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 Saffroy[e], Stanislas Carn[e], Nicholas Heming, M.D.[b,e], Jérôme Fleuriet, Ph.D.[e], Hélène Salvator, M.D.[a,f], Emmanuel Naline, Ph.D.[a,f], Louis-Jean Couderc, M.D.[a,f], Philippe Devillier, M.D.[a,f], 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] Hôpital 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, UVSQ, 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 Poincaré Hospital, Assistance Publique-Hôpitaux de Paris, Garches, France (P.M., G.S., S.C., N.H., J.F., D.A.)
[f] FHU SEPSIS (Saclay and Paris Seine Nord Endeavour to PerSonalize Interventions for Sepsis) (S.G.D., H.S., E.N., L-J.C., P.D., E.T., D.A.).

ARTICLE INFO

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•94-0•98, after cross-validation). The four most prominent volatile compounds in COVID-19 patients were methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal.

Interpretation: The real-time, non-invasive detection of methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal 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 Île de France (SESAME 2016), and Fondation Foch.

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

* Corresponding author.
E-mail address: stanislas.grassin-delyle@uvsq.fr (S. Grassin-Delyle).
# contributed equally to this work

Research in context

Evidence before this study

Early diagnosis of coronavirus disease 2019 (COVID-19) is of the utmost importance but remains challenging. Around 5% of patients with COVID-19 will develop acute respiratory distress syndrome (ARDS), septic shock and/or multiple organ failure; ideally, these patients should be identified as soon as possible. Breath analysis is an innovative, non-invasive, real-time, point-of-care technique for detecting volatile organic compounds (VOCs) in expired breath. It has potential for use in diagnosis and large-scale screening. However, it was not previously known whether patients with COVID-19 have a breath "signature" (also known as a "breathprint").

Added value of this study

Here, we show that breath analysis can discriminate between COVID-19 ARDS and non-COVID-19 ARDS. We characterized a VOC breathprint that was able to identify COVID-19 ARDS patients requiring invasive mechanical ventilation with high sensitivity and specificity. The four most prominent volatile compounds in the patients' breath were methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal. The COVID-19 breathprint did not depend on the severity of the ARDS or the patient's viral load.

Implications of all the available evidence

All the available evidence suggest that real-time, non-invasive breath analysis could enable the large-scale screening and thus earlier treatment of patients likely to develop severe forms of COVID-19.

immune response may not depend on the severity of the illness, the most severely ill patients show persistent elevations of blood inflammatory markers (such as IL-1α, IL-1β, IL-6, IL-10, IL-18 and TNF-α) 10 or so days after SARS-CoV-2 infection, with a very high risk of subsequent organ injury [5–7]. Proteomic and metabolomic studies of serum have described a COVID-19-specific molecular signature; severe and non-severe forms of COVID-19 differ with regard to amino acid metabolism and the expression of acute phase proteins [8]. Breath analysis is an innovative, non-invasive, real-time point-of-care technique for detecting volatile organic compounds (VOCs) with potential for use in diagnosis and large-scale screening [9,10]. Thousands of VOCs have been identified in human breath following infectious, inflammatory or pathological events [11,12]. It has been suggested that the analysis of exhaled breath can be used to diagnose tuberculosis, invasive fungal infections, and bacterial colonization of the respiratory tract [13–16], together with ARDS and ventilator-associated pneumonia in patients in the intensive care unit (ICU) [17–22]. Likewise, previous studies have suggested that VOC analysis is of value in the diagnosis of viral infections in patients with chronic obstructive pulmonary disease and of influenza infections in a swine model [23,24]. The airway and lung damage caused by SARS-CoV-2 [25] might conceivably result in the release of characteristic VOCs in the exhaled breath. To test this hypothesis, we determined the metabolomic breath signature in a group of ARDS patients with or without COVID-19 and who required invasive mechanical ventilation.

