Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study
Stanislas Grassin-Delyle, Camille Roquencourt, Pierre Moine, Gabriel Saffroy, Stanislas Carn, Nicholas Heming, Jérôme Fleuriet, Hélène Salvator, Emmanuel Naline, Louis-Jean Couderc, et al.

To cite this version:

HAL Id: cea-03086093
https://hal-kea.archives-ouvertes.fr/cea-03086093
Submitted on 22 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study

Stanislas Grassin-Delyle, Ph.D.a,b,f,#,* Camille Roquencourt, M.S.c,# Pierre Moine, M.D.b,e, Gabriel Saffroya, Stanislas Carnb, Nicholas Heming, M.D.b,e, Jérémy Fleuriet, Ph.D.c, Hélène Salvator, M.D.a,d, Emmanuel Naline, Ph.D.a,f, Louis-Jean Couderc, M.D.a,f, Philippe Devillier, M.D.a,d, Etienne A. Thévenot, Ph.D.d,f, Djillali Annane, M.D.b,e,f, for the Garches COVID-19 Collaborative GroupRECORDS Collaborators and Exhalomics® Collaborators

a Hospital Foch, Exhalomics®, Département des maladies des voies respiratoires, Suresnes, France (S.G.D., H.S., E.N., L-J.C., P.D.)

b Université Paris-Saclay, DVSQ, INSERM, Infection et inflammation, Montigny le Bretonneux, France (S.G.D., P.M., N.H, D.A.)

c CEA, LIST, Laboratoire Sciences des Données et de la Décision, GIF-sur-Yvette, France (C.R.)

d Département Médicaments et Technologies pour la Santé (DMTS), Université Paris-Saclay, CEA, INRAE, MetaboHUB, GIF-sur-Yvette, France (E.T.)

e Intensive Care Unit, Raymond Poincare Hospital, Assistance Publique-Hôpitaux de Paris, Garches, France (P.M., G.S., S.C., N.H., J.F., D.A.)

# contributed equally to this work

ABSTRACT

Background: Early diagnosis of coronavirus disease 2019 (COVID-19) is of the utmost importance but remains challenging. The objective of the current study was to characterize exhaled breath from mechanically ventilated adults with COVID-19.

Methods: In this prospective observational study, we used real-time, online, proton transfer reaction time-of-flight mass spectrometry to perform a metabolomic analysis of expired air from adults undergoing invasive mechanical ventilation in the intensive care unit due to severe COVID-19 or non-COVID-19 acute respiratory distress syndrome (ARDS).

Findings: Between March 25th and June 25th, 2020, we included 40 patients with ARDS, of whom 28 had proven COVID-19. In a multivariate analysis, we identified a characteristic breathprint for COVID-19. We could differentiate between COVID-19 and non-COVID-19 ARDS with accuracy of 93% (sensitivity: 90%, specificity: 94%, area under the receiver operating characteristic curve: 0.98-0.98, after cross-validation). The four most prominent volatile compounds in COVID-19 patients were methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanol.

Interpretation: The real-time, non-invasive detection of methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanol in exhaled breath may identify ARDS patients with COVID-19.

Funding: The study was funded by Agence Nationale de la Recherche (SoftwAiR, ANR-18-CE45-0017 and RHU4 RECORDS, Programme d'Investissements d'Avenir, ANR-18-RHUS-0004), Région Ile de France (SESAME 2016), and Fondation Foch.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Introduction

As of November 21st, 2020, about 57 million of people worldwide had been infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and about 1•4 million had died from coronavirus disease 2019 (COVID-19) [1]. Approximately 5% of patients with COVID-19 will develop acute respiratory distress syndrome (ARDS), septic shock, or multiple organ dysfunction [2]. Around the world, unprecedented research efforts are being focused on the prevention, early detection, diagnosis and management of this lethal disease. To date, only one antiviral drug (remdesivir) has been approved for the treatment of patients hospitalized for COVID-19 [3]. More recently, a large trial showed that dexamethasone at a daily dose of 6 mg for 10 days substantially reduced the risk of 28 day death (age-adjusted rate ratio [95% confidence interval (CI)]: 0•83 [0•75 to 0•93], particularly in patients with severe disease requiring invasive mechanical ventilation (rate ratio: 0•64 [0•51 to 0•81]) [4]. Although the early
Studies design and oversight

