Breath Metabolomics Provides an Accurate and Noninvasive Approach for Screening Cirrhosis, Primary, and Secondary Liver Tumors

Galen Miller-Atkins,1 Lou-Anne Acevedo-Moreno,2 David Grove,3 Raed A. Dweik,4 Adriano R. Tonelli,4 J. Mark Brown,5,6 Daniela S. Allende,7 Federico Aucejo,2 and Daniel M. Rotroff1

Hepatocellular carcinoma (HCC) and secondary liver tumors, such as colorectal cancer liver metastases are significant contributors to the overall burden of cancer-related mortality. Current biomarkers, such as alpha-fetoprotein (AFP) for HCC, result in too many false negatives, necessitating noninvasive approaches with improved sensitivity. Volatile organic compounds (VOCs) detected in the breath of patients can provide valuable insight into disease processes and can differentiate patients by disease status. Here, we investigate whether 22 VOCs from the breath of 296 patients can distinguish those with no liver disease (n = 54), cirrhosis (n = 30), HCC (n = 112), pulmonary hypertension (n = 49), or colorectal cancer liver metastases (n = 51). This work extends previous studies by evaluating the ability for VOC signatures to differentiate multiple diseases in a large cohort of patients. Pairwise disease comparisons demonstrated that most of the VOCs tested are present in significantly different relative abundances (false discovery rate \( P < 0.1 \)), indicating broad impacts on the breath metabolome across diseases. A predictive model developed using random forest machine learning and cross validation classified patients with 85% classification accuracy and 75% balanced accuracy. Importantly, the model detected HCC with 73% sensitivity compared with 53% for AFP in the same cohort. An added value of this approach is that influential VOCs in the predictive model may provide insight into disease etiology. Acetaldehyde and acetone, both of which have roles in tumor promotion, were considered important VOCs for differentiating disease groups in the predictive model and were increased in patients with cirrhosis and HCC compared to patients with no liver disease (false discovery rate \( P < 0.1 \)). Conclusion: The use of machine learning and breath VOCs shows promise as an approach to develop improved, noninvasive screening tools for chronic liver disease and primary and secondary liver tumors. (Hepatology Communications 2020;4:1041-1055).

Liver cancer is currently the fifth leading cause of cancer mortality among males, with an estimated 30,200 individuals succumbing annually in the United States.\(^{(1)}\) Hepatocellular carcinoma (HCC) accounts for 80% of all primary liver cancers, and the global incidence of HCC has tripled...
in the past 40 years due to nonalcoholic steatohepatitis (NASH), hepatitis C, and excessive alcohol consumption.\(^2\) In the United States, 2-year survival with HCC ranges from 30% to 44%.\(^{1,3}\) Early diagnosis and treatment is essential to curbing the high mortality associated with HCC and other liver tumors. Ultrasound imaging and measuring alpha-fetoprotein (AFP) are two commonly used approaches to screen for HCC; however, there are important limitations to both of these methodologies. While some patients with HCC may exhibit increased levels of AFP, it is estimated that 30%-50% of HCCs do not present with elevated AFP levels, resulting in too many false-negative diagnoses.\(^4\) Furthermore, AFP cannot differentiate HCC from other metastatic tumors that may also express AFP, such as germ cell tumors.\(^5\)

The liver remains a common site of metastasis for certain cancers such as colorectal cancer, also a leading cause of cancer deaths in the United States.\(^{1,6}\) Only 20%-30% of patients with colorectal cancer liver metastases (CRLM) are candidates for resection due to extrahepatic disease and other complicating factors.\(^6\) Therefore, it is critical that noninvasive, accurate, and cost-effective tools are developed that can diagnose chronic liver diseases, detect cancer in the liver, and track disease progression. In addition, the more accessible these tools are, the more they can be used to monitor signs of early disease development and response to treatment.

Metabolomics, the study of the biochemical products of metabolic processes, has shown promise in detecting metabolite biomarkers that can diagnose and predict disease, as well as assess treatment response.\(^7-10\) In addition, metabolomics has the ability to rapidly assess many metabolites from various noninvasive biospecimens (e.g., blood, saliva, breath, urine), which could provide biomarkers that are useful for screening diseases. Several metabolites have been found to be important in detecting oral cancers, cardiovascular disease, and predicting drug response.\(^11-13\) These studies indicate that metabolite profiles can detect disease from a variety of different biospecimens. This may be particularly relevant for HCC and other liver diseases in which metabolic alternations from liver damage may be reflected in detectable biochemical changes. Several studies have demonstrated the ability to discriminate people with healthy livers from those with liver disease(s) through principal component analysis (PCA), multiple logistic regression, or machine learning approaches.\(^14-17\) For example, Fitian et al. compared random forest, multivariate statistics, and other methodologies to identify and predict HCC in patients with hepatitis C virus and cirrhosis,\(^18\) and Liu et al.\(^19\) used a random forest model to predict HCC from serum and achieved 100% sensitivity despite some AFP values lower than 20 ng/mL. Breath metabolomics has also shown promising results for detecting colon cancer, breast cancer, infectious disease, asthma, and others.\(^19-25\)

Here, we present a model developed from breath-based metabolites (volatile organic compounds [VOCs]) to classify patients as being healthy, having pulmonary hypertension (as a disease control), cirrhosis, HCC, or CRLM in a large and well-characterized cohort of patients with liver disease. We compare and contrast our model’s prediction accuracy for detecting HCC to the current clinical standard, AFP, and examine the potential for breath metabolomics as a noninvasive screening tool for detecting liver diseases. Our hypothesis is that the combination of

**ARTICLE INFORMATION:**

From the \(^1\)Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH; \(^2\)Department of General Surgery, Cleveland Clinic, Cleveland, OH; \(^3\)Department of Inflammation and Immunity, Lerner Research Institute, Cleveland Clinic, Cleveland, OH; \(^4\)Respiratory Institute, Cleveland Clinic, Cleveland, OH; \(^5\)Department of Cardiovascular and Metabolic Sciences, Cleveland Clinic, Cleveland, OH; \(^6\)Department of Pathology, Cleveland Clinic, Cleveland, OH.

**ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:**

Daniel M. Rotroff, Ph.D., M.S.P.H.
Department of Quantitative Health Sciences
Lerner Research Institute, Cleveland Clinic
9500 Euclid Avenue, JJN3-01
Cleveland, OH 44195
E-mail: rotrofd@ccf.org
Tel.: +1-216-444-3399
metabolomics and machine learning can be a powerful tool to screen individuals for liver disease and liver cancers with high sensitivity and specificity.

Materials and Methods

STUDY PARTICIPANTS

Breath samples were collected from a total of 296 patients seen at the Cleveland Clinic (Cleveland, Ohio). Eligible participants were adult patients (>18 years of age) who underwent liver transplantation for HCC, surgical resection for liver tumors, or liver biopsy. In addition, we included patients with pulmonary hypertension who underwent right heart catheterization, patients attending treatment for hernia without history of liver disease or liver cancer (healthy control subjects), or family relatives without history of liver disease or liver cancer (for healthy control subjects). Importantly, initial disease diagnoses, made from clinical presentation, imaging and laboratory techniques, underwent a secondary confirmatory pathological diagnosis after tumor resection, rendering accuracy to initial disease classifications. Written informed consent was provided by all participants, and the study was approved by the Cleveland Clinic internal review board (IRB #10-347).

BREATH SAMPLE PROCUREMENT AND PROCESSING

Some subjects were nil per os for 2-8 hours prior to breath collection; however, no statistically significant differences in VOC profiles between fasting and nonfasting individuals were observed (data not shown). All subjects had a mouth rinse with tap water immediately before obtaining the breath sample to eliminate contamination from oral bacteria. Subjects were encouraged to exhale followed by inhalation to total lung capacity through a disposable mouth filter. The inhaled ambient air was filtered through a N7500-2 acid gas cartridge (North Safety Products, Smithfield, RI). The subjects then exhaled into a Mylar bag (Convertidora Industrial, Jalisco, Mexico) at a steady flow rate. Breath samples were analyzed within 2 hours of collection after incubation to 37°C for 10 minutes using the Selective Ion Flow Tube Mass Spectrometer (SIFT-MS) (Syft Technologies Ltd., Christchurch, New Zealand).

SIFT-MS

SIFT-MS works by creating reagent ions like H3O+, NO+, and O2+ in a microwave source. The reagent ions are selected individually by a quadrupole mass analyzer, and then ionize individual gases of a complex gaseous mixture like the breath. Here, we measured the relative concentrations of 22 VOCs in exhaled breath: 2-propanol, acetaldehyde, acetone, acetonitrile, acrylonitrile, benzene, carbon disulphide, dimethyl sulphide, ethanol, isoprene, pentane, 1-decene, 1-heptene, 1-nonene, 1-octene, 3-methylhexane, (E)-2-nonene, ammonia, ethane, hydrogen sulphide, triethylamine, and trimethylamine. The VOCs analyzed here are common compounds found in the human breath metabolome, and therefore should be easy to detect across platforms. Several of the 22 VOCs have previously been found to be associated with renal failure, alcoholic hepatitis, and inflammatory bowel disease.

METABOLITE DATA PROCESSING

Histograms were plotted for each metabolite to assess parametric assumptions. Because most metabolites demonstrated right skew and long tails, log transformation was used to normalize the data. PCA was performed to detect potential batch effects or outlying samples. Seven outlying samples were removed from further analyses and are described in Supporting Fig. S1. To avoid bias in model training, samples were also removed from the machine-learning analysis if they were missing data on more than 20% of the 22 metabolites. Any remaining missing metabolite data were mean-imputed before being used in the machine-learning models. All P values from the covariate and metabolite-disease association tests were adjusted for multiple comparisons using the Benjamini-Hochberg false-discovery rate (FDR) approach.

METABOLITE-DISEASE ASSOCIATIONS

A workflow describing the statistical analysis can be found in Fig. 1. Each of the 22 metabolites was tested for association with each patient cohort using logistic regression. Significant demographic variables and laboratory tests common between groups were included as model covariates to correct for potential confounding (Supporting Tables S1-S9). Association
Fig. 1. Workflow diagram for data analysis and predictive modeling. Abbreviation: MS, mass spectrometry.
tests were performed using both the original data set and an imputed data set, in which missing values were replaced with the mean of each feature. The imputed results are provided in Supporting Tables S17-S23. Metabolites and covariates with an FDR P value < 0.1 were considered to be statistically significant.

RANDOM FOREST ENSEMBLE CLASSIFICATION

A random forest ensemble classification approach was implemented to determine whether combinations of metabolites, with age and sex, could accurately classify patients by disease status. A description of the random forest model is provided in the Supporting Information. Models were developed that included (1) metabolites only, (2) demographic variables only (i.e., age, sex), and (3) metabolites and demographic variables. Random forest was implemented using the R package, RandomForest. (29)

First, we randomly selected 5% of the patients from each group to be excluded from model training to be used as a test set to assess model performance (n = 12). The remaining subjects were incorporated into a training and validation cohort. A grid search was performed on these remaining subjects to optimize the random forest hyperparameters. A leave-one-out cross validation (LOOCV) approach was implemented during the grid search, which iteratively removed an individual subject from the model training and then tested the model on the withheld subject. This process was repeated until each sample was used as a validation case. The grid search evaluated the optimal number of decision trees (ntrees) and the number of randomly selected variables to choose from at each node in the decision tree (mtry). Additional details regarding the grid search can be found in the Supporting Information. Hyperparameters 16 and 100 for mtry and ntrees, respectively, produced the highest classification accuracy and were subsequently used to develop the final model. The classification accuracy for each mtry and ntrees combination value is shown in Supporting Fig. S2.

