

RESEARCH LETTERS

Diagnostic Performance of a Noninvasive Breath Test for Colorectal Cancer: COBRA1 Study



Colorectal cancer (CRC) is the third most common cancer globally.¹ When diagnosed early, the 5-year survival rate is 92%,² yet 23% of CRCs are diagnosed at an advanced stage³ in the United Kingdom with a 5-year survival rate of 10%.² Early CRC has symptoms that are shared with common benign conditions.⁴ Colonoscopy capacity is limited, and referring all symptomatic patients for colonoscopy would overwhelm available resources. An intermediate triage test to identify patients at risk of CRC could streamline referral pathways. A breath test based on detecting volatile organic compounds (VOCs) has the ideal characteristics for a triage tool because it is noninvasive, simple to undertake, and acceptable to patients of all ethnicities.⁵

COBRA1 is a prospective, multicenter diagnostic study aimed to develop a breath test to detect CRC (Research Ethics Committee no. 17/EE01/12; Bowel Cancer Screening Program identification 189; clinicaltrials.gov identifier NCT03699163). The target sample size was 1463 patients (117 CRC, 1346 control subjects) using the power diagnostic test function from the MKmisc R package, considering type I error (alpha) of 0.05, power (1-beta) of 0.8, prevalence value of 0.08, and assumed difference of 0.1. The study recruited patients aged 18–90 years attending 7 London hospitals (Appendix 1) for a bowel cancer screening program colonoscopy because of a positive fecal occult blood test (n = 664), colonoscopy for other indications (n = 645), or surgical colorectal adenocarcinoma resection (n = 123) (June 2017 to February 2020). We recruited a mixed population to ensure an adequate number of cancer cases for model development. The reference test was colonoscopy ± histopathology. Exclusion criteria was concurrent chemotherapy.

Breath was collected by trained research nurses using the ReCIVA breath sampling device (Owlstone Medical Ltd, Cambridge, UK) onto thermal desorption tubes using standardized settings immediately before colonoscopy and surgery.⁶ Patients fasted for 4 hours before breath collection. Quality control measures were performed for breath collection and analysis (Appendix 2). Thermal desorption tubes were couriered to the VOC laboratory (Imperial College London) for same-day analysis using gas chromatography-mass spectrometry (Agilent Technologies, Cheshire, UK) equipped with a midpolar column (ZB-62, 460 m × 0.25 mm inner diameter × 1.40 μm df; Phenomenex Inc, Torrance, CA). When same-day analysis was not possible, thermal desorption tubes were stored at –80°C for subsequent analysis. Laboratory staff were blinded to disease status. Gas chromatography-mass spectrometry data were acquired using MassHunter software (B.07 SP1; Agilent Technologies) and processed using the custom-designed spectral deconvolution tool MSHub.⁷ Machine learning pipelines identified predictive features (VOCs and

clinical metadata) to develop a multivariate discriminant analysis model and receiver-operating characteristic (ROC) curves (Appendix 3).

Included in the analysis were 1432 patients (828 men), with a median age of 66.5 years (range, 18–90). No adverse events were reported. Of 1432 patients, 357 had a normal colonoscopy; 188 had benign pathology (hemorrhoids or diverticular disease); 106 had inflammatory bowel disease; 348, 67, and 204 patients had low-, intermediate- and high-risk polyps, respectively; and 162 had colorectal adenocarcinoma. Polyp risk stratification was based on 2002 British Society of Gastroenterology (BSG) polyp surveillance guidelines,⁸ adapted to include dysplasia status, and the BSG 2017 guidance on serrated polyps.⁹

Patient demographics, cancer characteristics, and exclusion details are presented in Supplementary Table 1. CRC patients were older than control subjects and more likely to have had previous CRC or heart disease or used laxatives, antibiotics, or anticoagulants. CRCs were recruited from surgical lists (n = 119), bowel cancer screening programs (n = 30), and other colonoscopy lists (n = 13). Of the CRCs, 64.2% were T3 or T4. Of 1432 patients, 855 reported at least 1 symptom at the time of breath sampling.

Across all breath samples, 1024 VOC product ions were detected. The top 99 most predictive features (97 VOCs, body mass index, and age) according to Random Forest Scores were annotated and identified using mass spectral libraries (NIST, version 2.0).¹⁰ We assessed the origin of VOCs using the Human Metabolome Database 2018.¹¹ Thirty-five VOCs were deemed likely to be exogenous and 37 were of unknown identity, leaving 25 endogenous VOCs for further statistical analysis (Supplementary Table 2). A diagnostic model comparing all CRC (n = 162) and non-CRC patients (n = 1270) based on 14 endogenous VOCs and body mass index predicted CRC with area under the ROC curve of 0.87, sensitivity of 79%, specificity of 86%, and negative predictive value of 97% (Figure 1A). A model using data from symptomatic patients only, taken from the same cohort (CRC, n = 146; non-CRC, n = 709), predicted CRC with an area under the ROC curve of 0.91, sensitivity of 83%, specificity of 88%, and negative predictive value of 96% (Figure 1B). Predictive VOCs were largely from the alkane-, alcohol-, ester-, and sulfur-containing chemical groups. Higher levels of dimethyl

Abbreviations used in this paper: CRC, colorectal cancer; ROC, receiver-operating characteristic; VOC, volatile organic compound.