This prospective study was part of the observational phase of the ongoing RECORDS trial (NCT04280497) and was conducted at the ICU of Raymond Poincaré Hospital (Garches, France). The RECORDS study protocol was approved by an ethics committee (Comité de Protection des Personnes EST I, Dijon, France; reference 20.03.10.51415) and the French National Agency for Healthcare Product Safety (ANSM, Paris, France). The study was registered with the European Union Drug Regulating Authorities Clinical Trials Database (EudraCT 2020-000296-21). Whenever possible, participants or their legally authorized next of kin provided written, informed consent before inclusion. In the remaining cases, patients provided their deferred, written, informed consent. This investigator-led study was publicly funded. All the authors had full and independent access to all data and vouch for the integrity, accuracy, and completeness of the data and analysis and for the adherence of the trial to the protocol.

Study participants

Adult patients (aged 18 or over) in ICUs were eligible for inclusion if they had ARDS and required invasive mechanical ventilation. ARDS was defined as all of the following: (i) acute onset, i.e., within one week of an apparent clinical insult, followed by progression of the respiratory syndrome, (ii) bilateral opacities on chest imaging not explained by another lung disease (e.g., pleural effusion, atelectasis, nodules etc.), (iii) no evidence of heart failure or volume overload, and (iv) \( \text{PaO}_2/\text{FiO}_2 \leq 300 \text{ mm Hg} \), and positive end expiratory pressure (PEEP) \( \geq 5 \text{ cm H}_2\text{O} \) [26]. The main exclusion criteria were pregnancy, an expectation of death within 48 h, and the withholding or withdrawal of treatment.

Study measurements and procedures

Variables recorded at baseline were patient demographics and anthropometrics, the source of infection, and the severity of illness (according to the Simplified Acute Physiology Score (SAPS) II and the Sequential Organ Failure Assessment (SOFA)) [27,28]. The following variables were recorded at baseline and daily during the hospital stay: core body temperature, vital signs, central hemodynamic data, standard laboratory data, microbiological and virologic data. Samples for routine surveillance of lower respiratory tract colonization were obtained every 72 h until the patient had been weaned off mechanical ventilation or had died. A nonbronchoscopic bronchoalveolar lavage was performed with three 20 mL aliquots of sterile 0.9% saline solution, with a view to collect at least 5–10 mL of effluent per sample. Samples of blood and nasopharyngeal, bronchial or bronchoalveolar lavage fluids were assayed for SARS-CoV-2 and other respiratory viruses with a PCR test, as described by the French National Reference Center for Respiratory Viruses (Institut Pasteur, Paris, France). We also recorded life-supportive therapies including mechanical ventilation, renal replacement therapy, intravenous fluids bolus and the administration of vasopressors, and adjunct therapies including corticosteroids, thiamine, vitamin C, other vitamins, nutritional supplements, blood products, anticoagulants, sedatives, stress ulcer prophylaxis, and anti-infective drugs.

Breath analysis

Each patient's expired air was analyzed daily in the morning until discharge. Measurements were made with a proton-transfer-reaction quadrupole time-of-flight mass spectrometer (Ionicon Analytik GmbH, Innsbruck, Austria) placed outside the patient room. Samples were obtained via a heated transfer line (length: 1.6 m) connected directly to the end of the endotracheal tube (i.e., without disconnection from the mechanical ventilator) and with an air flow of 50 mL/min. To eliminate the dependency on the oxygen concentration in the sample matrix, recordings were performed in patients with a fraction of inspired oxygen of 100% for at least 3 min [29]. The acquisition took 2 min. \( \text{H}_2\text{O}^+ \) was used as the primary ion and the instrument settings were as follows: source voltage, 120 V; drift tube pressure, 3\( \times \)8 mbar;...
drift tube temperature, 60 °C, and drift tube voltage, 959 V. The mass spectrum was acquired up to \( m/z = 392 \), with a time resolution of 0.1 s.