Mean classification accuracy, sensitivity, specificity, and balanced accuracy (BA) on the withheld validation subjects were used to evaluate the model’s predictive performance. Classification accuracy refers to the proportion of correctly identified subjects (true positives and true negatives) out of the total number of subjects. The best performing model was then evaluated using the withheld test cohort. There were not enough individuals in the test set (n = 12) to evaluate sensitivity and specificity. However, the classification accuracy of the test set was used to evaluate whether the LOOCV approach was overfit. The mean decrease Gini estimates across the cross-validation procedure were used to provide an estimate of the importance of each feature to the performance of the model.

RESULTS

PATIENT CHARACTERISTICS

Cohort descriptions and summary statistics are provided in Table 1. The cohort included healthy controls (n = 54), patients with cirrhosis (n = 30), HCC (n = 112), CRLM (n = 51), and pulmonary hypertension (n = 49). The mean age for each patient cohort were 58.8, 59.6, 66.7, 55.9, and 61.9 for healthy controls and patients with cirrhosis, HCC, CRLM, and pulmonary hypertension, respectively. The percent of females in each group ranged from 25% for patients with HCC to 71% for patients with pulmonary hypertension. Data on race were not collected for patients with pulmonary hypertension or healthy controls, but most patients with cirrhosis, HCC, and CRLM self-reported as Caucasian (83%, 71%, and 90%, respectively).

METABOLITE-DISEASE ASSOCIATIONS

Individual metabolite differences between healthy control subjects and subjects in each disease category were tested. In addition, metabolite differences between individuals with cirrhosis and HCC, individuals with cirrhosis and CRLM, and between individuals with HCC and CRLM, were also tested. All associations were adjusted for any significant clinical covariates between the two groups, such as age or sex, to account for any confounding factors. The full list of covariate association results is provided in Supporting Tables S1-S9. A table with the three most significant metabolite associations for each disease comparison are presented in Table 2. Sample sizes in Table 2 reflect the data after filtering
# TABLE 1. STUDY COHORT SUMMARY

| Characteristic                                      | Cirrhosis | HCC       | CRLM       | Healthy Control | Pulmonary Hypertension |
|-----------------------------------------------------|-----------|-----------|------------|------------------|------------------------|
| Total (n)                                           | 30        | 112       | 51         | 54               | 49                     |
| Mean age (min-max)                                  | 59.6 (37-79) | 66.7 (25-95) | 55.9 (26-82) | 58.8 (36-80)     | 61.9 (45-85)           |
| Sex                                                 |           |           |            |                  |                        |
| Male (%)                                            | 14 (46%)  | 84 (75%)  | 27 (50%)   | 27 (50%)         | 14 (29%)               |
| Female (%)                                          | 16 (53%)  | 28 (25%)  | 14 (26%)   | 14 (26%)         | 35 (71%)               |
| Race                                                |           |           |            |                  |                        |
| Caucasian (%)                                       | 25 (83%)  | 80 (71%)  |            |                  |                        |
| Black (%)                                           | 3 (10%)   | 21 (19%)  |            |                  |                        |
| Hispanic (%)                                        | 1 (3%)    | 1 (1%)    |            |                  |                        |
| Other (%)                                           | 1 (3%)    | 4 (4%)    | 2 (4%)     |                  |                        |
| Mean BMI (min-max)                                  | 27.66 (12.54-40.17) | 28.61 (15.6-50.76) | 26.3 (17.71-47.46) |                  |                        |
| Cirrhosis                                           |           |           |            |                  |                        |
| Yes (%)                                             | 30 (100%) | 75 (67%)  | 1 (2%)     | 1 (2%)           |                        |
| NASH (%)                                            | 2 (6%)    | 14 (13%)  | 0          |                  |                        |
| EtOH (%)                                            | 8 (26%)   | 26 (23%)  | 0          |                  |                        |
| HCV (%)                                             | 8 (26%)   | 52 (46%)  | 0          |                  |                        |
| HBV (%)                                             | 2 (6%)    | 2 (2%)    | 0          |                  |                        |
| Hemochromatosis (%)                                 | 4 (13%)   | 0         | 0          |                  |                        |
| Alpha 1 antithypsin deficiency (%)                  |           | 1 (1%)    | 0          |                  |                        |
| Wilson’s disease (%)                                |           | 0         |            |                  |                        |
| Other (%)                                           | 9 (90%)   | 15 (13%)  | 1 (2%)     |                  |                        |
| No (%)                                              | 31 (28%)  | 48 (94%)  |            |                  |                        |
| Child-Pugh score                                    |           |           |            |                  |                        |
| 1                                                   | 1 (3%)    |           |            |                  |                        |
| 5                                                   | 1 (3%)    | 8 (7%)    | 14 (27%)   |                  |                        |
| 6                                                   |            | 3 (3%)    | 1 (1%)     |                  |                        |
| 7                                                   |            | 1 (1%)    | 1 (1%)     |                  |                        |
| 8                                                   |            |           | 1 (1%)     |                  |                        |
| Diabetes mellitus (%)                               | 10 (33%)  | 37 (33%)  | 1 (1%)     |                  |                        |
| Hypertension (%)                                    | 14 (46%)  | 62 (55%)  | 12 (24%)   |                  |                        |
| Coronary artery disease (%)                         | 5 (16%)   | 17 (15%)  | 2 (2%)     |                  |                        |
| Hyperlipidemia (%)                                  | 8 (26%)   | 28 (25%)  | 6 (12%)    |                  |                        |
| Psychiatric disorder (%)                            | 10 (33%)  | 14 (13%)  | 5 (10%)    |                  |                        |
| Other cancer history                                |           |           |            |                  |                        |
| B-cell lymphoma (%)                                 | 2 (6%)    |           |            |                  |                        |
| Granulosa cell tumor (%)                            | 1 (3%)    |           |            |                  |                        |
| COPD/Asthma/OSA                                      | 8 (26%)   | 22 (20%)  | 5 (10%)    |                  |                        |
| Thyroid                                             | 4 (13%)   | 8 (7%)    | 2 (2%)     |                  |                        |
| Other PH                                            | 3 (10%)   | 43 (38%)  | 16 (31%)   |                  |                        |
| Encephalopathy                                      |           |           |            |                  |                        |
| Grade 1-2 (%)                                       | 27 (90%)  | 1 (1%)    | 1 (2%)     |                  |                        |
| Grade 3-4 (%)                                       | 1 (3%)    |           |            |                  |                        |
| None (%)                                            | 1 (3%)    | 103 (92%) | 43 (84%)   |                  |                        |
| Mean hemoglobin (SEM)                               | 11.6 (0.403) | 12.7 (0.222) | 12.1 (0.339) |                  |                        |
| Mean platelets (SEM)                                | 142 (15.8) | 180 (10.5) | 188 (11.9) |                  |                        |
| Mean ALP (SEM)                                      | 163 (20.6) | 141 (9.7) | 142 (19.7) |                  |                        |
| Mean AST (SEM)                                      | 43.8 (3.77) | 78.6 (10.7) | 71.9 (24.2) |                  |                        |
| Mean ALT (SEM)                                      | 32 (4.46)  | 76.3 (11.7) | 58.4 (17.4) |                  |                        |
| Mean bilirubin (SEM)                                | 1.12 (0.17) | 0.998 (0.09) | 0.929 (0.259) |                  |                        |
for outliers and removing missing data. The full list of association results is given in Supporting Tables S10–S16. Association tests were also performed using an imputed data set. The association results for the imputed data set are provided in Supporting Tables S17–S23.