Most current article

© 2022 The Author(s). Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

0016-5085

<https://doi.org/10.1053/j.gastro.2022.06.084>

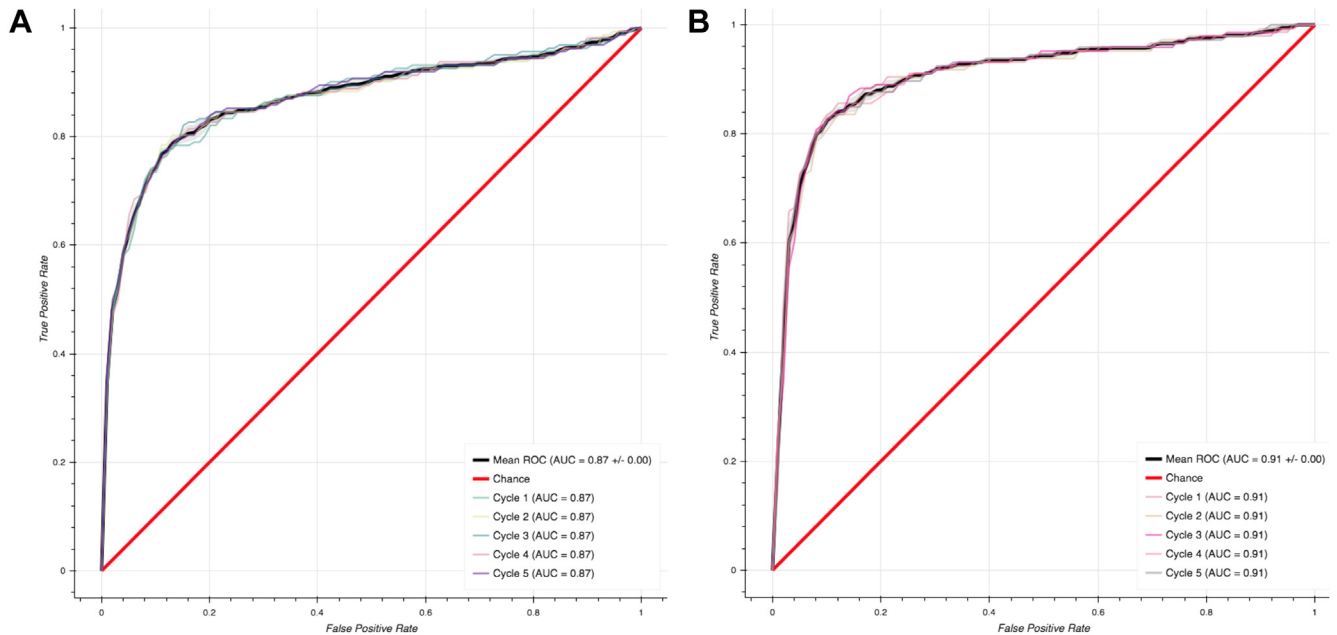


Figure 1. ROC curves illustrating the predictive power of a model based on 15 features (14 endogenous VOCs and body mass index), using the support vector machine (SVM) method, in determining (A) CRC (n = 162) from non-CRC (n = 1270) and (B) symptomatic CRC (n = 146) from symptomatic non-CRC (n = 709). AUC, area under the curve.

sulfide and 2-ethoxypropane discriminated right-sided (cecum to transverse colon) from left-sided tumors ($P = .002$ and $P = .045$, respectively). Polyps of any risk category (n = 619) could be predicted with an area under the ROC curve of 0.67, sensitivity of 66%, and specificity of 58% when compared with patients with no polyps and no CRC (n = 651) based on 16 endogenous VOCs and age. A model based on high-risk polyps did not improve prediction.

An acknowledged limitation of this study is the use of a selected population including patients known to have CRC to recruit enough cancer cases for model development, resulting in differences in age, medication use, and bowel preparation. It precluded symptom-based analysis (although this was not a study aim). Bowel preparation and age were not predictive features or confounding factors in the VOC-based model. Use of medications could not be examined in the machine learning analysis.

