Supplementary Information

Development and validation of a cross-species breath metabolomics platform for translational VOC analysis

Zhiheng Yu[1], Yongyan Ji[1], Zijun Meng[1], Xiang Li[1] *

[1] Shanghai Key Laboratory of Air Quality and Environmental Health, Department of Environmental Science & Engineering, Fudan University, Shanghai 200438, P.R. China.

*Corresponding author. Email: lixiang@fudan.edu.cn

This SI file (Page S1-S11) includes:

Text S1-S5

Figure S1-S6

Table S1-Table S3

Tables of Content

Text S1. Inclusion and Exclusion Criteria for Human Participants.

To minimize potential confounding effects on breath metabolomic profiles, participants meeting any of the following criteria were excluded from the study. Individuals who had undergone abdominal surgery within the preceding three months were excluded. Participants with comorbid systemic diseases were also excluded, including chronic respiratory disorders (such as asthma or chronic obstructive pulmonary disease), hepatic or renal dysfunction (defined as alanine aminotransferase or aspartate aminotransferase levels exceeding twofold the upper limit of normal, or abnormal serum creatinine levels), or a history of malignant disease. Women who were pregnant or breastfeeding at the time of sampling were excluded. In addition, individuals with long-term smoking or alcohol abuse histories, or those who had received antibiotics or probiotic supplements within four weeks prior to breath sample collection, were not eligible for enrollment.

Text S2. Urine sample preparation and VOC extraction.

Urine samples were obtained from 26 patients with IBD and 17 healthy volunteers and stored at −40 °C until analysis. Prior to measurement, samples were thawed at 4 °C overnight. For VOC extraction, 2 mL of urine was transferred into a 20 mL glass headspace vial, followed by the addition of 1 μL of internal standard solution (20 μg/mL). Volatile compounds were extracted from the headspace using a HiSorb sorptive extraction probe. The probe was positioned in the vial headspace above the liquid phase to prevent direct contact with the urine matrix. Extraction was performed at 60 °C for 30 min. After extraction, the HiSorb probe was placed into a thermal desorption tube and analyzed by TD–GC–MS/MS.

Text S3. Animal Husbandry and Habituation Protocol

All mice were housed under specific pathogen–free (SPF) conditions in individually ventilated cages, maintained at a constant ambient temperature of 22 °C with a 12-h light/12-h dark cycle. During the three-day pre-modeling period, animals were provided ad libitum access to sterilized drinking water and standard irradiated chow. Following disease induction, both the DSS and EEN groups received DSS solution as the sole source of drinking fluid, and food was withdrawn during the dark phase. During the light phase, mice in the DSS group were provided with standard irradiated chow, whereas mice in the EEN group received exclusive enteral nutrition in place of

regular chow.

To minimize stress associated with physical restraint during breath sampling, a three-day habituation protocol was implemented prior to data collection. On the first day, the restraining device was placed inside the home cage, allowing mice to freely explore and acclimate to its presence and odor. On the second day, mice were gently guided to voluntarily enter the restrainer to familiarize themselves with the confined space. On the third day, a mock sampling procedure was performed, during which mice were placed into the sealed sampling apparatus and supplied with purified air for 15 min, closely mimicking the conditions of actual breath collection.

Text S4. Detailed Instrumental Parameters for TD–GC–MS/MS Analysis

Thermal desorption–gas chromatography–tandem mass spectrometry (TD–GC–MS/MS) analyses were performed using a Markes TD100-xr thermal desorption system (Markes International Ltd., UK) coupled to a GC–MS/MS platform.

Thermal desorption (TD) conditions.

The TD system was operated as follows. For primary tube desorption, a dry purge was applied at a flow rate of 100 mL/min for 10 min, followed by thermal desorption at 300 °C with a carrier gas flow of 30 mL/min for 5 min. Volatile compounds were focused on a cold trap (model U-T11GPC-2S) maintained at 30 °C. For secondary desorption, a pre-purge step of 2 min was applied prior to trap desorption at 300 °C for 5 min, with a maximum heating rate exceeding 100 °C/s. Secondary desorption was conducted in splitless mode. The transfer line temperature was set to 180 °C.