Methods

Study 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 commitee (*Comite 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) PaO$_2$/FiO$_2$ $\leq$ 300 mm Hg, and positive end expiratory pressure (PEEP) $\geq$ 5 cm H$_2$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. H$_3$O$^+$ was used as the primary ion and the instrument settings were as follows: source voltage, 120 V; drift tube pressure, 3•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_3O^+, $H_2O\text{-}H_3O^+$, 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•05) 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 25th and June 25th, 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 which 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•08, 111•12, 135•09, and 143•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

Table 1
Patient characteristics and treatments

	COVID-19 ARDS	Non-COVID-19 ARDS	*p* value
Number of patients *(n)*	28	12	-
Males/females *(n)*	20/8	6/6	0•28
Age (years)	61 [55-72]	72 [54-79]	0•75
Body weight (kg)	80•0 [66•6-87•6]	86•5 [65•3-94•1]	0•71
Height (cm)	170 [164-175]	173 [169-175]	0•55
Body mass index (kg/m^2)	26•3 [23•7-32•4]	28•9 [23•0-30•9]	0•79
SAPS II score in the first 24 hours	62 [49-68]	46 [40-57]	0•051
SOFA score in the first 24 hours	11 [7-12]	8 [5-12]	0•37
Comorbidities: (*n* (%))			
26- high blood pressure	11 (39)	6 (50)	0•73
26- chronic obstructive pulmonary disease	2 (7)	1 (8)	>0•99
26- ischemic cardiac disease	5 (18)	3 (25)	0•68
26- cancer	2 (7)	3 (25)	0•15
Treatments before admission: (*n* (%))			
26- glucocorticoids	1 (4)	3 (25)	0•073
26- conversion enzyme inhibitors	5 (18)	1 (8)	0•54
26- angiotensin antagonists	2 (7)	2 (16)	0•57
Interventions after admission: (*n* (%))			
26- catecholamines	17 (61)	4 (33)	0•17
26- renal replacement therapy	9 (32)	0 (0)	0•038
Treatments after admission: (*n* (%))			
26- hydroxychloroquine	27 (96)	1 (8)	<0•0001
26- remdesivir	2 (7)	0 (0)	>0•99
26- lopinavir/ritonavir	7 (25)	0 (0)	0•081
26- glucocorticoids	11 (39)	6 (50)	0•73
26- fludrocortisone	1 (4)	4 (33)	0•022
26- eculizumab	12 (43)	4 (33)	0•73
Body temperature at first sample (°C)	37•4 [36•5-38•3]	37•3 [36•8-37•8]	0•84
Respiratory rate at first sample (breaths per min)	26 [25-28]	20 [18-23]	<0•0001
Tidal volume at first sample (mL)	420 [400-475]	438 [400-490]	0•99
Fraction of inspired oxygen at first sample (%)	80 [50-100]	48 [31-68]	0•007
Positive end-expiratory pressure at first sample (cm H$_2$O)	10 [8-13]	5•5 [5-8]	0•0002
Serum creatinine at first sample (μM)	74 [56-137]	67 [44-86]	0•30
Serum C-reactive protein at first sample (mg/L)	195 [175-268]	76 [23-119]	0•002

Continuous data are presented as the median [IQR].

four compounds at *m/z* 99•08, 111•12, 135•09, and 143•15 were respectively methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal.

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) [27,28] measured during the first 24 h in the ICU (Table 2, |r| < 0•4).