COBRA1 achieved its aim of constructing a diagnostic model of a VOC-based breath test to detect CRC, with promising results. The high negative predictive value of the breath test suggests the possibility of use as a triage tool. The strength of this study lies in its multicenter design, large sample size, comprehensive quality control measures, and selecting only endogenous VOCs for model development. The feasibility of multicenter breath collection with centralized sample analysis is also demonstrated. These results support further evaluation of this technology for detecting CRC in an unselected screening-eligible population for CRC screening, either alone or in combination with other tests, such as the fecal immunochemical test.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org and at <https://doi.org/10.1053/j.gastro.2022.06.084>.

GEORGIA WOODFIELD*

Department of Surgery and Cancer
Imperial College London
London, United Kingdom

ILARIA BELLUOMO*

Department of Surgery and Cancer
Imperial College London
London, United Kingdom

IVAN LAPONOGOV

Department of Surgery and Cancer
Imperial College London
London, United Kingdom

KIRILL VESELKOV

Department of Surgery and Cancer
Imperial College London
London, United Kingdom

COBRA1 WORKING GROUP

AMANDA J. CROSS

Department of Surgery and Cancer
Imperial College London
London, United Kingdom

GEORGE B. HANNA

Department of Surgery and Cancer
Imperial College London
London, United Kingdom

References

1. Ferlay J, et al. <https://www.who.int/news-room/fact-sheets/detail/cancer> and <https://gco.iarc.fr/today>.
2. Cancer Research UK. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#heading-Three>.
3. Cancer Research UK. <https://crukcanerintelligence.shinyapps.io/EarlyDiagnosis/>.
4. Adelstein B-A, et al. *BMC Gastroenterol* 2011;11:65.
5. Woodfield G, et al. *BMJ Open* 2021;11:e044691.
6. Doran SLF, et al. *J Breath Res* 2017;12:016007.
7. Aksenov AA, et al. *Nat Biotechnol* 2021;39:169–173.
8. Atkin WS, et al. *Gut* 2002;51(Suppl 5):v6–v9.
9. East JE, Atkin WS, Bateman AC, et al. *Gut* 2017;66:1181–1196.
10. Shen VK, et al. NIST Standard Reference Simulation Website. <https://www.nist.gov/programs-projects/nist-standard-reference-simulation-website>.
11. Wishart DS, et al. *Nucleic Acids Res* 2018;46:D608–D617.

*Authors share co-first authorship.

Correspondence

Address correspondence to: George B. Hanna, PhD, Department of Surgery and Cancer, 1st Floor, B Block, Hammersmith Hospital, London, W12 0HS, United Kingdom. e-mail: g.hanna@imperial.ac.uk.

Acknowledgments

We thank the National Institute for Health Research–affiliated breath testing teams at St Mark’s, Charing Cross, St George’s, Homerton, West Middlesex, Chelsea and Westminster, and St Mary’s Hospitals for their patient recruitment and sample collection across 7 sites for this study.

The full study protocol has been provided to the journal as a separate file and can be provided on request. Individual deidentified participant data will not be shared but are available on request to authors.

The COBRA1 Working Group includes Piers R. Boshier,¹ GengPing Lin,¹ Antonis Myridakis,¹ Oscar Ayrton,¹ Patrik Španěl,^{1,2} Alberto Vidal-Diez,¹ Andrea Romano,¹ John Martin,³ Laura Marelli,⁴ Chris Groves,⁵ Kevin Monahan,^{1,6} Christos Kontovounisios,^{1,7} Brian P. Saunders,^{1,8} from the ¹Department of Surgery and Cancer, Imperial College London, London, United Kingdom; ²J. Heyrovský Institute of Physical Chemistry of the Czech Academy of Sciences, Prague, Czech Republic; ³Department of Gastroenterology, Charing Cross Hospital, London, United Kingdom; ⁴Department of Gastroenterology, Homerton University Hospital, London, United Kingdom; ⁵Department of Gastroenterology, St George’s Hospital, London, United Kingdom; ⁶Department of Gastroenterology, West Middlesex University Hospital, London, United Kingdom; ⁷Department of Surgery,

Chelsea and Westminster Hospital, London, United Kingdom; and ⁸Department of Gastroenterology, St Mark’s Hospital and Academic Institute, London, United Kingdom.

