ERJ open research
THE BEST IN OPEN ACCESS BASIC, TRANSLATIONAL & CLINICAL RESPIRATORY RESEARCH

Early View

Original article

Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography-mass spectrometry

Wadah Ibrahim, Rebecca L. Cordell, Michael J. Wilde, Matthew Richardson, Liesl Carr, Ananga Sundari Devi Dasi, Beverley Hargadon, Robert C. Free, Paul S. Monks, Christopher E. Brightling, Neil J. Greening, Salman Siddiqui

Please cite this article as: Ibrahim W, Cordell RL, Wilde MJ, *et al*. Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography-mass spectrometry. *ERJ Open Res* 2021; in press (https://doi.org/10.1183/23120541.00139-2021).

Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography-mass spectrometry

Wadah Ibrahim[1,2,*], Rebecca L. Cordell[3,*], Michael J. Wilde[3], Matthew Richardson[1,2], Liesl Carr[1,2], Ananga Sundari Devi Dasi[1,2], Beverley Hargadon[1,2], Robert C. Free[1,2], Paul S. Monks[3], Christopher E. Brightling[1,2], Neil J. Greening[1,2] , Salman Siddiqui[1,2], on behalf of the EMBER consortium

[1]Department of Respiratory Sciences, University of Leicester, Leicester, UK

[2]Institute for Lung Health, NIHR Leicester Biomedical Research Centre (Respiratory theme), Glenfield Hospital, Groby Road, Leicester LE3 9QP

[3] School of Chemistry, University of Leicester, Leicester, UK

*Equal contribution to work

Correspondence:

Professor Salman Siddiqui BM, FRCP, PHD

Clinical Professor of Airways Disease
Chief Investigator EMBER molecular pathology node & NIHR BEAT-Severe Asthma
National Respiratory Stratified Medicines Champion

College of Life Sciences, University of Leicester
NIHR Biomedical Research Centre (Respiratory theme)
Institute for Lung Health, Leicester

Email: ss338@le.ac.uk
t (direct): +44 116 250 2873
t (secretary): +44 116 256 3841

Abstract

Background: The ongoing COVID-19 pandemic has claimed over two and a half million lives worldwide so far. SARS-CoV-2 infection is perceived to be seasonally recurrent and a rapid non-invasive biomarker to accurately diagnose patients early-on in their disease course will be necessary to meet the operational demands for COVID-19 control in the coming years.

Objective: To evaluate the role of exhaled breath volatile biomarkers in identifying patients with suspected or confirmed COVID-19 infection, based on their underlying PCR status and clinical probability.

Methods: A prospective, real-world, observational study recruiting adult patients with suspected or confirmed COVID-19 infection. Breath samples were collected using a standard breath collection bag, modified with appropriate filters to comply with local infection control recommendations and samples were analysed using gas chromatography-mass spectrometry (TD-GC-MS).

Findings: 81 patients were recruited between April 29th to July 10th, 2020, of whom 52/81 (64%) tested positive for COVID-19 by RT-PCR. A regression analysis identified a set of seven exhaled breath features (benzaldehyde, 1-propanol, 3, 6-methylundecane, camphene, beta-cubebene, Iodobenzene, and an unidentified compound) that separated PCR positive patients with an area under the curve (AUC): 0.836, sensitivity: 68%, specificity: 85%.

Conclusions: GC-MS detected exhaled breath biomarkers were able to identify PCR positive COVID-19 patients. External replication of these compounds is warranted to validate these results.

Introduction:

The ongoing COVID-19 pandemic has claimed over two and a half million lives worldwide so far by March 2021 [1]. Rapid and accurate diagnostic testing of COVID-19 is invaluable to allow rapid isolation, healthcare pathways and access to definitive therapies. Final diagnosis of COVID-19 infection relies heavily on real-time-reverse-transcriptase polymerase chain reaction (rRT-PCR) positivity [2] as well as clinical symptoms, radiological features and blood biomarkers. The false-negative rate for SARS-CoV-2 RT-PCR remains highly variable with rates reaching up to 67% false negative within the first 5 days post exposure [3]. It is strongly recommended that PCR results are carefully interpreted and correlated with pre-test probability, clinical symptoms, blood biomarkers, and digital pathology, particularly when used as basis for removing precautions put in place to prevent onward transmission.