TABLE 2. TOP 3 ASSOCIATION RESULTS BETWEEN METABOLITES AND EACH PAIRWISE COMBINATION OF DISEASE STATUSES

Comparison  
N*  
Metabolite  
Mean Difference in Relative Abundance  
Log (Odds Ratio [SEM])  
OR (95% CI)  
P Value  
FDR P Value  
Healthy vs. HCC†  
136  
(E)-2-Nonene  
0.35  
7.4 (1.4)  
1.60 (100-25,000)  
1.53 × 10⁻⁷  
3.36 × 10⁻⁶  
Ethane  
15.52  
9.4 (2)  
12,000 (250-560,000)  
1.84 × 10⁻⁶  
1.69 × 10⁻⁵  
Benzene  
1.13  
2.3 (0.49)  
10 (3.9-27)  
2.3 × 10⁻⁶  
1.69 × 10⁻⁵  
Healthy versus cirrhosis  
77  
Methylhexane  
10.07  
6.3 (1.3)  
540 (40-7,300)  
2.12 × 10⁻⁶  
2.65 × 10⁻⁵  
Decene  
0.49  
6.8 (1.5)  
900 (52-15,000)  
2.73 × 10⁻⁶  
2.65 × 10⁻⁵  
Acrylonitrile  
0.39  
5.7 (1.2)  
310 (27-3,700)  
3.99 × 10⁻⁶  
2.65 × 10⁻⁵  
Healthy versus PH‡  
85  
(E)-2-Nonene  
0.26  
5.7 (1.2)  
310 (27-3,600)  
3.77 × 10⁻⁶  
8.29 × 10⁻⁵  
Acetaldehyde  
9.92  
6.9 (1.6)  
1000 (47-23,000)  
1.11 × 10⁻⁵  
8.87 × 10⁻⁵  
Ethane  
10.85  
8.2 (1.9)  
3500 (91-140,000)  
1.12 × 10⁻⁵  
8.87 × 10⁻⁵  
Healthy versus CRLM  
97  
(E)-2-Nonene  
0.35  
2.8 (0.67)  
16 (4.3-59)  
3.29 × 10⁻⁵  
0.0007238  
Acetaldehyde  
2.31  
3.6 (1)  
8 (5.3-270)  
0.0003  
0.003355  
Triethyl amine  
0.07  
3 (0.89)  
21 (3.6-120)  
0.0066  
0.00484  
Cirrhosis versus HCC§  
128  
Acetone  
−344.03  
−1.3 (0.32)  
0.267 (0.14-0.5)  
7.53 × 10⁻⁵  
0.0006  
Acetaldehyde  
−14.07  
−2 (0.51)  
0.13 (0.048-0.35)  
6.5 × 10⁻⁵  
0.0006  
Dimethyl Sulfide  
−4.55  
−1.4 (0.35)  
0.25 (0.12-0.5)  
8.19 × 10⁻⁵  
0.0006  
HCC versus CRLM||  
136  
Isoprene  
−16.86  
−1.5 (0.47)  
0.21 (0.085-0.54)  
0.00117  
0.013  
Pentane  
−6.39  
−2.2 (0.67)  
0.12 (0.031-0.43)  
0.0013  
0.013  
Acetone  
−13.01  
−0.93 (0.31)  
0.39 (0.22-0.72)  
0.0024  
0.013  
Cirrhosis versus CRLM¶  
71  
Methylhexane  
−8.97  
−4.9 (1.2)  
0.0076 (0.0007-0.088)  
9.31 × 10⁻⁵  
0.0008  
Isoprene  
−42.52  
−3.2 (0.84)  
0.041 (0.008-0.21)  
0.00015  
0.0008  
Trimethyl amine  
−0.07  
−2.8 (0.74)  
0.062 (0.014-0.26)  
0.000167  
0.0008

*Sample sizes reflect data after removing outliers and missing data.
†Age was included as a covariate.
‡Sex was included as a covariate.
§Age and sex were included as covariates.
||Age, hypertension, and albumin were included as covariates.
¶Albumin and diabetes mellitus were included as covariates.
Abbreviations: CI, confidence interval; OR, odds ratio; PH, portal hypertension.

Healthy Versus HCC

All of the results between healthy controls and HCC described subsequently were adjusted for age. Interestingly, 18 of the 22 identified metabolites were significantly increased in patients with HCC compared...
to healthy control subjects (FDR \( P < 0.1 \)). Of these 18 metabolites, (E)-2-nonene (FDR \( P = 3.03 \times 10^{-5} \)), ethane (FDR \( P = 1.08 \times 10^{-4} \)), and benzene (FDR \( P = 1.08 \times 10^{-4} \)) were the most significantly increased. Hydrogen sulfide (FDR \( P = 1.12 \times 10^{-3} \)) was the only metabolite that was significantly decreased in patients with HCC compared to healthy control subjects.