CRedit Authorship Contributions

Georgia Woodfield, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Supporting; Funding acquisition: Equal; Investigation: Equal; Methodology: Equal; Project administration: Equal; Writing – original draft: Equal; Writing – review & editing: Equal). Ilaria Belluomo, PhD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Funding acquisition: Supporting; Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Supporting; Supervision: Equal; Writing – original draft: Supporting; Writing – review & editing: Equal). Piers R Boshier, PhD (Conceptualization: Supporting; Formal analysis: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Supporting; Supervision: Supporting; Writing – original draft: Supporting; Writing – review & editing: Equal). GengPing Lin, MD (Formal analysis: Supporting; Investigation: Equal; Methodology: Supporting; Project administration: Equal; Writing – review & editing: Supporting). Antonis Myridakis, PhD (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Software: Supporting; Supervision: Supporting; Visualization: Supporting; Writing – review & editing: Supporting). Oscar Ayrton, MSc (Investigation: Supporting; Methodology: Supporting; Project administration: Supporting; Resources: Supporting; Software: Supporting; Writing –review & editing: Supporting). Ivan Laponogov, PhD (Data curation: Supporting; Formal analysis: Lead; Methodology: Supporting; Software: Lead; Supervision: Supporting; Validation: Supporting; Writing –review & editing: Supporting). Kirill Veselkov, PhD (Data curation: Supporting; Formal analysis: Lead; Methodology: Supporting; Software: Lead; Supervision: Supporting; Validation: Supporting; Writing –review & editing: Supporting). Patrik Španěl, Dr. rer. nat (Data curation: Equal; Formal analysis: Equal; Investigation: Supporting; Methodology: Supporting; Resources: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Alberto Vidal-Diez, PhD (Data curation: Supporting; Formal analysis: Supporting; Methodology: Supporting; Software: Supporting; Writing – review & editing: Supporting). Andrea Romano, PhD (Formal analysis: Supporting; Investigation: Supporting; Software: Supporting; Writing – review & editing: Supporting). John Martin, MD (Project administration: Supporting; Resources: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Laura Marelli, MD (Project administration: Supporting; Resources: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Chris Groves, MD (Project administration: Supporting; Resources: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Kevin Monahan, PhD (Project administration: Supporting; Resources: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Christos Kontovounisios, PhD (Project administration: Supporting; Resources: Supporting; Supervision: Supporting; Writing – review & editing: Supporting). Brian P Saunders, MD (Supervision: Supporting; Visualization: Supporting; Writing –original draft: Supporting; Writing – review & editing: Supporting). Amanda J Cross, PhD (Conceptualization: Supporting; Formal analysis: Supporting; Methodology: Supporting; Resources: Supporting; Supervision: Supporting; Writing – original draft: Supporting; Writing – review & editing: Equal). George B Hanna, PhD (Conceptualization: Lead; Data curation: Supporting; Formal analysis: Equal; Funding acquisition: Lead; Investigation: Equal; Methodology: Lead; Project administration: Supporting; Resources: Equal; Software: Equal; Supervision: Lead; Writing – original draft: Supporting; Writing – review & editing: Equal).

Conflicts of Interest

These authors disclose the following: George B. Hanna is founder of a cancer diagnostic company. Georgia Woodfield, Ilaria Belluomo, Piers R. Boshier, and George B. Hanna are named on patents related to volatile biomarkers. Brian P. Saunders has received research funding, speaker fees, and loan equipment from Olympus and speaker fees and loan equipment from Fuji. The remaining authors disclose no conflicts.

Funding

This research was supported by grants from the Rosetrees and Stonegate Trusts and Imperial Health Charity and funding originating from the Imperial Biomedical Research Center, HCA Healthcare UK and by infrastructure funding from the National Institute for Health Research In Vitro Diagnostics Co-operative London at Imperial College Healthcare NHS Trust. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Appendix 1: Recruitment

Patients attending for colonoscopy were recruited from Charing Cross, St Mark's, St George's and Homerton Hospitals in London. Surgical patients were recruited from St Mary's, West Middlesex and Chelsea, and Westminster Hospitals in London. Breath samples were obtained on the same day before colonoscopy or elective surgery. Polyps were stratified into low-risk (1–2 subcentimeter tubular adenomas with low-grade dysplasia or subcentimeter serrated polyps without dysplasia), intermediate-risk (3–4 subcentimeter or one >1-cm tubular adenoma with low-grade dysplasia or >1-cm serrated polyps without dysplasia), and high-risk (≥ 5 subcentimeter adenomas, ≥ 3 adenomas if 1 was >1 cm, or any adenoma with high-grade dysplasia or villous change or any serrated polyp with dysplasia) categories.

Appendix 2: Quality Control

Quality Control for Cleaning Thermal Desorption Tubes

After conditioning (TC20; Markes Ltd, Llantrisant, UK), 30 tubes were randomly selected and assessed using proton transfer reaction time of flight mass spectrometry. If any VOC abundance was >1 ppb, tubes were rechecked and excluded if concentrations remained >1.5 ppb.

Quality Control for ReCIVA Breath Collection

We assessed VOC contamination in the CASPER filtration system (Owlstone Medical) every 3 months and replaced the filter if contamination was detected or after 450 hours of use; the flow rate by each pump every 4 weeks using a flowmeter to test the flow through an empty thermal desorption (TD) tube attached to each of tube inlets to ensure the intended 200 mL/min flow rate for each tube position; and recorded breath collection parameters by ReCIVA software for all breath samples (temperature, flow rates of both pumps, and pressure of the facemask against the patient's face as a surrogate for a good seal). We interrogated the h5 files generated by the ReCIVA software using an in-house generated script written with R programming language¹ to create a graphic representation of quality control parameters. We visually assessed all outputs to exclude inadequate samples.