There are several tests that offer increased turnaround at the expense of reduced sensitivity, these include Rapid Antigen Tests and Lateral Flow Assays [4, 5]. The trade off in sensitivity is justified when mass high throughput testing may be required at frequent intervals e.g. at airports, schools and hospitals, however, this has important implications for patients, healthcare professionals, and COVID-19 policymakers, highlighting the imperative of developing early, non-invasive, more sensitive diagnostic tools.

Exhaled breath analysis has attracted notable scientific and clinical interest in recent years. Volatile organic compounds (VOCs) have the potential to mirror various metabolic processes both locally within the respiratory system and systemically, via blood circulation [6, 7]. VOCs have been utilised as diagnostic, prognostic and treatment response biomarkers for various respiratory illness, including infections [8-12]. The rapid, cost effective and non-invasive

nature make VOCs a strong candidate as a potential COVID-19 biomarkers, substantiated by preliminary studies highlighting its diagnostic potential [13, 14].

In this pilot study, we evaluated exhaled breath volatile compounds as potential biomarkers for COVID-19 infection, based on the underlying PCR status and clinical probability.

Materials and methods:

Study design and participants:

This was a prospective, real-world observational study conducted at Leicester, United Kingdom. The National Research Ethics Service Committee East Midlands approved the study protocol (REC number: 16/LO/1747). Patients admitted to hospital with suspected COVID-19 infection were approached and provided informed consent. Those who agreed to take part, provided a breath samples within 24 hours of admission to hospital and samples were analysed using thermal desorption coupled gas chromatography-mass spectrometry (TD-GC-MS).

All participants underwent testing for SARS-CoV-2 using PCR of nasopharyngeal swab, as part of the local streaming and clinical care pathways.

Clinical probability:

In addition to PCR testing participants were classified by clinical probability of COVID-19 infection. This was with the aim of capturing patients with false negative PCR swabs, or those with delayed presentation who had pneumonitis but no longer shedding viral load.

We classified the study subjects as high clinical probability of SARS-CoV-2 infection if they fulfilled either 2 major or 1 major and 3 minor diagnostic criteria **(Table 1).**

These criteria were contrived by a panel of senior clinicians, supported by peer-reviewed published evidence [15-21].

Exhaled breath sample collection:

Breath sampling was optimised to minimise cross-contamination and infection risk. The breath sampling apparatus included a mouthpiece with integrated HEPA filter (GVS) and a t-piece with directional valves (Intersurgical), connected to a polypropylene shutoff valve with 1/4" push connections (RS Components) via a 22F-10 mm connector (Intersurgical). The valve was connected on the other side, via a piece of 1/4" PFA tubing of minimal length, to the large port of a 3 L Tedlar bag with dual stainless steel fittings with a piece of pre-conditioned silicon tubing surrounding the fitting to ensure a secure connection. The tubing and shut-off valves were pre-conditioned by heating for 4 hours at 60°C and the Tedlar bags flushed with nitrogen and vacuum evacuated six times prior to sampling to reduce background VOCs. The pre-assembled, single-use device was handed to the patient at bedside and patients were asked to perform tidal breathing, opening the shut-off valve when providing the sample and closing it again once the bag was inflated. Average sampling time was less than a minute.

The mouthpiece and all connections above the shut-off valve were disconnected and disposed on the ward. The remaining bag and valve assembly were wiped down and double bagged before being taken immediately for loading onto sorbent tubes in triplicate (Carbograph 1TD, Markes International). The unused fitting on the Tedlar bag was

connected to a handheld Elf pump (Escort® Elf, Sigma Aldrich, Dorset, UK) using sealed Tygon tubing, and the sample within the bag drawn through the sorbent tube under negative pressure at a flow rate of 0.5 L/min for two minutes, loading a total of 1 L of breath onto each of the three sorbent tubes. The sorbent tubes were sealed with brass caps (Swagelok) and stored in the fridge for up to 48 hours before being transferred to the lab for analysis. Blank samples were taken from Tedlar bags filled with air (BTCA grade, BOC) at the beginning of each sampling week.

Sample analysis:

All sample tubes were loaded with internal standard (0.6 µL of 20 µg/mL toluene-d_8, phenanthrene-d_{10} (Sigma Aldrich) and n-octane-d_{18} (Cambridge Isotope Laboratories) mixture in methanol (Sigma Aldrich) in a stream of nitrogen at a flow rate of 100 mL/min for 2 min, purging off the excess solvent.