Gas chromatography (GC) conditions.

Chromatographic separation was achieved using a DB-624 capillary column (60 m × 0.25 mm i.d. × 1.4 μm film thickness; Agilent J&W Scientific). High-purity helium (99.999%) was used as the carrier gas at a constant flow rate of 1.0 mL/min. The oven temperature program was as follows: an initial temperature of 40 °C held for 5 min, ramped to 160 °C at 5 °C/min, then increased to 230 °C at 10 °C/min and held for 21 min.

Mass spectrometry (MS) conditions.

The ion source temperature was set to 230 °C, the quadrupole temperature to 150 °C, and the transfer line temperature to 250 °C. A solvent delay of 2 min was applied. Data acquisition was performed in full-scan mode (m/z 30–350) for qualitative analysis, while selected ion monitoring (SIM) or multiple reaction monitoring (MRM)

modes were used for quantitative analysis.

Text S5. Data Preprocessing and Machine Learning Methods

Data Cleaning and Preprocessing Details

For exhaled VOCs data, missing values were imputed using the half-minimum method, in which missing or undetected values are replaced by half of the minimum non-zero value observed for each feature. For visualization in heatmaps and bubble plots, data were scaled to the range 0–1 using Min-Max normalization. Mouse data were transformed prior to PCA analysis using $\log_{10}(\text{Concentration}+\varepsilon)$, where ε is a small constant to avoid errors caused by zero values.

Machine Learning Models

The following algorithms were applied in the analysis: Random Forest (RF), XGBoost, Support Vector Machine (SVM), Lasso Regression, Bridge Regression (a general penalized regression method that includes Lasso and Ridge as special cases), Gradient Boosting Machine (GBM), and K-Nearest Neighbors (KNN).

Table S1. DAI Scoring Criteria

Score	Weight Loss(%)	Stool Consistency	Rectal Bleeding
0	None	Normal, well-formed stool	No blood
1	1-5%	Slightly soft stool	Occult blood positive
2	5-10%	Clearly soft or mucoid stool	Mild visible blood
3	10-15%	Watery stool	Moderate bleeding
4	＞15%	Severe diarrhea / unformed stool	Severe bleeding

Table S2. Linear dynamic ranges, calibration concentration ranges, and correlation coefficients (R^2) for the targeted VOCs.

Compound Name	Calibration Concentration Range (ppm)	Linear Dynamic Range (ppm)	Correlation Coefficient (R^2)
Acetic acid	0.5-20	0.5-20	0.9970
Propionic acid	0.5-20	0.5-20	0.9912
Dimethyl sulfide	0.5-20	0.5-20	0.9924
Isobutyric acid	0.5-20	0.5-20	0.9913
Butyric acid	0.5-20	0.5-20	0.9935
Valeric acid	0.5-20	0.5-20	0.9968
D-limonene	0.5-20	0.5-20	0.9974
Benzonitrile	0.5-20	0.5-20	0.9960
Phenol	0.5-20	0.5-20	0.9981
Geranyl acetone	0.5-20	0.5-20	0.9928
Propionaldehyde	0.5-20	0.5-20	0.9853
Acetone	100-2000	100-2000	0.9974
Isopropanol	0.5-20	0.5-20	0.9967
Methacrolein	0.5-20	0.5-20	0.9924
Ethyl acetate	0.5-20	0.5-20	0.9726
Tert pentanol	0.5-20	0.5-20	0.9957
Butanol	0.5-20	0.5-20	0.9873
2-pentanone	0.5-20	0.5-20	0.9932
Acetoin	0.5-20	0.5-20	0.9975
Hexanal	0.5-20	0.5-20	0.9896
1,2-propanediol	0.5-20	0.5-20	0.9959
2-heptanone	0.5-20	0.5-20	0.9946
Heptanal	0.5-20	0.5-20	0.9920
Benzaldehyde	0.5-20	0.5-20	0.9785
Octanal aldehyde	0.5-20	0.5-20	0.9942
Nonanal	0.5-20	0.5-20	0.9871
Decanal	0.5-20	0.5-20	0.9952
Decane	0.5-20	0.5-20	0.9906
Undecane	0.5-20	0.5-20	0.9934
Dodecane	0.5-20	0.5-20	0.9927
Tridecane	0.5-20	0.5-20	0.9847
Tetradecane	0.5-20	0.5-20	0.9968
Pentadecane	0.5-20	0.5-20	0.9931