**Data and statistical analysis**

Patient characteristics were expressed as the median [interquartile range (IQR)] for continuous variables and the frequency (percentage) for categorical variables. Patients with and without COVID-19 were compared using Fisher’s exact test for categorical variables, and a t-test or the Mann-Whitney test for normally and non-normally distributed continuous variables (as evaluated with the d’Agostino-Pearson test), respectively.

Mass spectrometry data were processed with the *ptairMS* R package (https://github.com/camilleroquencourt/ptairMS) and included mass calibration, expiratory phase detection on the CO\(_2\) extracted ion chromatogram, peak detection and quantification with background subtraction, normalization, alignment, isotope identification, and imputation of missing values. All concentration values were quoted in ppb [30]. After aligning each individual peak, ions detected in more than 70% of at least one group (COVID vs. non-COVID-19 ARDS) were kept; this resulted in 81 features. Missing values (corresponding to ions in exhaled breath that were not detected by the preprocessing algorithm) were imputed with the *ptairMS* package, which returns to the raw data and integrates the noise at the exact missing \( m/z \). Data were then log2-transformed and standardized. Outliers (patients with a z-score >3 for at least five features) were deleted. In the remaining patients, saturated ions (acetone, H\(_2\)O\(^+\), H\(_2\)O-H\(_3\)O\(^+\), oxygen) and isotopes were deleted to leave a final table of 65 features. For the univariate analysis, a Wilcoxon test was performed and \( p \)-values were adjusted to control for the false discovery rate [31]. For multivariate analysis, data were analyzed first with principal component analysis and then with machine learning algorithms with different mathematical backgrounds (orthogonal partial least-squares discriminant analysis (OPLS-DA), linear support vector machine (SVM), elastic net, and random forest (RF); summarized in Table S1) with the R packages ropls, e1071, and caret [32–35]. A 10-fold, stratified cross-validation was repeated four times (in order to avoid overfitting the small number of data points), and features were selected with the elastic net and RF approaches. The models’ parameters were tuned to optimize the accuracy of cross-validation. Features were ranked according to the specific metrics of each modeling method (\( p \)-values from the Wilcoxon test, absolute loading values from PCA, the variable importance in projection from OPLS-DA, the coefficient values from the elastic net and SVM models, and the feature importance from the RF model). An aggregated ranking was then computed by maximizing the sum of the Spearman correlation with each of the metric rankings (RankAggreg R package) [36]. The correlations between the metric rankings and the aggregated rank are shown in Fig. S3. To limit the risk of overfitting, we aggregated several metrics from statistical models with different mathematical backgrounds. The effects of tidal volume, serum C-reactive protein (CRP) level, body temperature, and the number of days spent in the ICU were investigated in a correlation test with the three first components of the PCA (using a Pearson’s test for continuous variables and a chi-squared test for categorical variables) to detect putative factors with a strong impact on the VOC concentrations which may interfere with the prediction of the COVID-19 status (Fig. S1). For the positive end-expiratory pressure (PEEP) and respiratory rate (the median levels of which differed for each COVID-19 status), we performed a Pearson correlation test within each group (as described in the Supplementary Material and Fig. S2). No significant correlations were detected by any of these tests.