Reference standards:

A 100 µg/mL multi component air standard (p/n 4M7537-U Sigma Aldrich, Dorset, UK) was used to monitor instrument performance and for peak identity confirmation, and was diluted in methanol to give a final concentration of 10 µg/mL. 1 µl of standard was loaded onto sorbent tubes into a stream of N_2 (zero grade, BOC) at 100 ml/min and purged for 1 minute, before addition of internal standard as described above. Benzaldehyde, 1-propanol and octanal (Sigma Aldrich) and SPEX Certiprep Can-Terp Mix (Fisher Scientific, Loughborough, UK) standards were diluted in the same way and loaded onto tubes to confirm the identity of additional selected peaks.

A reference solution used to monitor retention behaviour was prepared from a 40 mg/L C_8-C_{20} saturated alkanes certified reference material (Sigma Aldrich) combined with a 2000 µg/mL aromatics calibration standard (NJDEP EPH 10/08 Rev.2, Thames Restek). The mixture was diluted in methanol (SupraSolv grade, Sigma Aldrich) to give final concentrations of 10 µg/mL and 20 µg/ml for n-alkanes and aromatics, respectively.

TD-GC-MS Analysis

GC-MS analysis was carried out on an Agilent 7820A with 5977B MS (Agilent Technologies Ltd, Stockport, UK). A DB-5MS capillary column 60 m x 0.25 mm x 0.25 µm (from Agilent Technologies) was used with a 1 ml/min column flow rate using helium (N6.0, BOC) as a carrier gas. The GC conditions were as follows: the column starting temperature was 35 °C which was then raised to 130 °C at 2.8 °C/min, to 220 °C at 4 °C/min and then to 320°C at 25 °C/min, where it was held for a further 10 min.

The mass spectrometer was operated in full scan mode from 40-350 amu. The transfer line to the mass spectrometer was heated to 300 °C; the source temperature was maintained at 230 °C and the quadrupole at 150 °C.

The instrument was interfaced with a Markes TD-100 xr thermal desorption unit (Markes International Ltd, Llantrisant, UK). Tubes were pre-purged with carrier gas for 1 min at 50 mL/min and then desorbed at 300 °C for 5 min with a flow of 50 mL/min onto a 'hydrophobic, general purpose' trap (Markes International Ltd, Llantrisant, UK) held at −10 °C. The trap was then purged for 2 min at 2 mL/min before being heated at the maximum heating rate to 300 °C for 5 min, with a spit flow of 2 ml/min.

Data analysis:

A custom library was built in Unknowns (Agilent Technologies) from random selection of ten patient samples, covering all patient groups, based on their age, PCR status and clinical probability. From this library a method was built in MassHunter Quantitative analysis (Agilent Technologies) using the base peak as quantifier and two qualifier ions for identity confirmation. All peaks were manually screened to ensure correct integration, and siloxane peaks were removed prior to further analysis. Peaks were removed if <60% of samples did not exceed the mean blank integrated peak area plus three standard deviations. Three patient samples were discarded owing to the high levels of exogenous contamination.

Statistical analysis

Feature selection was performed by applying LASSO regression to the final VOCs peak table with the dependent variable as PCR status first, then clinical probability, extracting a set of relevant features for each model.

A logistic regression model was then fitted with the dependent variable as PCR status first, then clinical probability; with the independent variables set as the respective relevant features selected by the LASSO model above. Partial Least Squares Discriminant Analysis (PLS-DA) and Principal Component Analysis (PCA) were then applied to the two sets of features, with the dependent variable as SARS-CoV-2 PCR status and clinical probability respectively.

All analyses were performed using R 4.0.0 [22]. LASSO regression was performed using glmnet [23], PCA was performed using the mixOmics library [24], and PLS-DA was performed using the ropls library [25].

Results

From April 29th to July 10th, 2020, a total of 81 adult participants, mean age (SD): 56.5 (15.1), 59% males, with suspected or confirmed COVID-19 infection were recruited. Details of all study participants and their clinical characteristics are detailed in (**Table 2**).

Identification of SARS-CoV-2 in breath by PCR status:

52 (64%) of the participants had a positive PCR test at the time of admission **(Table 2).**

A set of seven features were extracted that had non-zero regression coefficients in at least 70 out of 100 runs of 10-fold cross validation of the LASSO model. Compound identities were confirmed using the Metabolomics Standards Initiative (MSI) [26]. These were: the oxygenates benzaldehyde, 1-propanol (both MSI level 1); a hydrocarbon, 3,6-dimethylundecane (MSI level 2); and two terpenes, camphene and beta-cubebene (MSI level 1 and 2 respectively). Iodobenzene was also extracted, which is likely of exogenous origin, and a compound that was unable to be identified owing to co-elution with a much larger peak.