Table S3. List of 71 targeted VOCs and their CAS numbers

Names	CAS	Names	CAS
Isopentane	78-78-4	Decanal	112-31-2
Isoprene	78-79-5	Cyclohexane	110-82-7
Acetone	67-64-1	1-(Methylthio)-propane	3877-15-4
Cyclopentane	287-92-3	α-Pinene	2437-95-8
Hexane	110-54-3	α-Methylstyrene	98-83-9
2,4-Dimethylpentane	108-08-7	Decane	124-18-5
Ethyl acetate	141-78-6	Undecane	1120-21-4
2-Methylhexane	591-76-4	Dodecane	112-40-3
1-Butanol	71-36-3	Tridecane	629-50-5
2-Pentanone	107-87-9	Tetradecane	629-59-4
3-Methyl-heptane	589-81-1	Pentadecane	629-62-9
Acetoin	513-86-0	Camphene	79-92-5
1,2-Propanediol	57-55-6	2-Methylfuran	534-22-5
Cyclopentanone	120-92-3	Benzene	71-43-2
Isononane	3221-61-2	3-(Methylthio)-1-propene	10152-76-8
2-Heptanone	110-43-0	Dimethyl disulfide	624-92-0
6-Methyl-5-heptene-2-one	110-93-0	Methyl thiocyanate	556-64-9
γ-Terpinene	99-85-4	Toluene	108-88-3
2-Nonanone	821-55-6	3-Methylthiophene	616-44-4
Geranyl acetone	3796-70-1	3-Furaldehyde	498-60-2
Pentane	109-66-0	Ethylbenzene	100-41-4
Propanal	123-38-6	p-Xylene	106-42-3
Methacrolein	78-85-3	Furfural	98-01-1
Heptane	142-82-5	o-Xylene	108-38-3
Propanoic acid	79-09-4	Cumene	98-82-8
Octane	111-65-9	Dimethyl trisulfide	3658-80-8
Isobutyric acid	79-31-2	Benzaldehyde	100-52-7
Hexanal	66-25-1	Benzonitrile	100-47-0
Butyric acid	107-92-6	3-Thiophenecarboxaldehyde	498-62-4
Nonane	111-84-2	2-Thiophenecarboxaldehyde	98-03-3
Heptanal	111-71-7	Phenol	108-95-2
Valeric acid	109-52-4	Acetophenone	98-86-2
Cyclohexanone	108-94-1	m-Cresol	108-39-4
Octanal	124-13-0	4-Ethylphenol	123-07-9
Hexanoic acid	142-62-1	D-Limonene	138-86-3
Nonanal	124-19-6		

Table S4. Abbreviations and full names of the 33 core volatile organic compounds (VOCs) selected for mouse breath analysis.

No.	Compound Name	No.	Compound Name
1	Acetic acid(AA)	2	Propionic acid(PA)
3	Dimethyl sulfide(DMS)	4	Isobutyric acid(IBA)
5	Butyric acid(BA)	6	Valeric acid(VA)
7	D-limonene(D-Lim)	8	Benzonitrile(BN)
9	Phenol(PhOH)	10	Geranyl acetone(GA)
11	Propionaldehyde(PPA)	12	Acetone(Ace)
13	Isopropanol(IPA)	14	Methacrolein(MACR)
15	Ethyl acetate(EA)	16	Tert pentanol(t-PeOH)
17	Butanol(BuOH)	18	2-pentanone(2-PN)
19	Acetoin(ACN)	20	Hexanal(HEX)
21	1,2-propanediol(1,2-PD)	22	2-heptanone (2-HP)
23	Heptanal(HPT)	24	Benzaldehyde(BZD)
25	Octanal aldehyde(OCT)	26	Nonanal(NON)
27	Decanal(DEC)	28	Decane(C10)
29	Undecane(C11)	30	Dodecane(C12)
31	Tridecane(C13)	32	Tetradecane(C14)
33	Pentadecane(C15)		