Quality Control for VOC Presence in TD Tubes

We used a threshold-based system to quantify acetone ($m/z = 58$, $RT = 8.97$) within each TD tube using gas chromatography-mass spectrometry, indicating the presence of enough breath sample for analysis. Acetone was selected as the reference compound because it is always present in human breath.² Samples with an acetone abundance of <4,000,000 area (raw gas chromatography-mass spectrometry data) were excluded. We identified this

threshold by comparing 121 breath samples (500 mL) collected using identical protocol and analytical parameters as COBRA1 against 152 nonbiologic control subjects: empty conditioned TD tubes; 500-mL room air samples collected onto TD tubes using ReCIVA, following a similar procedure to that for patient breath; and TD tubes, previously conditioned and then loaded with a standard mixture not containing any of the tested compounds. All but 5 of 121 breath samples analyzed had a breath acetone level above the identified threshold (Mann-Whitney U test, $P < .0001$).

Quality Control for Gas Chromatography-Mass Spectrometry Analysis

We used retention time, peak shape, and peak area to assess consistency and accuracy of instrument analysis of 5 TD tubes loaded with a certified standard mixture using a permeation unit (ES 4050P; Eco Scientific, Gloucestershire UK).³ The standard mixture consisted of benzene (63 ppb), phenol (90 ppb), butyric acid (20 ppb), pentanoic acid (5 ppb), hexanoic acid (5 ppb), decanal (4 ppb), and butanal (5 ppb), maintained at 30°C and with a nitrogen flow of 0.9 L/min.

Supplementary References

1. R Core Team. <https://www.r-project.org>.
2. de Lacy Costello B, et al. *J Breath Res* 2014;8:014001.
3. Romano A, et al. *Analytic Chem* 2018;90:10204–10210.

Appendix 3: Data Processing

A machine learning pipeline using Python¹ and Scikit-learn² processed all data (up to 1024 chemical ions per breath sample and metadata for each patient) and identified discriminatory “features” of CRC and non-CRC using both analysis of variance and random forest-based scores. The data were normalized, variance stabilized, and log transformed as part of the machine learning pipeline. Random forest, alphanet, support vector machine (SVM), lasso, and elastic machine learning prediction methods were used independently to compare every combination and permutation of pathology group. The same analyses were repeated for patients aged 40–59, 45–65, 50–69, and 70–89 years and all ages together to investigate whether age was confounding VOC data. The prediction models considered patient-related factors between groups (age, number of hours of fasting, body mass index, ethnic origin, gender, smoking status, weekly alcohol consumption, type of bowel preparation taken before colonoscopy/surgical resection, and family history of CRC) and sampling-related factors (storage time of the TD tube from conditioning to breath sampling, storage time after-sampling until mass spectrometry analysis, and number of days the TD tube was stored in the freezer [if applicable]). Medications were not inputs into the model, because answers were too numerous and heterogeneous; if every answer was coded, “missing data” would have been very common and would have

introduced bias, whereas only sporadic mention of a variable would have added unnecessary noise to the dataset and not have been useful.

ROC curves were used to determine the accuracy of the diagnostic test in classifying those with and without colorectal disease. The ROC curves were generated based on 25 runs: 5 repeats of 5-fold stratified K-fold splits with reshuffling between splits. This meant that samples were shuffled and then split into 5 groups. Each group was then used in turn as a test set, whereas the other 4 were the training set. Feature selection and model building (machine learning) were performed on a training set each time (80% of the data) and then applied to the test set (20% of the data) to produce the statistics. This was repeated 5 times, and then the results from different runs were averaged to get ROC curves and error estimates. This way avoided

overfitting of the model and any bias of an “unlucky” split if the data happened to have been split in a nonrepresentative way. Features were not weighted in any way but filtered to leave the most discriminating features in each run. If a feature was independently selected to be a differentiating feature regardless of how the data was split, the selection score would be higher. A higher score therefore meant that the feature in question was more likely to be a true feature differentiating CRC and non-CRC.