Discussion

This study provided proof of concept for the measurement of VOCs and the determination of a specific VOC breathprint in the exhaled breath from patients with COVID-19-related ARDS requiring invasive me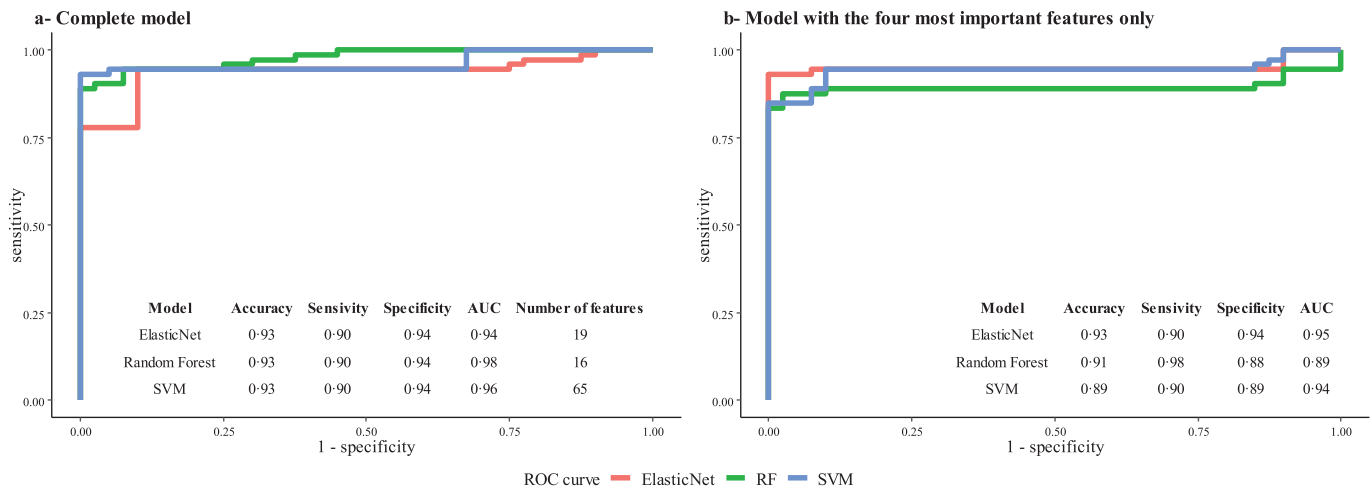chanical ventilation in the ICU. This breathprint was independent of the severity of illness and the viral load. Four VOCs (methylpent-2-enal, 2,4-octadiene 1-chloroheptane, and nonanal) may discriminate between COVID-19 and non-COVID-19 ARDS.

We applied a highly sensitive, rapid, non-invasive, real-time mass spectrometry breath analysis [37,38]. This contrasts with offline technologies, which require a sampling step and remote, time-consuming analytical steps [21,22]. 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) [20].

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

a- PCA

b- OPLS-DA

COVID-19 disease — - — +
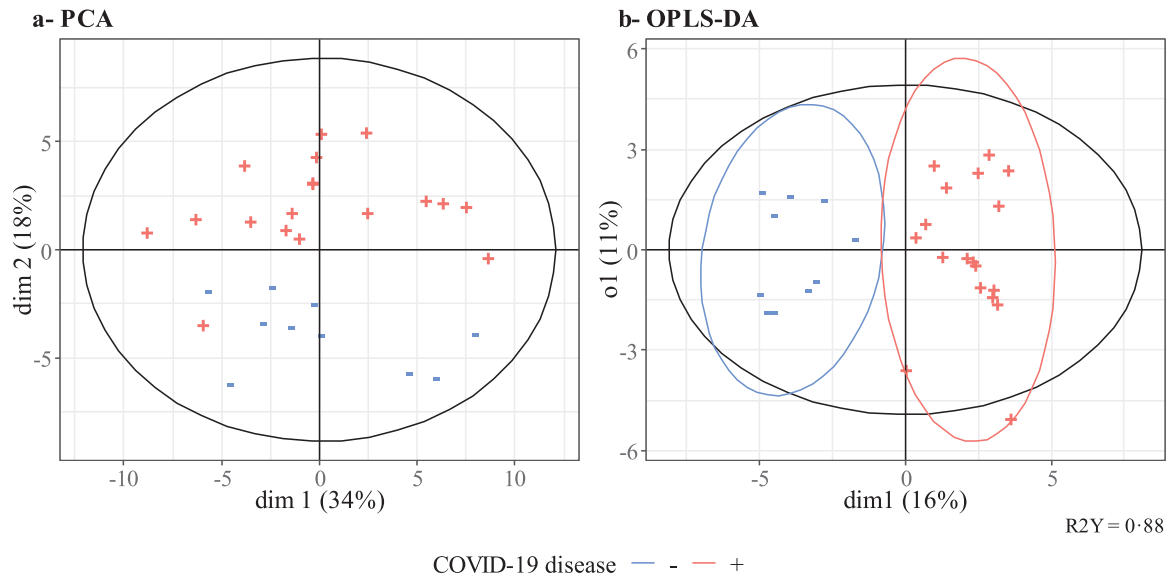
Fig. 1. Multivariate analysis. Principal component analysis (left) and orthogonal partial least squares - discriminant analysis (right) of the breath signature in intubated, mechanically ventilated ICU patients with a positive (red) or negative (blue) PCR test for SARS-CoV-2.