Table S5. Histological scoring criteria for colonic damage(0–3 points for each item, with a total score of 0–15)

Criteria	Observation	Score 0	Score 1	Score 2	Score 3
Inflammatory cell infiltration	Target Cells: Lymphocytes (small round nuclei, dark staining), Neutrophils (multi-lobed nuclei, pale pink cytoplasm), Macrophages (large, vacuolated cytoplasm).	None or rare	Approx. 5–20 cells	Approx. 20–50 cells	>50 cells or large clusters
Mucosal / Crypt damage	Focus: Epithelial integrity and disruption or loss of crypt apices.	Normal	Mild epithelial loss or slight disruption of crypt apices	Moderate crypt destruction or focal extensive epithelial loss	Extensive epithelial denudation, ulceration, or extensive crypt destruction
Crypt architectural distortion	Focus: Regularity of crypt arrangement and morphology.	Regular, U-shaped, or columnar	Mild distortion or discontinuity	Significant shortening, bending, or blind-ending formation	Severe loss of structure, abundant blind sacs, and abnormal dilation
Submucosal edema / Hemorrhage	Focus: Edema (widened connective tissue spaces/vacuoles) and Hemorrhage (scattered red blood cells or deposits, extravascular).	No significant edema or hemorrhage	Mild vacuolization or scattered red spots	Moderate vacuolization with few bleeding spots	Extensive vacuolization/spacing with significant hemorrhagic deposition
Muscularis and serosal involvement	Focus: Inflammatory infiltration, edema, or fibrosis-like changes in the muscle layer.	No involvement	Rare, scattered at the margin of the muscle layer	Multiple infiltrations or marked vacuolization / edema	Extensive infiltration / edema, blurred muscle fiber interfaces, or fibrosis (pink collagen deposition)

Table S6. Qualitative and semi-quantitative comparison of the FaunaScope with conventional breath/gas sampling methods.

Feature/Parameter	FaunaScope	Tedlar Bags	Conventional Animal Exposure Chambers
Primary Material	Poly(methyl methacrylate），PMMA	Polyvinyl fluoride (PVF) film	Polycarbonate / Acrylic
Background Emissions	Minimal	High	Moderate to High
Dead Volume	Minimized (~ 600 mL)	Variable (usually 1–5 L)	Massive (Typically ＞ 3 L)
Sample Dilution / SNR	Low dilution / High SNR	Moderate	Severe dilution / Low SNR
References	N/A	[1], [2]	[3]

Supplementary References:

[1] K. Westphal, D. Dudzik, M. Waszczuk-Jankowska, B. Graff, K. Narkiewicz and M. Markuszewski, Metabolites, 2023, 13.

[2] J. Kwak, M. Fan, J. A. Martin, D. K. Ott and C. C. Grigsby, Anal Sci, 2017, 33, 147–152.

[3] K. Hintzen, A. Smolinska, A. Mommers, N. Bouvy, F. van Schooten and T. Lubbers, Journal of Breath Research, 2022, 16.