A longitudinal univariate analysis of the most important features was performed with a mixed effects model. The fixed effect represents the change in the VOC concentration as a function of the period of mechanical ventilation, with only one measurement per patient per day. We chose a spline function (sum of four b-spline functions basis of degree three uniformly distributed over time) for the fixed effect and an intercept per patient for the random effect. Intergroup differences in trends and means were assessed with an F-test (\( p \)-value <0.005) adjusted for the false discovery rate. The test compares the residuals of models with and without COVID status as a predictor. Correlations between VOC concentrations, the SAPS II, the SOFA score, and the viral load were analyzed using Pearson’s correlation test, after adjustment for the false discovery rate.

**Role of the funding source**

The funding source had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. The corresponding author confirms that he had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Results**

**Patients**

Between March 25\(^{th}\) and June 25\(^{th}\), 2020, 40 patients (of whom 28 had confirmed COVID-19-related ARDS) were included in the study and a total of 303 measurements were made. Compared with the patients with non-COVID-19 ARDS, the patients with COVID-19 ARDS had (i) a higher respiratory rate, FiO\(_2\), PEEP and CRP on admission, (ii) a higher incidence of treatment with hydroxychloroquine and a lower incidence of treatment with fludrocortisone after admission, and (iii) a greater likelihood of renal replacement therapy (Table 1).

**Metabolomic analysis of exhaled breath**

We first used an untargeted metabolomic strategy to discover the signature associated with COVID-19 ARDS. To this end, we used the first breath sample collected after admission. Twelve of the 40 participants had been hospitalized for more than 10 days at the start of the sampling period and so were excluded from this first part of the study. Hence, we analyzed 18 patients with COVID-19 ARDS and 10 with non-COVID-19 ARDS. The study groups’ demographic characteristics are summarized in Table S2. A principal component analysis and an orthogonal partial least-squares discriminant analysis showed that COVID-19 was associated with a specific signature in the expired air, i.e., the breathprint could discriminate between COVID-19 ARDS and non-COVID-19 ARDS cases (Fig. 1). The use of three machine learning algorithms yielded an accuracy of 93% for all three classifiers, based on the selection of 19, 16 or 65 features for the elastic net, random forest, and support vector machine algorithms, respectively (in a 10-fold stratified cross-validation, repeated four times). The corresponding receiver operating characteristic curves are shown in Fig. 2a. A Wilcoxon test with \( p \)-value correction for the false discovery rate highlighted VOCs that significantly distinguished between the two groups (\( p <0.05 \)). We checked that none of the other external covariates impacted the VOC concentrations and interfered with the model’s predictions (see the Supplementary Material). To determine how VOCs were most discriminant for COVID-19 ARDS, we performed a rank aggregation based on the various metrics from the previously mentioned models and the hypothesis tests. The four most relevant features in the rank aggregation were at \( m/z \) 99\( \times \)08, 111\( \times \)12, 135\( \times \)09, and 143\( \times \)15 (Fig. 3a). Using these four features only, the elastic net, random forest, and support vector machine algorithms yielded an accuracy of between 89% and 93% (Fig. 2b). We therefore investigated the expression of these VOCs in the whole study population throughout the period of mechanical ventilation (Fig. 3b). We observed that the VOC concentrations (i) were significantly higher in the breath of patients with COVID-19 ARDS than in the breath of patients with non-COVID-19 ARDS, and (ii) tended to decrease over the first 10 days of hospitalization. The putative annotations for the
invasive mechanical ventilation in the ICU. This breathprint was exhaled breath from patients with COVID-19-related ARDS requiring VOCs and the determination of a specific VOC signature. We then performed a longitudinal analysis of expired air in ARDS patients, which allowed us to confirm the VOC signature and to track the changes over time in the VOC concentrations. Two of the four prominent VOCs (methylpent-2-enal and nonanal) were not significantly correlated with the bronchoalveolar fluid viral load or with the severity of illness (i.e., the SAPS II and SOFA score) measured during the first 24 h in the ICU (Table 2, r < 0.4).