### Healthy Versus Cirrhosis

Eighteen of the 19 metabolites associated with HCC were also significantly different between healthy controls and patients with cirrhosis. However, unlike patients with HCC, trimethylamine (FDR \( P = 2.8 \times 10^{-4} \)) and propanol (FDR \( P = 2.35 \times 10^{-3} \)) were significantly increased in patients with cirrhosis compared to healthy controls.

### Healthy Versus Pulmonary Hypertension

For healthy controls and patients with pulmonary hypertension, 16 of 22 metabolites were significantly different between these groups (FDR \( P < 0.1 \)). (E)-2-nonene (FDR \( P = 2.80 \times 10^{-4} \)), acetaldehyde (FDR \( P = 1.94 \times 10^{-4} \)), and ethane (FDR \( P = 1.94 \times 10^{-4} \)) were the most significantly increased in patients with pulmonary hypertension compared to healthy controls. Hydrogen sulfide (FDR \( P = 0.0354 \)) was the only metabolite significantly decreased in patients with pulmonary hypertension compared to healthy controls.

### Healthy Versus CRLM

Of the 11 metabolites that were significantly increased in patients with CRLM compared to healthy controls, (E)-2-nonene (FDR \( P = 2.80 \times 10^{-4} \)), acetaldehyde (FDR \( P = 0.001 \)), and triethyl amine (FDR \( P = 0.002 \)) were the most significantly increased. Hydrogen sulfide (FDR \( P = 0.003 \)), trimethylamine (FDR \( P = 0.062 \)), and acetone (FDR \( P = 0.086 \)) were significantly decreased in patients with CRLM compared to healthy controls.

### Cirrhosis Versus HCC

Eighteen of 22 metabolites were significantly different between individuals with cirrhosis and those with HCC (FDR \( P < 0.1 \)). Acetone (FDR \( P = 6.7 \times 10^{-4} \)), acetaldehyde (FDR \( P = 6.7 \times 10^{-4} \)), and dimethyl sulfide (FDR \( P = 6.7 \times 10^{-4} \)) were the three most significantly enriched in patients with cirrhosis. Ethanol (FDR \( P = 0.089 \)) was the only metabolite that was significantly enriched in patients with HCC compared to those with cirrhosis.

### RANDOM FOREST MODEL PERFORMANCE

Three separate random forest models were constructed: (1) metabolites only, (2) age and sex only, and (3) metabolites, age, and sex. Model predictivity was assessed using classification accuracy, sensitivity, specificity, and BA for each disease category (Fig. 2). Models with metabolites, age, and sex (overall \( BA = 75\% \)) broadly outperformed models with metabolites only (overall \( BA = 68\% \)) or age and sex only (overall \( BA = 63\% \)) for each disease category, and had an overall classification accuracy of 85%. The one exception was for cirrhosis, which was worse with both metabolites and demographic variables (BA = 68%) than the model with only metabolites (BA = 68%) (Table 3). The results of the model combining metabolites, age, and sex for each specific disease group are detailed next.

### Healthy Controls

The model correctly identified 35 of 46 healthy controls and 198 of 205 true negatives, resulting in the highest classification accuracy (93%) of all the groups tested, with 76% and 97% sensitivity and specificity, respectively (BA = 86%). Although the model accurately identified healthy individuals and distinguished them from individuals with disease, the positive predictive value (PPV), the number of true positives out of all identified positives was 83%, suggesting that some of the predicted positives were false positives. The negative predictive value (NPV), the number of true negatives out of all identified negatives was 95%, indicating very few false negatives.

### Cirrhosis

The model’s ability to detect cirrhosis was the worst of all the groups tested, detecting only 8 of 27 cases. This resulted in a low sensitivity of 40%,
Fig. 2. Bar plots for random forest results grouped by disease status and model type: age and sex only, metabolites only, or metabolites, age, and sex. Disease status is color-coded (pink, pulmonary hypertension; white, healthy; gray, HCC; green, CRLM; and red, cirrhosis).
but a high specificity of 96% (BA = 68%) and an overall classification accuracy of 90%. The PPV was 50%, suggesting that half of those identified as cirrhosis by the model were false positives. Of the 10 false positives predicted to be cirrhosis, 50% of them belonged to the HCC group. Overall, the model accurately classified true negatives, those without cirrhosis (NPV = 94%), but was less able to distinguish cirrhosis from HCC and pulmonary hypertension.

HCC
The model identified 67 of the 92 individuals with HCC, resulting in an overall classification accuracy of 72%. The sensitivity and specificity for HCC was 73% and 71%, respectively, with an overall BA of 72%. There were 46 false-positive HCC predictions made (PPV = 59%), with 35% of those false positives belonging to the CRLM group. The NPV was 82%, suggesting that the model correctly identified most of the negatives.

CRLM
The model correctly identified 22 of 43 CRLM cases and 195 out of 208 true negatives, resulting in an overall classification accuracy of 86%. The sensitivity and specificity for CRLM were 51% and 94%, yielding a balanced accuracy of 72%. There were 13 false-positive CRLM predictions made by the model, resulting in a PPV of 63%. Of the 13 false positives, 69% of them belonged to the healthy group. There were 21 false negatives, resulting in a NPV of 90%. Similar to other disease categories, the model more accurately classified true negatives.

Pulmonary Hypertension
The model identified 26 of 45 pulmonary hypertension cases and 191 of 206 true negatives, resulting in an overall classification accuracy of 86%. The model had a balanced accuracy for pulmonary hypertension of 75%, with a sensitivity and specificity of 58% and 93%, respectively. There were 15 false positives, 33% of which were in the HCC group. The model resulted in 19 false negatives, resulting in a NPV of 91%. Most of the predicted negatives were correctly identified.

Test Set
To evaluate potential overfitting and assess final model performance, a hold-out group should be used as an independent cohort. To that end, 5% of each
patient cohort (n = 12) was withheld from model training in order to determine the classification accuracy of the LOOCV model. The model combining metabolites, sex, and age classified patients in the test set with an accuracy of 83%. The classification accuracies for specific groups were 100%, 66%, 75%, 92%, and 83% for cirrhosis, CRLM, HCC, healthy, and pulmonary hypertension, respectively. The model performance on the test set did not indicate any signs of overfitting from the LOOCV model, although larger cohorts are needed to more effectively evaluate model performance on an independent data set.

