 1130−1139.
(37) Abraham, M. H.; Gola, J. R. M.; Gil-Lostes, J.; Acree, W. E.; Cometto-Muniz, J. E. Determination of solvation descriptors for terpene hydrocarbons from chromatographic measurements. *J. Chromatogr A* **2013**, *1293*, 133−141.
(38) Meulenberg, C. J. W.; Vijverberg, H. P. M. Empirical relations predicting human and rat tissue: air partition coefficients of volatile

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

organic compounds. *Toxicol. Appl. Pharmacol.* **2000**, *165* (3), 206−216.
(39) Mochalski, P.; King, J.; Kupferthaler, A.; Unterkofler, K.; Hinterhuber, H.; Amann, A. Human Blood and Plasma Partition Coefficients for C4-C8 n-alkanes, Isoalkanes, and 1-alkenes. *International Journal of Toxicology* **2012**, *31* (3), 267−275.
(40) Konoz, E.; Golmohammadi, H. Prediction of air-to-blood partition coefficients of volatile organic compounds using genetic algorithm and artificial neural network. *Anal. Chim. Acta* **2008**, *619* (2), 157−164.
(41) Luan, F.; Liu, H. T.; Ma, W. P.; Fan, B. T. QSPR analysis of air-to-blood distribution of volatile organic compounds. *Ecotoxicology and Environmental Safety* **2008**, *71* (3), 731−739.
(42) Dong, J.; Cao, D. S.; Miao, H. Y.; Liu, S.; Deng, B. C.; Yun, Y. H.; Wang, N. N.; Lu, A. P.; Zeng, W. B.; Chen, A. F. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of Cheminformatics* **2015**, *7*, 60.
(43) Fan, D. Q.; Yu, R.; Fu, S. Y.; Yue, L.; Wu, C. F.; Shui, Z. H.; Liu, K. N.; Song, Q. L.; Sun, M. J.; Jiang, C. Y. Precise design and characteristics prediction of Ultra-High Performance Concrete (UHPC) based on artificial intelligence techniques. *Cement & Concrete Composites* **2021**, *122*, 104171.
(44) Ghanei-Nasab, S.; Hadizadeh, F.; Foroumadi, A.; Marjani, A. A QSAR Study for the Prediction of Inhibitory Activity of Coumarin Derivatives for the Treatment of Alzheimer's Disease. *Arabian Journal for Science and Engineering* **2021**, *46* (6), 5523−5531.
(45) Galimberti, F.; Moretto, A.; Papa, E. Application of chemometric methods and QSAR models to support pesticide risk assessment starting from ecotoxicological datasets. *Water Res.* **2020**, *174*, 115583.
(46) Agapiou, A.; Amann, A.; Mochalski, P.; Statheropoulos, M.; Thomas, C. L. P. Trace detection of endogenous human volatile organic compounds for search, rescue and emergency applications. *Trac-Trend Anal Chem.* **2015**, *66*, 158−175.
(47) Szabó, A.; Unterkofler, K.; Mochalski, P.; Jandacka, M.; Ruzsanyi, V.; Szabó, G.; Mohácsi, A.; Teschl, S.; Teschl, G.; King, J. Modeling of breath methane concentration profiles during exercise on an ergometer. *J. Breath Res.* **2016**, *10* (1), 017105.
(48) Jia, Z. A.; Thavasi, V.; Venkatesan, T.; Lee, P. Y. Breath Analysis for Lung Cancer Early Detection-A Clinical Study. *Metabolites* **2023**, *13* (12), 1197.
(49) Weber, R.; Streckenbach, B.; Welti, L.; Inci, D.; Kohler, M.; Perkins, N.; Zenobi, R.; Micic, S.; Moeller, A. Online breath analysis with SESI/HRMS for metabolic signatures in children with allergic asthma. *Frontiers in Molecular Biosciences* **2023**, *10*, 1154536.
(50) Wang, N.; Pugliese, G.; Carrito, M.; Moura, C.; Vasconcelos, P.; Cera, N.; Li, M.; Nobre, P.; Georgiadis, J. R.; Schubert, J. K.; Williams, J. Breath chemical markers of sexual arousal in humans. *Sci. Rep-Uk* **2022**, *12* (1), 6.
(51) Porto-Figueira, P.; Pereira, J. A. M.; Cámara, J. S. Exploring the potential of needle trap microextraction combined with chromatographic and statistical data to discriminate different types of cancer based on urinary volatomic biosignature. *Anal. Chim. Acta* **2018**, *1023*, 53−63.
(52) Das, S.; Pal, M. Review-Non-Invasive Monitoring of Human Health by Exhaled Breath Analysis: A Comprehensive Review. *J. Electrochem. Soc.* **2020**, *167* (3), 037562.
(53) Webster, E. M.; Qian, H.; Mackay, D.; Christensen, R. D.; Tietjen, B.; Zaleski, R. Modeling Human Exposure to Indoor Contaminants: External Source to Body Tissues. *Environ. Sci. Technol.* **2016**, *50* (16), 8697−8704.
(54) Valcke, M.; Krishnan, K. Assessing the impact of the duration and intensity of inhalation exposure on the magnitude of the variability of internal dose metrics in children and adults. *Inhalation Toxicology* **2011**, *23* (14), 863−877.
(55) Wang, H.; Wang, H. M.; Wang, K. L.; Xiong, J. Y.; Huang, S. D.; Wolfson, J. M.; Koutrakis, P. Characterization of chemical transport in human skin and building material. *J. Hazard Mater.* **2023**, *458*, 131917.
(56) Gentry, P. R.; Covington, T. R.; Clewell, H.; Anderson, M. E. Application of a physiologically based pharmacokinetic model for reference dose and reference concentration estimation for acetone. *J. Toxicol. Env. Health Part A* **2003**, *66* (23), 2209−2225.
(57) Pinthong, N.; Thepanondh, S.; Kondo, A. Source Identification of VOCs and their Environmental Health Risk in a Petrochemical Industrial Area. *Aerosol and Air Quality Research* **2022**, *22* (2), 210064.
(58) Chen, J.; Ji, Y. Y.; Liu, Y. Q.; Cen, Z. N.; Chen, Y. W.; Zhang, Y. X.; Li, X. W.; Li, X. Exhaled volatolomics profiling facilitates personalized screening for gastric cancer. *Cancer Letters* **2024**, *590*, 216881.
(59) Liu, Y. Q.; Ji, Y. Y.; Chen, J.; Zhang, Y. X.; Li, X. W.; Li, X. Pioneering noninvasive colorectal cancer detection with an AI-enhanced breath volatilomics platform. *Theranostics* **2024**, *14* (11), 4240−4255.
(60) Tsou, P. H.; Lin, Z. L.; Pan, Y. C.; Yang, H. C.; Chang, C. J.; Liang, S. K.; Wen, Y. F.; Chang, C. H.; Chang, L. Y.; Yu, K. L.; Liu, C. J.; Keng, L. T.; Lee, M. R.; Ko, J. C.; Huang, G. H.; Li, Y. K. Exploring Volatile Organic Compounds in Breath for High-Accuracy Prediction of Lung Cancer. *Cancers* **2021**, *13* (6), 1431.
(61) Kramer, C.; Mochalski, P.; Unterkofler, K.; Agapiou, A.; Ruzsanyi, V.; Liedl, K. R. Prediction of blood:air and fat:air partition coefficients of volatile organic compounds for the interpretation of data in breath gas analysis. *J. Breath Res.* **2016**, *10* (1), 017103.

https://doi.org/10.1021/acs.est.4c04575
Environ. Sci. Technol. XXXX, XXX, XXX−XXX