a- Complete model

b- Model with the four most important features only

Model	Accuracy	Sensitivity	Specificity	AUC	Number of features
ElasticNet	0·93	0·90	0·94	0·94	19
Random Forest	0·93	0·90	0·94	0·98	16
SVM	0·93	0·90	0·94	0·96	65

Model	Accuracy	Sensitivity	Specificity	AUC
ElasticNet	0·93	0·90	0·94	0·95
Random Forest	0·91	0·98	0·88	0·89
SVM	0·89	0·90	0·89	0·94

ROC curve — ElasticNet — RF — SVM

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%.

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 (offline *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

a- First acquisition per patient
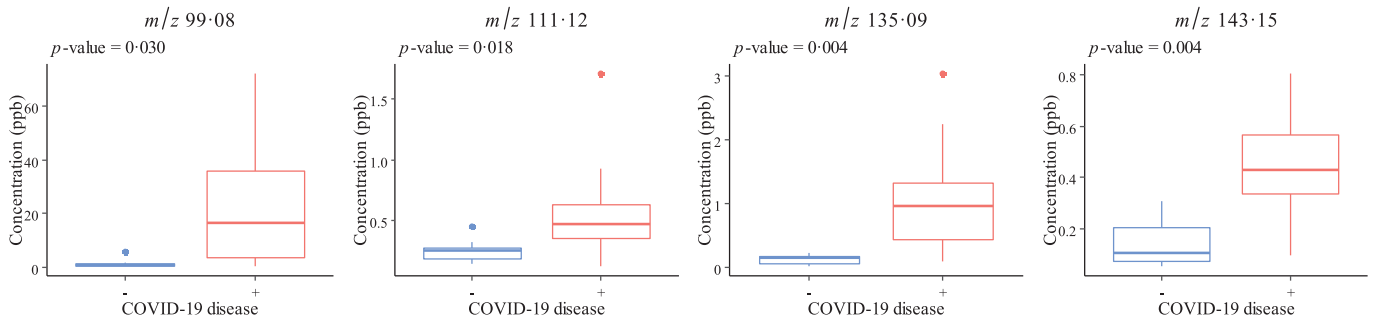
b- Longitudinal modeling
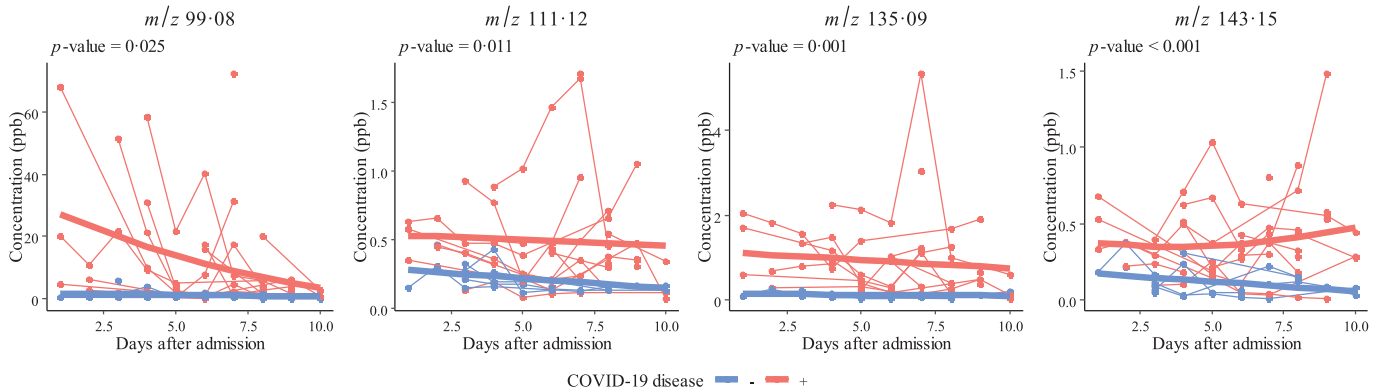
Fig. 3. Longitudinal analysis of VOCs in expired breath. The four features (*m/z* 99•08, 111•12, 135•09, and 143•15) contributing the most to the models were assessed in the first sample available for each patient (a) and over time (b) during the ICU stay for intubated, mechanically ventilated patients with COVID-19 ARDS (in red, n = 28) or non-COVID-19 ARDS (in blue, n = 12). All the points for a given patient are connected, and the bold lines correspond to the fixed effect of the mixed model for each group. *p*-values come from a Wilcoxon test (a) and an F-test (b).