Figure 3 shows the relative ranking of each model feature based on the mean decrease Gini estimate for the model using metabolites, age, and sex. Age had the highest mean decrease in Gini score, indicating that it was the most important discriminating feature, followed by ethane, (E)-2-nonene, acetaldehyde, and acetone. The least important variables were ammonia, nonene, triethyl amine, decene, and acetonitrile.

**Discussion**

HCC represents a significant burden in global cancer-related deaths (1) (same box here for me). Improving detection through the identification of new biomarkers will be crucial to reducing the incidence of HCC and other liver-related comorbidities. Given the metabolic role of the liver, metabolomics is an ideal technology to detect liver diseases through the resulting perturbations in metabolic pathways. Metabolomics is increasingly being explored as a tool to find diagnostic biomarkers from blood, serum, urine, or breath samples for numerous diseases. Machine learning has the potential to discover complex relationships and patterns with metabolites and other features that can be used to construct a biomarker signature to detect the presence of disease. Discriminatory metabolites from breath samples, VOCs, may provide an opportunity for a noninvasive approach that could lead to earlier and more precise detection. Although previous studies have identified biomarker signatures that accurately distinguish healthy individuals from patients with HCC or cirrhosis, this study develops a model to differentiate healthy individuals from chronic liver diseases, a primary liver cancer, a secondary liver cancer, and a disease control (pulmonary hypertension). Rather than focusing solely on identifying HCC from cirrhosis or healthy individuals from those with HCC, the model presented here can accurately differentiate among multiple liver diseases. The addition of more disease categories broadens the utility of the model and helps us understand model limitations and disease misclassification that might occur with the real-world deployment of a similar screening tool. Effective models must be
sensitive to avoid missing patients with a disease (false negatives) and are specific to avoid needless follow-up testing (false positives). Given the noninvasive nature of breath collection and high specificity of the proposed model, it has the potential to yield clinical utility as a screening tool. By correctly identifying the true negatives, the model could identify patients who would be unlikely to benefit from additional testing. However, it is important to note that our model was also more sensitive at detecting HCC than AFP, indicating that it may also be better at detecting patients with HCC. Those patients who are likely within one of the disease cohorts may be followed up with confirmatory testing.

Here, we present a model that uses a signature of 22 breath VOCs to distinguish patients who are healthy, have HCC, cirrhosis, pulmonary hypertension, or CRLM. Importantly, the model was 73% sensitive for detecting HCC, which was substantially better than AFP, the current gold-standard biomarker, which had a sensitivity of 53% and specificity of 88% in the same cohort. AFP is a glycoprotein and an important biomarker of HCC; however, there are important limitations to its use in clinical practice for screening HCC. First, AFP is only secreted in approximately 50% of HCC tumors, resulting in too many false negatives. Second, while AFP levels above 400 ng/mL are generally considered diagnostic, this threshold may only be 60% sensitive. Incorporating AFP into the model may improve prediction of AFP. However, because blood samples were collected as part of the patient’s standard of care, AFP is not available for patients of other disease groups. Although age and sex were important variables in the models, as shown in the Gini scores (Fig. 3), the predictive accuracy increased considerably with the addition of metabolites in most cases. Despite controlling for confounding factors such as age and sex, significant differences in metabolite concentrations were seen between disease groups. The association results and predictive models both suggest that breath VOCs may have important roles as biomarkers.

Ideally, pathway analysis or a similar approach could be used to systematically investigate enriched biochemical pathways in certain disease groups that could point to a potential mechanism. However, because nearly all of the 22 VOCs tested were significantly different among groups, we are unable to identify a set of relatively perturbed pathways. Technologies that analyze a broader spectrum of VOCs will be needed to gain this type of mechanistic insight into these diseases. Nevertheless, we can use the Gini scores to approximate VOC importance, which may point to potential biomarkers or underlying disease mechanisms.

Many of the VOCs identified as being important contributors to discriminating the disease groups may serve a role in neoplastic development or progression, or as an indicator of dysfunction in the liver due to the presence of a tumor. For example, acetaldehyde, which had the third highest Gini score for the VOCs, was significantly increased in all disease groups compared to healthy controls. Acetaldehyde, according to the Kyoto Encyclopedia for Genes and Genomes, is involved in numerous biochemical pathways, including glycolysis/glucose metabolism, pyrvalate metabolism, dioxin degradation, and others. Acetaldehyde is genotoxic and is detoxified by the enzyme acetaldehyde dehydrogenase (ALDH), and studies have shown that individuals with polymorphisms (e.g., ALDH2*2)
have reduced activity of ALDH, resulting in accumulation of acetaldehyde. Furthermore, these individuals are known to have significantly higher relative risk for developing alcohol-related esophageal cancers and upper aerodigestive tract cancers.\(^{(37,38)}\)

Sulfur-containing compounds, dimethyl sulfide and carbon disulfide, were significantly elevated in patients with cirrhosis, HCC, and pulmonary hypertension compared with healthy controls (FDR \(P < 0.1\)). Carbon disulfide was also significantly increased in patients with CRLM compared with healthy controls. On the other hand, hydrogen sulfide was significantly decreased in patients with HCC, pulmonary hypertension, and CRLM compared with healthy controls. Decreased hydrogen sulfide was also observed in patients with cirrhosis compared to healthy controls, but was not statistically significant (FDR \(P = 0.14\)). Increased concentrations of dimethyl sulfide in the breath of patients with cirrhosis has been reported previously in multiple studies.\(^{(39-41)}\) Interestingly, a previous study using the same SIFT-MS device found that sulfur-containing compounds were decreased in childhood chronic liver disease compared with healthy controls.\(^{(42)}\) In the same study, (E)-2-nonenal was significantly decreased in children with chronic liver disease compared with healthy controls.\(^{(42)}\) However, here, (E)-2-nonenal was significantly increased in all disease groups compared with healthy controls (FDR \(P < 0.1\)).