Correlation with viral load and severity scores

The viral load in bronchoalveolar fluid was measured for 18 patients. The median [IQR] value in the first sample was 7 × 2 [6 × 2 - 8 × 4] log eq. copies/mL. The VOC concentrations in the first sample were not significantly correlated with the bronchoalveolar fluid viral load or with the severity of illness (i.e., the SAPS II and SOFA score) measured during the first 24 h in the ICU (Table 2, r < 0.4). Four VOCs (methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal) may discriminate between COVID-19 and non-COVID-19 ARDS.

We applied a highly sensitive, rapid, non-invasive, real-time mass spectrometry breath analysis. This contrasts with offline technologies, which require a sampling step and remote, time-consuming analytical steps. Implementation of a non-targeted strategy (as described here) is mandatory for the discovery of novel biomarkers. The subsequent diagnostic validation and clinical implementation can be based on less cumbersome technologies, such as mass spectrometers dedicated to targeted analyses or portable “electronic noses” with a set of sensors that are relatively selective for different families of VOCs (as previously used in patients with ARDS).

The first (cross-sectional) part of the present study enabled us to identify a specific signature. We then performed a longitudinal analysis of expired air in ARDS patients, which allowed us to confirm the VOC signature and to track the changes over time in the VOC concentrations. Two of the four prominent VOCs (methylpent-2-enal and nonanal) were respectively methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal.

### Table 1

<table>
<thead>
<tr>
<th>Patient characteristics and treatments</th>
<th>COVID-19 ARDS</th>
<th>Non-COVID-19 ARDS</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients (n)</td>
<td>28</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>Males/females (n)</td>
<td>20/8</td>
<td>6/6</td>
<td>0.28</td>
</tr>
<tr>
<td>Age (years)</td>
<td>61 [55-72]</td>
<td>72 [54-79]</td>
<td>0.75</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>80 [66-87]</td>
<td>86 [65-94]</td>
<td>0.71</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>170 [164-175]</td>
<td>173 [169-175]</td>
<td>0.55</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>26.3 [23-32]</td>
<td>26.9 [23-30]</td>
<td>0.79</td>
</tr>
<tr>
<td>SAPS II score in the first 24 hours</td>
<td>62 [49-68]</td>
<td>46 [30-57]</td>
<td>0.051</td>
</tr>
<tr>
<td>SOFA score in the first 24 hours</td>
<td>11 [7-12]</td>
<td>8 [5-12]</td>
<td>0.37</td>
</tr>
<tr>
<td>Comorbidities: (n (%))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26- high blood pressure</td>
<td>2 (7)</td>
<td>1 (8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>26- chronic obstructive pulmonary disease</td>
<td>5 (18)</td>
<td>3 (25)</td>
<td>0.68</td>
</tr>
<tr>
<td>26- ischemic cardiac disease</td>
<td>2 (7)</td>
<td>3 (25)</td>
<td>0.15</td>
</tr>
<tr>
<td>26- cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments before admission: (n (%))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26- glucocorticoids</td>
<td>1 (4)</td>
<td>3 (25)</td>
<td>0.073</td>
</tr>
<tr>
<td>26- conversion enzyme inhibitors</td>
<td>2 (7)</td>
<td>2 (16)</td>
<td>0.57</td>
</tr>
<tr>
<td>26- angiotensin antagonists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interventions after admission: (n (%))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26- catecholamines</td>
<td>17 (61)</td>
<td>4 (33)</td>
<td>0.17</td>
</tr>
<tr>
<td>26- renal replacement therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatments after admission: (n (%))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26- hydroxychloroquine</td>
<td>27 (96)</td>
<td>1 (8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>26- remdesivir</td>
<td>2 (7)</td>
<td>0 (0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>26- lopinavir/ritonavir</td>
<td>21 (73)</td>
<td>6 (50)</td>
<td>0.73</td>
</tr>
<tr>
<td>26- glucocorticoids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26- fluocortisone</td>
<td>12 (43)</td>
<td>4 (33)</td>
<td>0.73</td>
</tr>
<tr>
<td>26- eculizumab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body temperature at first sample (°C)</td>
<td>37.4 [36.5-38.3]</td>
<td>37.1 [36.8-37.8]</td>
<td>0.84</td>
</tr>
<tr>
<td>Respiratory rate at first sample (breaths per min)</td>
<td>26 [25-28]</td>
<td>20 [18-23]</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Tidal volume at first sample (mL)</td>
<td>420 [400-475]</td>
<td>438 [400-490]</td>
<td>0.99</td>
</tr>
<tr>
<td>Fraction of inspired oxygen at first sample (%)</td>
<td>80 [50-100]</td>
<td>48 [31-68]</td>
<td>0.007</td>
</tr>
<tr>
<td>Positive end-expiratory pressure at first sample (cm H2O)</td>
<td>10 [8-13]</td>
<td>5±5 [5-8]</td>
<td>0.0002</td>
</tr>
<tr>
<td>Serum creatinine at first sample (µM)</td>
<td>74 [56-137]</td>
<td>67 [44-86]</td>
<td>0.30</td>
</tr>
<tr>
<td>Serum C-reactive protein at first sample (mg/L)</td>
<td>195 [175-268]</td>
<td>76 [23-119]</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Continuous data are presented as the median [IQR].
nonanal) are aldehydes, while 2,4-octadiene is an alkadiene. These three compounds are known to be expressed in breath [39,40], while 1-chloroheptane is probably not endogenous. Nonanal is a sub-product of the destruction of the cell membrane as a result of oxidative stress; reactive oxygen species may be generated by various type of inflammatory, immune and structural cell in the airways [41]. In studies of expired air from patients with ARDS, Schubert et al. found abnormally low isoprene concentrations and Bos et al. reported abnormally high concentrations of octane, acetaldehyde and 3-methylheptane [21,22]. Differences in study populations (non-COVID-19 vs. COVID-19 ARDS) and analytical methods (of line vs. online) might explain the differences between the VOCs identified in the present study and those identified in previous studies of ARDS [21,22]. Although there may be an association between VOCs and disease, the underlying biochemistry has not been fully characterized.