Supplementary References

1. Rossum GV, et al. Python 3 reference manual. Scotts Valley, CA: CreateSpace; 2009.
2. [Fabian Pedregosa GV, et al. J Mach Learn Res 2011; 12:2825–2830.](#)

Supplementary Table 1.Detailed Patient Demographics of CRC and Control Groups

Parameter	CRC patients only (n = 162)	All control subjects (non-CRC) (n = 1270)	<i>P</i> between CRC and control patients
No. of patients approached for study	2334 (166 from theatre, 2168 from endoscopy)		
No. of patients declined to participate	73 (17 from theatre, 56 from endoscopy)		
No. of patients consented but not breath sampled (research nurse judged there was insufficient time preprocedure)	406 (10 from theatre, 396 from endoscopy)		
No. of patients who had breath sampled	1855 (139 from theatre, 1716 from endoscopy)		
No. of patients excluded due to inadequate reference test (incomplete or canceled colonoscopy)	65		
No. of patients excluded before laboratory analysis due to inadequate breath sample (wrong settings used, illegible labeling, human error: tubes not sealed properly)	61		
No. of patients excluded due to failure of ReCIVA quality control (poor breath trace)	50		
No. of patients excluded as samples lost during gas chromatography-mass spectroscopy analysis (human error, instrument fault)	81		
No. of patients excluded due to failure of quality control for VOC presence in TD tubes	166		
No. of patients with high-quality breath samples and reference test and therefore included in the final analysis	1432		
Gender			
Female	61 (37.7)	538 (42.4)	.360
Male	101 (62.3)	727 (57.2)	
Unrecorded	0	5 (0.4)	
Age, y			
Median (IQR)	66.5 (17)	63 (14)	<.001
Minimum to maximum	30–90	18–87	
Unrecorded	0	7	
Body mass index, kg/m ²			
Median (IQR)	26 (7)	26 (8)	.674
Minimum to maximum	18–41	14–48	
Unrecorded	36	661	
Ethnicity			
Arab	5 (3.1)	40 (3.1)	.568
Asian/Asian British	22 (13.6)	198 (15.6)	
Black/African/Caribbean/Black British	19 (11.7)	106 (8.3)	
Other ethnicity	8 (4.9)	38 (3.0)	
White British/European	104 (64.2)	858 (67.6)	
Unrecorded	4 (2.5)	30 (2.4)	
Smoking status			
Current	14 (8.6)	171 (13.5)	.080
Ex	45 (27.8)	397 (31.3)	
Never	102(63)	679 (53.5)	
Unrecorded	1 (0.6)	23 (1.8)	
Alcohol intake status			
Current	88 (54.4)	802 (63.1)	.107
Ex	11 (6.8)	62 (4.9)	
Never	61 (37.7)	380 (29.9)	
Unrecorded	2 (1.2)	26 (2)	

Supplementary Table 1. Continued

Parameter	CRC patients only (n = 162)	All control subjects (non-CRC) (n = 1270)	P between CRC and control patients
Bowel preparation			
Total patients receiving bowel preparation	124 (76.5)	1263 (99.4)	<.001
No bowel preparation	37 (22.8)	3 (0.2)	
Unrecorded	1 (0.6)	4 (0.3)	
Time fasted, h			
Minimum	4	7	<.001
Maximum	36	72	
Median (IQR)	22 (10)	24 (4)	
Unrecorded	46	189	
Reason for attendance			
Colonoscopy: CRC screening	30 (18.5)	634 (49.9)	<.001 (explained by deliberate enrichment of CRC patients from theatre)
Colonoscopy: rescope for polyp removal	1 (0.6)	26 (2.0)	
Colonoscopy: surveillance inflammatory bowel disease/polyps/family history	3 (1.8)	241 (19.0)	
Colonoscopy: symptoms, 2-week wait	5 (3.1)	91 (7.2)	
Colonoscopy: symptoms, urgent	4 (2.5)	123 (9.7)	
Colonoscopy: symptoms, routine	0	147 (11.6)	
Theatre resection patient for likely CRC	119 (73.5)	4 (0.3)	
Unknown reason	0 (0)	4 (0.3)	
Past medical history			
Previous CRC	13 (8.0)	33 (2.6)	<.001
Previous cancer excluding CRC	14 (8.6)	80 (6.3) ^a	.257
Celiac disease	1 (0.6)	2 (0.2)	.228
Past bowel resection	5 (3.1)	44 (3.5)	.803
Barrett's esophagus	0	6 (0.5)	.381
High blood pressure	57 (35.2)	434 (34.2)	.798
Known heart disease	28 (17.3)	132 (10.4)	.009
Diabetes	31 (19.1)	186 (14.6)	.133
Renal impairment	8 (4.9)	37 (2.9)	.164
Chronic obstructive pulmonary disease	6 (3.7)	41 (3.2)	.749
Asthma	14 (8.6)	89 (7.0)	.448
Liver impairment	3 (1.9)	48 (3.8)	.212
Medications			
Proton pump inhibitor	31 (19.1)	235 (18.5)	.846
Nonsteroidal anti-inflammatory drug	18 (11.1)	176 (13.9)	.336
Laxative	11 (6.8)	27 (2.1)	.002
Antibiotics	8 (4.9)	18 (1.4)	.002
Ranitidine	3 (1.9)	15 (1.2)	.471
Clopidogrel	5 (3.1)	24 (1.9)	.309
Blood thinner ^b	14 (8.6)	44 (3.5)	.002
Immunosuppressants ^c	10 (6.2)	54 (4.3)	.265
Patient-reported symptoms (n = 855: 146 CRC patients and 709 control subjects),			
Bowel symptoms ^d	67	409	.021
% of total study patients			
Weight loss	28	67	.264
Abdominal pain	40	152	.737
Rectal bleeding	55	232	.918
Other symptoms ^e	23	79	.170
Tumor site			
Left sided: rectum to splenic flexure	100 (61.7)	NA	NA
Right sided: transverse colon to cecum	62 (38.3)	NA	NA