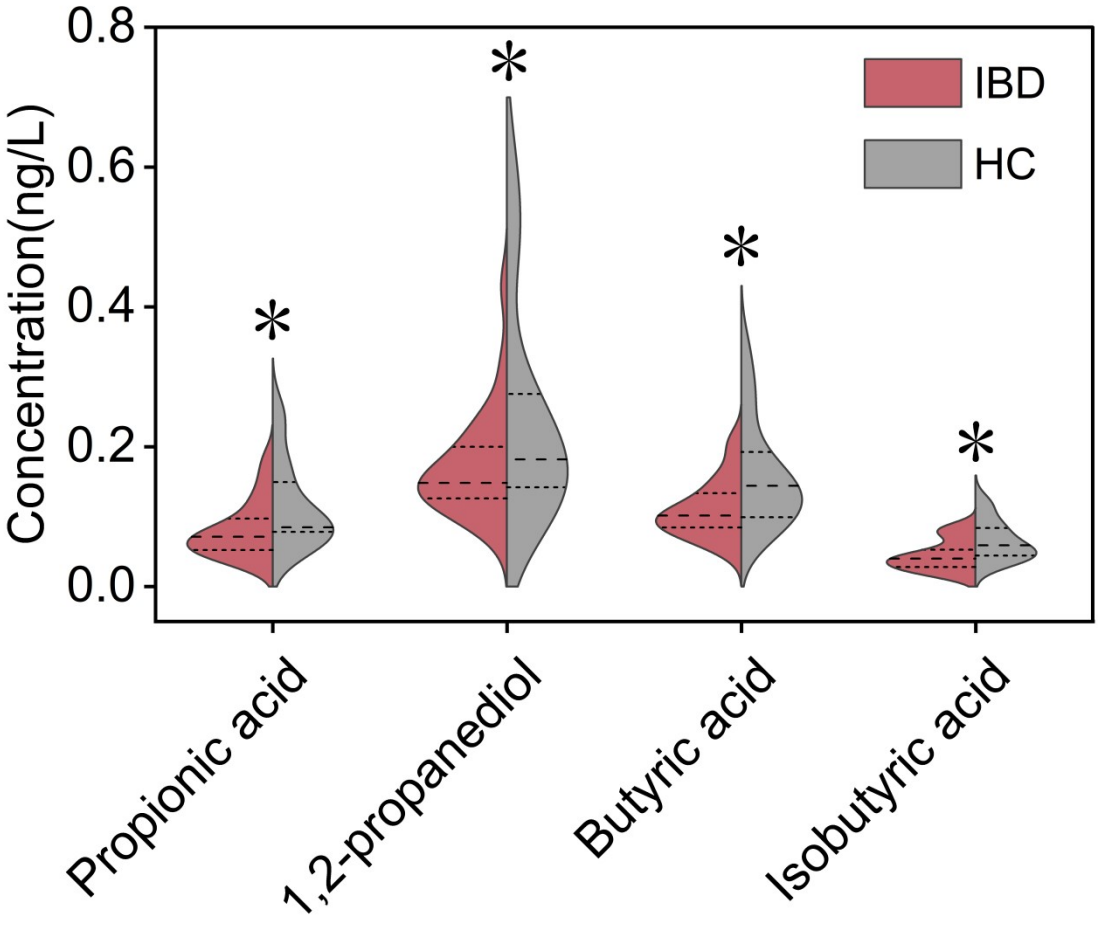

Fig. S1. **Comparison of short-chain fatty acid (SCFA) concentrations in urine between IBD patients and healthy controls (HC).** The violin plot shows the urinary concentration distribution of four key metabolic biomarkers: propionic acid,1,2-propanediol, butyric acid, and isobutyric acid between the two groups. The results indicate that, compared to the healthy control group, the concentrations of these four SCFAs and their derivatives are significantly lower in the urine of IBD patients. This trend is highly consistent with the results from breath samples, further confirming the reliability of these biomarkers. (* indicates $p < 0.05$,).

Figure S2. Macroscopic and histological characteristics of the colon in DSS- and EEN-treated mice. A, Colon length measurement shows significant shortening in DSS mice, while EEN-treated mice maintain near-normal colon length. B, Histological scores indicate that EEN intervention markedly reduces microscopic colonic damage. Data are presented as mean ± SD.