In line with previous reports, the VOC concentrations measured here were not correlated with the severity of illness (as judged by the SAPS II and the SOFA score) [21]. This finding suggest that the exhaled breath signature is a marker of COVID-19 per se, rather than of the severity of illness. Likewise, the VOC concentrations were not correlated with viral load, suggesting that this signature may be a marker of the disease related to SARS-CoV-2 rather than of virus carriage.

Our interpretation of the present data may have been limited by differences between the COVID-19 and non-COVID-19 ARDS subgroups. Patients with COVID-19 ARDS cohort had higher respiratory rate, FiO₂, PEEP and CRP values on admission. The respiratory rate, PEEP and CRP were not found to interfere with the VOC predictive signature, and all the patients were sampled when breathing 100% FiO₂ (to avoid mass spectrometry interference by oxygen) [29]. Similarly, patients with COVID-19 ARDS were more likely to have been treated with hydroxychloroquine. However, this drug was administered to the patients after their first sample had been analyzed. Although the VOC concentrations decreased over time, the treatments did not change, and there was no correspondence between

---

**Fig. 1.** Multivariate analysis. Principal component analysis (left) and orthogonal partial least squares - discriminant analysis (right) of the breath signature in intubated, mechanically ventilated ICU patients with a positive (red) or negative (blue) PCR test for SARS-CoV-2.