Ketones have been shown to play an important role in the promotion of tumor growth and metastasis for various cancers.\(^{(43-46)}\) Acetone is produced during the decarboxylation of ketone bodies, and can be increased substantially in individuals with certain health conditions, such as alcoholism and diabetes, and is not genotoxic.\(^{(47)}\) Acetone was considered the fourth most important metabolite in our model and was significantly increased in cirrhosis, HCC, and pulmonary hypertension groups compared with healthy controls (FDR \(P < 0.1\)). However, acetone was significantly decreased in patients with CRLM compared to healthy controls (FDR \(P < 0.1\)).

Although this study represents an important step toward using VOC biomarkers for screening of chronic liver diseases, additional studies are needed to better characterize this technology in the context of these diseases. It is not currently known how the concentrations of the VOCs detected here change throughout disease progression. Although this study was not powered to investigate associations with comorbidities, future work would also benefit from investigating the impact of common comorbidities on prediction results. Future work would also benefit from comparing and contrasting early and late-stage HCC, as well as samples from individuals before and after resection. These additional samples may help predict early metabolomics signs of recurrence and increase the model's utility as a screening tool. Incorporating known biomarkers of HCC and colorectal cancer, such as AFP and carcinoembryonic antigen, respectively, may improve predictions. However, because blood samples for these samples were collected as part of their standard of care, these markers were not tested for most of the disease groups, and this will be an important investigation for future work. Because good medical practice involves performing tests or collecting information most likely to benefit patients, this type of missing data is common in health care data, and extends to covariates that were available for testing here and, like many studies, may result in bias. Pulmonary hypertension can sometimes result in congestive hepatopathy, which could potentially affect VOCs related to liver function. Future studies may consider a longitudinal design to follow up patients over time, although this will be challenging due to the slow progression of chronic liver disease. In addition, understanding the functional roles and biochemical pathways involved in the production of these VOCs may yield new biomarkers or therapeutic targets. Expanding the cohort to include early-stage liver diseases, such as NASH or nonalcoholic fatty liver disease, will be important, as 25% of the population is thought to have these diseases and they are often asymptomatic.\(^{(48)}\)

The combination of machine learning and VOC metabolomics presents a promising approach to biomarker discovery and noninvasive disease screening. With advancements in breath analysis technology, integrating VOC metabolomics and machine learning may help to provide accurate and noninvasive screening tests for multiple liver diseases and provide earlier detection of primary and secondary liver cancers.

REFERENCES

1) Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.
2) Rawla P, Sunkara T, Muralidharan P, Raj JP. Update in global trends and aetiology of hepatocellular carcinoma. Wspolczesna Onkol 2018;22:141-150.
3) Golabi P, Fazel S, Ogtonsuren M, Sayiner M, Locklear CT, Younossi ZM. Mortality assessment of patients with hepatocellular carcinoma according to underlying disease and treatment modalities. Med (United States) 2017;96:e5904.

4) Debruyne EN, Delanghe JR. Diagnosing and monitoring hepatocellular carcinoma with alpha-fetoprotein: new aspects and applications. Clin Chim Acta 2008;395:19-26.

5) Wang B, Chen D, Chen Y, Hu Z, Cao M, Xie Q, et al. Metabonomic profiles discriminate hepatocellular carcinoma from liver cirrhosis by ultraperformance liquid chromatography-mass spectrometry. J Proteome Res 2012;11:1217-1227.

6) Ruers T, Van Coevorden F, Punt CJA, Prieur J-PEN, Borel-Rinkes I, Ledermann JA, et al. Local treatment of unseetable colorectal liver metastases: results of a randomized phase II trial. J Natl Cancer Inst 2017;109:dx015.

7) Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 2016;15:473-484.

8) Kaddurah-Daouk R, Krishnan KRR. Metabolomics: a global biochemical approach to the study of central nervous system diseases. Neuropsychopharmacology 2009;34:173-186.

9) Toledo JB, Arnold M, Kastenmüller G, Chang R, Baillie RA, Han X, et al. Metabolic network failures in Alzheimer's disease: a biochemical road map. Alzheimer's Dement 2017;13:965-984.

10) Rottroff DM, Oki NO, Liang X, Yee SW, Stocker SL, Corum DG, et al. Pharmacometabolomic assessment of metformin in non-diabetic, African Americans. Front Pharmacol 2016;7:135.

11) Kaddurah-Daouk R, Weinshilboum R. Metabolic signatures for drug response phenotypes: pharmacometabolomics enables precision medicine. Clin Pharmacol Ther 2015;98:71-75.

12) Mikkonen JJW, Singh SP, Herrala M, Lappalainen R, Myllymaa S, Kullaa AM. Salivary metabolomics in the diagnosis of oral cancer and periodontal diseases. J Periodontal Res 2016;51:431-437.

13) Usher JR, Elmariah S, Gerszten RE, Dyck JRB. The emerging role of metabolomics in the diagnosis and prognosis of cardiovascular disease. J Am Coll Cardiol 2016;68:2850-2870.

14) Soga T, Sugimoto M, Homma M, Mori M, Igarashi K, Kishikura K, et al. Serum metabolomics reveals γ-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. J Hepatol 2011;55:896-905.

15) Acharjee A, Ament Z, West JA, Stanley E, Griffin JL. Integration of metabolomics, lipidomics and clinical data using a machine learning method. BMC Bioinf 2016;17:440.

16) Huang X, Zeng J, Zhou L, Hu C, Yin P, Lin X. A new strategy for analyzing time-series data using dynamic networks: identifying prospective biomarkers of hepatocellular carcinoma. Sci Rep 2016;6:23248.

17) Poozevara G, Lozano S, Cuissart B, Bureau R, Bureau P, Croisnamie V, et al. A computational selection of metabolite biomarkers using emerging pattern mining: a case study in human hepatocellular carcinoma. J Proteome Res 2017;16:2240-2249.