Table 2
Correlations between VOC concentrations and the SAPS II, SOFA score and viral load.

VOC (*m/z*)	SAPS II score		SOFA score		Viral load	
	r	*p*-value	r	*p*-value	r	*p*-value
99•08	0•04	0•88	0•36	0•13	0•08	0•70
111•12	0•02	0•93	0•28	0•25	-0•14	0•48
135•09	0•05	0•85	0•35	0•14	-0•0004	1•00
143•15	0•12	0•62	0•27	0•25	-0•23	0•24

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 analyzed 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., P.D., 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/camilleroquencourt/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

1 COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://coronavirus.jhu.edu/map.html (accessed October 19th, 2020).
2 Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA 2020;324(8):782–93.
3 Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of COVID-19 - final report. N Engl J Med 2020.
4 Recovery_Collaborative_Group, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with COVID-19 - preliminary report. N Engl J Med 2020.
5 Lucas C, Wong P, Klein J, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature 2020;584(7821):463–9.
6 Hadjadj J, Yatim N, Barnabei L, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science 2020;369(6504):718–24.
7 Kuri-Cervantes L, Pampena MB, Meng W, et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. Sci Immunol 2020;5(49).
8 Shen B, Yi X, Sun Y, et al. Proteomic and metabolomic characterization of COVID-19 patient sera. Cell 2020;182(1) 59-72 e15.
9 Kataoka H, Saito K, Kato H, Masuda K. Noninvasive analysis of volatile biomarkers in human emanations for health and early disease diagnosis. Bioanalysis 2013;5 (11):1443–59.
10 Rattray NJ, Hamrang Z, Trivedi DK, Goodacre R, Fowler SJ. Taking your breath away: metabolomics breathes life in to personalized medicine. Trends Biotechnol 2014;32 (10):538–48.
11 Amann A, de Lacy Costello B, Miekisch W, et al. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. J Breath Res 2014;8(3):034001.
12 de Lacy Costello B, Amann A, Al-Kateb H, et al. A review of the volatiles from the healthy human body. J Breath Res 2014;8(1):014001.
13 Koo S, Thomas HR, Daniels SD, et al. A breath fungal secondary metabolite signature to diagnose invasive aspergillosis. Clin Infect Dis 2014;59(12):1733–40.
14 Nakhleh MK, Jeries R, Gharra A, et al. Detecting active pulmonary tuberculosis with a breath test using nanomaterial-based sensors. Eur Respir J 2014;43(5):1522–5.
15 Coronel Teixeira R, Rodriguez M, Jimenez de Romero N, et al. The potential of a portable, point-of-care electronic nose to diagnose tuberculosis. J Infect 2017;75(5):441–7.
16 Suarez-Cuartin G, Giner J, Merino JL, et al. Identification of pseudomonas aeruginosa and airway bacterial colonization by an electronic nose in bronchiectasis. Respir Med 2018;136:111–7.
17 Schnabel R, Fijten R, Smolinska A, et al. Analysis of volatile organic compounds in exhaled breath to diagnose ventilator-associated pneumonia. Sci Rep 2015;5:17179.
18 Filipiak W, Beer R, Sponring 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(1):016004.
19 Schnabel RM, Boumans ML, Smolinska A, et al. Electronic nose analysis of exhaled breath to diagnose ventilator-associated pneumonia. Respir Med 2015;109(11):1454–9.