**Fig. 2.** Receiver operating characteristic curves for models classifying patients with COVID-19 vs. non-COVID-19 ARDS. a. Complete model. The use of three machine learning algorithms (elastic net, support vector machine (SVM), and random forest (RF)) yielded an accuracy of up to 93%, with a 10-fold cross validation repeated four times and based on the selection of 19 features (elastic net), 16 features (random forest) or all 65 features (support vector machine) from the full dataset. After internal cross-validation, the sensitivity was 90% and the specificity was 94%. b. Model with the four most important features only. After internal cross-validation, the sensitivity ranged from 90% to 98% and the specificity ranged from 88% to 94%.
the VOCs described in the present study and the molecular masses of the known metabolites of hydroxychloroquine. Lastly, the sample size of this pilot study was limited and these observations will require confirmation with an external validation cohort.

In conclusion, we determined a COVID-19-specific breath metabolomic signature in patients with ARDS requiring invasive mechanical ventilation. Knowledge of this specific breathprint might enable the development of rapid, non-invasive, point-of-care tests for large-scale COVID-19 screening.

Acknowledgements

The authors would like to thank all the staff members at the intensive care unit at Raymond Poincaré Hospital for their collaboration, Professor Marie-Anne Wettl and Professor Elyanne Gault for providing virological data, and Dr David Fraser (Biotech Communication SARL, Ploudalmézeau, France) for copy-editing assistance. The study was funded by Agence Nationale de la Recherche (SoftwAIR, ANR-18-CE45-0017 and RHU4 RECORDS, Programme d’Investissements d’Avenir, ANR-18-RHUS-0004), Région Île de France (SESAME 2016), and Fondation Foch.

Table 2

<table>
<thead>
<tr>
<th>VOC (m/z)</th>
<th>SAPS II score</th>
<th>SOFA score</th>
<th>Viral load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p-value</td>
<td>r</td>
</tr>
<tr>
<td>99*08</td>
<td>0.04</td>
<td>0.0488</td>
<td>0.13</td>
</tr>
<tr>
<td>111*12</td>
<td>0.02</td>
<td>0.0493</td>
<td>0.25</td>
</tr>
<tr>
<td>135*09</td>
<td>0.05</td>
<td>0.0405</td>
<td>0.14</td>
</tr>
<tr>
<td>143*15</td>
<td>0.12</td>
<td>0.062</td>
<td>0.23</td>
</tr>
</tbody>
</table>

r: Pearson’s correlation coefficient.

the VOCs described in the present study and the molecular masses of the known metabolites of hydroxychloroquine. Lastly, the sample size of this pilot study was limited and these observations will require confirmation with an external validation cohort.

In conclusion, we determined a COVID-19-specific breath metabolomic signature in patients with ARDS requiring invasive mechanical ventilation. Knowledge of this specific breathprint might enable the development of rapid, non-invasive, point-of-care tests for large-scale COVID-19 screening.

Contributors

S.G.D. and D.A. conceived the study. S.G.D., P.M. and C.R. defined parameters for mass spectrometry breath analysis. S.G.D., G.S., S.C., J. F., H.S., E.N., L-J.C., P.D., P.M., D.A. performed the experiments and/or interpreted results. S.G.D., P.M., N.H., D.A. collected epidemiological and clinical data. P.M. and N.H. assisted in patient recruitment. C.R. and E.T. developed software and analyzed the data. S.G.D. and C.R. checked the underlying data. S.G.D., C.R. and D.A. drafted the manuscript. E.T., P.M., L-J.C., H.S., E.N., G.S., S.C., N.H., J.F. revised the manuscript. All authors read and approved the final version of the manuscript.

Declaration of Competing Interests

DA has received a grant from Agence Nationale de la Recherche to conduct the RECORDS program, of which this study is part (ANR-18-RHUS-0004). S.G.D, C.R., E.T. and D.A. are named as inventors on a patent application covering breath analysis in COVID-19. The authors declare no other conflicts of interest.

Data sharing statement

The study protocol and the datasets generated during and/or analysed during the current study, including deidentified participant data will be available with publication from the corresponding author on reasonable request. The ptairMS R package used for data analysis is publicly available at https://github.com/camillerouencourt/ptairMS
Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.103154.

References