Supplementary Table 1. Continued

Parameter	CRC patients only (n = 162)	All control subjects (non-CRC) (n = 1270)	P between CRC and control patients
Clinical tumor stage			
1	20 (12.3)	NA	NA
2	36 (22.2)	NA	NA
3	63 (38.9)	NA	NA
4	41 (25.3)	NA	NA
Missing data	2 (1.2)	NA	NA
Clinical nodal stage			
0	89 (54.9)	NA	NA
1	49 (30.2)	NA	NA
2	21 (13.0)	NA	NA
3	1 (0.6)	NA	NA
Missing data	2 (1.2)	NA	NA
Clinical metastasis stage			
0	138 (85.2)	NA	NA
1	22 (13.6)	NA	NA
Missing data	2 (1.2)	NA	NA
Differentiation of tumor			
Well differentiated	1 (0.6)	NA	NA
Moderately differentiated	129 (79.6)	NA	NA
Poorly differentiated	26 (16.0)	NA	NA
Missing data	6 (3.7)	NA	NA

Values are n (%) unless otherwise defined. NA, not applicable.

^aFourteen of these patients had >1 previous cancer.

^bBlood thinners included warfarin, therapeutic doses of low-molecular-weight heparin, and direct oral anticoagulants.

^cImmunosuppressants included steroids, biologics, and any other immunosuppressive disease-modifying agents. Patients were not taking chemotherapy at the time of breath testing.

^dDefinition of bowel symptoms was change in bowel habit, diarrhea, fecal urgency, or constipation.

^eDefinition of other symptoms was broad: dyspepsia, bloating, rectal pain, flatulence, rectal itching, and anemia. Note: The true number of anemic patients was unknown, because hemoglobin levels were not recorded.

Supplementary Table 2. Top 15 Features (Endogenous-only VOCs and Clinical Features) Capable of Differentiating CRC Patients From Control Patients in Model 1 (CRC vs Non-CRC [n = 1432, CRC = 162]) and Model 2 (CRC vs Non-CRC Symptomatic Patients Only [n = 855, CRC = 146])

Feature ^a	Retention time (min)	Proportion of deconvoluted peak (%)	Random forest score	Tentative identification	Chemical Abstracts Service registry number	Proposed explanation/mechanism
<i>Model 1: CRC (n = 162) vs non-CRC (n = 1270)</i>						
1	22.02	72	0.0102	Propyl ester of propanoic acid (propyl propionate)	106-36-5	Ester metabolism has previously been linked to cancer, where 2 of the most-studied human carboxylesterase enzymes (CES1 and CES2) are markedly altered in cancerous tissue (PMID 30245959). Esters in breath have previously been found to be predictive for CRC (PMID:26212114 and PMID: 24820062).
2	9.27	85	0.0100	Dimethyl sulfide	75-18-3	Dimethyl sulfide is a plausible biomarker, as it is produced by anaerobic bacteria in the gut, with disruption of the microbiome in CRC as the potential mechanism for its increased production. Fecal dimethyl sulfide within a panel of 4 other VOCs has also been shown to discriminate high-risk polyp patients from control subjects (PMID: 26086914). Bacteria are also known to metabolize branched-chain alkanes (PMID: 4852318, PMID: 24829093) found to be characteristic of CRC in the current study.
3	17.48	15	0.0036	1-Penten-3-ol	616-25-1	Increased alcohol abundance in CRC breath may be explained by the “Warburg” effect whereby glycolytic metabolism increases due to the increased proliferation characteristic of cancer cells. This leads to a shift toward anaerobic respiration of glucose. A wide variety of alcohols are endogenous and are found in the breath of humans, making this a plausible biomarker (PMID: 24421258). The human liver and gastrointestinal tract also contain alcohol dehydrogenase enzymes (PMID: 8244116), which would affect alcohol abundance in any blood returning from the gastrointestinal tract through the liver.
4	31.14	33	0.0038	3,4-Dimethyl- 1,5-cyclooctadiene	21284-05-9	Dienes were found in the breath of healthy humans (PMID: 24421258) and therefore could be a plausible biomarker to be altered in disease.