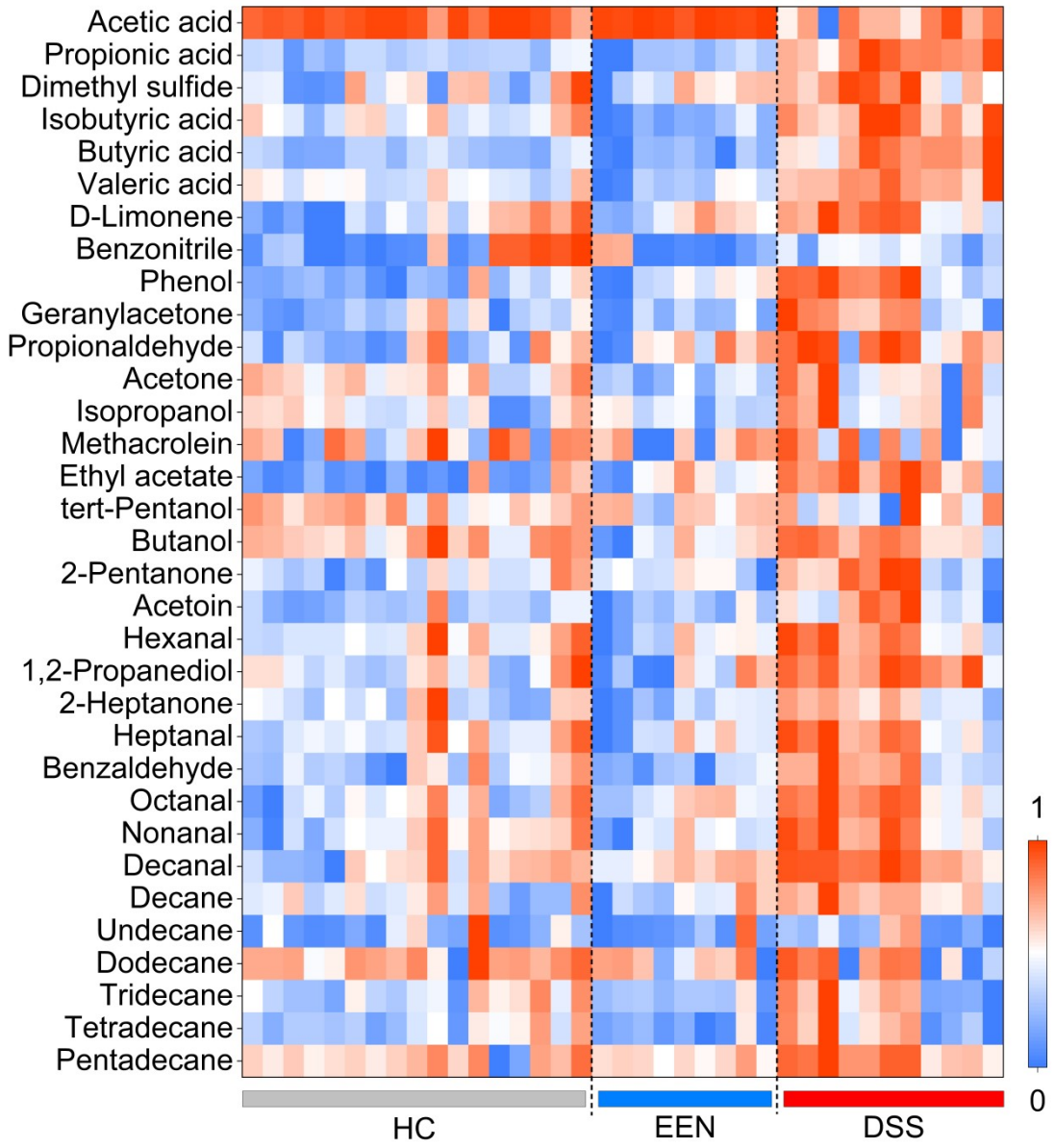

Figure S3. Heatmap of the full spectrum of breath metabolites in HC, DSS, and EEN mice. This heatmap displays the relative abundance of all detected volatile organic compounds in the breath of mice from HC group, DSS group, and EEN group. The figure clearly reveals that DSS-induced colitis caused extensive and significant disturbances in the overall breath metabolic profile of the mice.