18) Fitzian AI, Nelson DR, Liu C, Xu Y, Ararat M, Cabrera R. Integrated metabolomic profiling of hepatocellular carcinoma in hepatitis C cirrhosis through GC/MS and UPLC/MS-MS. Liver Int 2014;34:1428-1444.

19) Liu Y, Hong Z, Tan G, Dong X, Yang G, Zhao L, et al. NMR and LC/MS-based global metabolomics to identify serum biomarkers differentiating hepatocellular carcinoma from liver cirrhosis. Int J Cancer 2014;135:658-668.

20) Baranska A, Majagic Z, Smolinska A, Dallinga JW, Jonkers DMAE, Tigheelaar EF, et al. Volatile organic compounds in breath as markers for irritable bowel syndrome: a metabolomic approach. Aliment Pharmocol Ther 2016;44:45-56.

21) Bos LDJ, Schultz MJ, Sterk PJ. Exhaled breath profiling for diagnosing acute respiratory distress syndrome. BMC Pulm Med 2014;14:72.

22) Filipiak W, Beer R, Sponning A, Filipiak A, Ager C, Schiefecker A, et al. Breath analysis for in vivo detection of pathogens related to ventilator-associated pneumonia in intensive care patients: a prospective pilot study. J Breath Res 2015;9:016004.

23) Oakley-Girvan I, Davis SW. Breath based volatile organic compounds in the detection of breast, lung, and colorectal cancers: a systematic review. Cancer Biomark 2017;21:29-39.

24) Peralbo-Molina A, Calderón-Santiago M, Priego-Capote F, Jurado-Gámez B, Luque De Castro MD. Identification of metabolomics panels for potential lung cancer screening by analysis of exhaled breath condensate. J Breath Res 2016;10:026002.

25) Van Oort PMP, De Bruin S, Hans W, Knobel HH, Schultz MJ, Bos LD. Exhaled breath metabolomics for the diagnosis of pneumonia in intubated and mechanically-ventilated intensive care unit (ICU)-patients. Int J Mol Sci 2017;18:449.

26) Patel S, Ahmed S. Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. J Pharm Biomed Anal 2015;107:63-74.

27) Smith D, Španel P. Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. Mass Spectrom Rev 2005;24:661-700.

28) Hochberg B. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995;57:589-600.

29) Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18-22.

30) Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet 2018;391:1301-1314.

31) Nisho K, Ichikawa T. Recent topics on α-fetoprotein. Hepatol Res 2013;43:820-825.

32) Ayuso J, Rinola J, Vilana R, Burrel M, Darnell A, García-Criado A, et al. Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines. Eur J Radiol 2018;101:72-81.

33) Rich N, Singal AG. Hepatocellular carcinoma tumour markers: current role and expectations. Best Pract Res Clin Gastroenterol 2014;28:843-853.

34) Lee E, Edward S, Singal AG, Laviere MS, Volk M. Improving screening for hepatocellular carcinoma by incorporating data on levels of α-fetoprotein, over time. Clin Gastroenterol Hepatol 2013;11:437-440.

35) Kanelihasi M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40:D109-D114.

36) Cogliano VJ, Baan R, Straif K, Grosse Y, Lauby-Secretan B, El Ghissassi F, et al. Preventable exposures associated with human cancers. J Natl Cancer Inst 2011;103:1827-1839.

37) Secretan B, Straif K, Baan R, Grosse Y, El Ghissassi F, Bouvard V, et al. A review of human carcinogens—part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol 2009;10:1033-1034.

38) Yokoyama A, Omori T. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and role for esophageal and head and neck cancers. Alcohol 2005;35:175-185.

39) van den Velde S, Nevens F, Vanhee P, Vansteenbergh D, Quirynen M. GC-MS analysis of breath odor compounds in liver patients. J Chromatogr B Anal Technol Biomed Life Sci 2008;875:344-348.

40) Hiroshi K, Masaya H, Nariyoshi S, Makoto M. Evaluation of volatile sulfur compounds in the expired alveolar gas in patients with liver cirrhosis. Clin Chim Acta 1978;85:279-284.

41) Tangerman A, Baranska A, Majagic Z, Smolinska A, Dallinga JW, Jonkers DMAE, Tigheelaar EF, et al. Volatile organic compounds in breath as markers for irritable bowel syndrome: a metabolomic approach. Aliment Pharmocol Ther 2016;44:45-56.
42) Eng K, Alkhouri N, Cikach F, Patel N, Yan C, Grove D, et al. Analysis of breath volatile organic compounds in children with chronic liver disease compared to healthy controls. J Breath Res 2015;9:026002.

43) Martínez-Outschoorn UE, Prisco M, Ertel A, Tsirigos A, Lin Z, Pavlides S, et al. Ketones and lactate increase cancer cell “stemness”, driving recurrence, metastasis and poor clinical outcome in breast cancer: achieving personalized medicine via metabologenomics. Cell Cycle 2011;10:1271-1286.

44) Bonuccelli G, Tsirigos A, Whitaker-Menezes D, Pavlides S, Pestell RG, Chiavarina B, et al. Ketones and lactate “fuel” tumor growth and metastasis: evidence that epithelial cancer cells use oxidative mitochondrial metabolism. Cell Cycle 2010;9:3506-3514.

45) Shukla SK, Gebregiworgis T, Purohit V, Chaika NV, Gunda V, Radhakrishnan P, et al. Metabolic reprogramming induced by ketone bodies diminishes pancreatic cancer cachexia. Cancer Metab 2014;2:18.

46) Pavlides S, Tsirigos A, Migneco G, Whitaker-Menezes D, Chiavarina B, Flomenberg N, et al. The autophagic tumor stroma model of cancer: role of oxidative stress and ketone production in fueling tumor cell metabolism. Cell Cycle 2010;9:3485-3505.

47) U.S. Environmental Protection Agency. Toxicological review of acetone. May 2003. https://cfpub.epa.gov/ncea/iris/iris_documents/documents/toxreviews/0128tr.pdf. Accessed March 3, 2020.

48) Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018;15:11-20.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep4.1499/suppinfo.