Supplementary Table 2. Continued

Feature ^a	Retention time (min)	Proportion of deconvoluted peak (%)	Random forest score	Tentative identification	Chemical Abstracts Service registry number	Proposed explanation/mechanism
5	NA	NA	0.0062	BMI, metadata	NA	High BMI is a known risk factor for development of CRC (Cancer Research UK data). However, late-stage cancers of all types can also lead to low BMI due to the increased catabolism associated with severe illness. With regards to breath specifically, BMI is known to affect breath acetone levels (PMID: 29516396, PMID: 21725144), although neither acetone nor any other ketone was found to be discriminatory for CRC in the machine learning prediction model.
6	32.69	58	0.0029	2-Propenyl ester of acetic acid (allyl acetate)	591-87-7	See ester explanation for Feature 1
7	40.12	88	0.0052	Branched tetradecane	Unknown	Many alkanes, methylated alkanes, and cycloalkanes have been named as discriminatory markers for CRC (PMID 26212621, PMID 26212114, PMID: 30796770). Alkanes may be present in breath due to oxidative stress and lipid peroxidation that occurs in cancer and inflammation. (PMID: 18465793) There is also a link with microbial activity and branched alkanes; bacteria are known to be able to metabolize branched-chain alkanes (PMID: 4852318, PMID: 24829093). Altered gut microbiome in CRC could therefore potentially alter the branched alkane abundance.
8	23.52	29	0.0036	Overlapping ester similar to 2-propenyl ester of acetic acid (allyl acetate)	591-87-7	See ester explanation for Feature 1
9	10.41	100	0.0013	2-Methyl- 2-propanol	75-65-0	See alcohol explanation for Feature 3
10	32.24	85	0.0019	4-Ethyl-1-octyn-3-ol (branched alcohol, molecular weight 130 or 152)	5877-42-9	See alcohol explanation for Feature 3
11	31.69	62	0.0027	2,2,4-Trimethyl-3-pentanol	5162-48-1	See alcohol explanation for Feature 3

Supplementary Table 2. Continued

Feature ^a	Retention time (min)	Proportion of deconvoluted peak (%)	Random forest score	Tentative identification	Chemical Abstracts Service registry number	Proposed explanation/mechanism
12	4.75	100	0.0039	Cyclopropane	75-19-4	Cyclopropane has been documented in the breath and feces of healthy humans (PMID: 24421258). Cyclopropane has also been found in higher concentrations in the breath of breast cancer patients compared with healthy control subjects (PMID: 21383471).
13	11.67	12	0.0042	2-Ethoxypropane	625-54-7	See alkane explanation for Feature 7
14	40.52	86	0.0040	2-Phenoxy-ethanol	122-99-6	See alcohol explanation for Feature 3
15	15.48	65	0.0003	Heptane	142-82-5	See alkane explanation for Feature 7
<i>Model 2: Symptomatic CRC (n = 146) vs symptomatic non-CRC (n = 709)</i>						
1	22.02	72	0.0193	Propyl ester of propanoic acid (propyl propionate)	106-36-5	See above explanation
2	9.27	85	0.0110	Dimethyl sulfide	75-18-3	See above explanation
3	17.48	15	0.0038	1-Penten-3-ol	616-25-1	See above explanation
4	40.12	88	0.0054	Branched tetradecane	629-59-4	See above explanation
5	31.14	33	0.0031	3,4-Dimethyl- 1,5-cyclooctadiene	21284-05-9	See above explanation
6	4.75	100	0.0040	Cyclopropane	75-19-4	See above explanation
7	23.52	29	0.0025	Overlapping ester similar to 2-propenyl ester of acetic acid (allyl acetate)	591-87-7	See above explanation
8	32.69	58	0.0008	2-Propenyl ester of acetic acid (allyl acetate)	591-87-7	See above explanation
9	NA	NA	0.0028	BMI, metadata	NA	See above explanation
10	40.52	86	0.0033	2-Phenoxy-ethanol	122-99-6	See above explanation
11	11.67	12	0.0040	2-Ethoxypropane	625-54-7	See above explanation
12	38.74	58	0.0014	Branched tridecane	629-50-5	See alkane explanation for Feature 7 under Model 1
13	31.69	62	0.0026	2,2,4-Trimethyl-3-pentanol	5162-48-1	See above explanation
14	32.24	85	0.0006	4-Ethyl-1-octyn-3-ol (branched alcohol, molecular weight 130 or 152)	5877-42-9	See above explanation
15	10.41	100	0.0009	2-Methyl- 2-propanol	75-65-0	See above explanation

Features are ranked according to random forest order of discriminatory value. Random forest feature scorings usually should add to 1 when all features that contributed to the model are included. However, the table demonstrates the random forest score for the top 15 most promising and realistic biomarker candidates only (endogenous only) and is the reason the random forest scores depicted here do not sum to 1. The numerical values of the scores can only be interpreted in relation to the other feature scores, where true important features should have high scores relative to the noise features.

^aFeatures are listed in order of importance to the predictive model.