20 Bos LD, Schultz MJ, Sterk PJ. Exhaled breath profiling for diagnosing acute respiratory distress syndrome. BMC Pulm Med 2014;14:72.
21 Bos LD, Weda H, Wang Y, et al. Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. Eur Respir J 2014;44 (1):188–97.
22 Schubert JK, Muller WP, Benzing A, Geiger K. Application of a new method for analysis of exhaled gas in critically ill patients. Intensive Care Med 1998;24(5):415–21.
23 van Geffen WH, Bruins M, Kerstjens HA. Diagnosing viral and bacterial respiratory infections in acute COPD exacerbations by an electronic nose: a pilot study. J Breath Res 2016;10(3):036001.
24 Traxler S, Bischoff AC, Sass R, et al. VOC breath profile in spontaneously breathing awake swine during Influenza A infection. Sci Rep 2018;8(1):14857.
25 Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in COVID-19. N Engl J Med 2020;383(2):120–8.
26 Force ADT, Ranieri VM, Rubenfeld GD, et al. Acute respiratory distress syndrome: the Berlin definition. JAMA 2012;307(23):2526–33.
27 Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270 (24):2957–63.
28 Vincent JL, Moreno R, Takala J, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the european society of intensive care medicine. Intensive Care Med 1996;22(7):707–10.
29 Trefz P, Pugliese G, Brock B, Schubert JK, Miekisch W. Effects of elevated oxygen levels on VOC analysis by means of PTR-ToF-MS. J Breath Res 2019;13(4):046004.
30 Hansel A, Jordan A, Holzinger R, Prazeller P, Vogel W, Lindinger W. Proton transfer reaction mass spectrometry: on-line trace gas analysis at the ppb level. Int J Mass Spectrom Ion Process 1995;149-150:609–19.
31 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57(1):289–300.
32 Breiman L. Random forests. Machine Learning 2001;45(1):5–32.
33 Thevenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. J Proteome Res 2015;14(8):3322–35.
34 Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V. Feature selection for SVMs. Adv Neural Inform Process Syst 2000;13.
35 Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B 2005;67(2):301–20.
36 Pihur V, Datta S, Datta S. RankAggreg, an R package for weighted rank aggregation. BMC Bioinformatics 2009;10:62.
37 Trefz P, Schmidt M, Oertel P, et al. Continuous real time breath gas monitoring in the clinical environment by proton-transfer-reaction-time-of-flight-mass spectrometry. Anal Chem 2013;85(21):10321–9.
38 Brock B, Kamysek S, Silz J, Trefz P, Schubert JK, Miekisch W. Monitoring of breath VOCs and electrical impedance tomography under pulmonary recruitment in mechanically ventilated patients. J Breath Res 2017;11(1):016005.
39 van de Kant KD, van Berkel JJ, Jobsis Q, et al. Exhaled breath profiling in diagnosing wheezy preschool children. Eur Respir J 2013;41(1):183–8.
40 Corradi M, Pignatti P, Manini P, et al. Comparison between exhaled and sputum oxidative stress biomarkers in chronic airway inflammation. Eur Respir J 2004;24 (6):1011–7.
41 Rahman I. Oxidative stress, chromatin remodeling and gene transcription in inflammation and chronic lung diseases. J Biochem Mol Biol 2003;36(1):95–109.