The distribution of the first discriminant function and the first two principal components are shown in **(Figure 1a-b).** Wilcoxon test was applied to the first discriminant function and the first PCA scores **(Figure 1c-d).**

The AUC for the first discriminant function score was 0.83 (95% CI: 0.74-0.92), sensitivity was 0.68 (95% CI 0.55-0.80), specificity was 0.85 (95% CI 0.72-0.98), positive predictive value (PPV) was 0.89 (95% CI 0.79-0.99) and negative predictive value (NPV) was 0.60 (95% CI 0.44-0.75) **(Figure 1e)**. The AUC for the first PCA was 0.79 (95% CI: 0.69-0.90), sensitivity was 0.70 (95% CI 0.57-0.82), specificity was 0.78 (95% CI 0.63-0.93), PPV was 0.85 (95% CI 0.74-0.96) and NPV was 0.59 (95% CI 0.43-0.75) **(Figure 1f).**

Identification of SARS-CoV-2 in breath by clinical probability:

55/81 (68%) subjects were classified as 'high clinical probability' for COVID-19 of whom 38/55 (69%) have subsequently had positive PCR results.

A set of 11 features were extracted that had non-zero regression coefficients in at least 10 out of 100 runs of 10-fold cross validation of the LASSO model. The feature selection threshold of 10 out 100 runs of 10-fold cross validation is weak, suggesting that these features are less likely to be stable.

The 11 features were a mixture of hydrocarbons - cyclohexene (identified to MSI level 2), 3-heptene (MSI level 2), pentadecane (MSI level 1) and 4-ethenyl-1,2-dimethyl-benzene, (MSI level 2); oxygenates - octanal (MSI level 1), benzaldehyde (MSI level 1), 2,2-dimethyl 1-

propanol (MSI level 2), 1-propanol (MSI level 1), acetoin (MSI level 2) and acetic acid methyl ester (MSI level 2); and the chloro-carbon tetrachloroethylene (MSI level 2).

The distribution of the first discriminant function and the first two principal components are shown in **(Figure 2a-b).** Wilcoxon test was applied to the first discriminant function and the first PCA scores **(Figure 2c-d).**

The AUC for the first discriminant function score was 0.65 (95% CI: 0.52-0.78), positive predictive value (PPV) was 0.82 (95% CI 0.70-0.95) and negative predictive value (NPV) was 0.41 (95% CI 0.27-0.56) **(Figure 2e)**. The AUC for the first PCA was 0.55 (95% CI: 0.69-0.90), PPV was 0.81 (95% CI 0.69-0.92) and NPV was 0.44 (95% CI 0.28-0.60) **(Figure 2f).**

Discussion:

In this pilot study we evaluated the validity of using exhaled breath VOCs in identifying patients with COVID-19 infection based on their underlying PCR status and clinical probability, using GC-MS. This study provided proof of concept for exhaled breath measurement in patients with severe COVID-19 infection, while maintaining infection-control standard precautions. Furthermore, it demonstrated that VOC biomarker profiling can identify COVID-19 patients based on their underlying PCR status with good accuracy.

This study is the first to incorporate GC-MS, the gold standard in breath/VOC discovery analysis, recruiting the largest number of COVID-19 patients to date using an advanced offline technology. The high analytical standards applied in our study and the increased peak capacity resulted in the detection of 389 feature that exceed background levels by 3 standard deviations in 60% of samples.

With the current level of understanding of COVID-19 pathophysiology, there are several indicators that a SARS-CoV-2 infection would be detectable in the exhaled breath VOC pattern. The disease has been reported to cause a systemic inflammatory response [27], which supports the hypothesis that metabolism is influenced in more than one way and the distribution of breath volatile metabolites is substantially changed [28].

The discriminatory features identified in our study belonged to various chemical groups and, with the exception of the unidentified peak, were present in higher concentrations in the breath of PCR positive patients. These included 1-propanol, a previously observed marker of bacterial pneumonia [29], lung cancer [30, 31] and asthma [10]; and 3,6-dimethylundecane a likely derivative of lipid peroxidation and potentially background air [32]. The concentrations in breath of numerous alkanes have been reported to have altered levels in a range of respiratory diseases including asthma [10, 33] and COPD [34].

Benzaldehyde is normally considered to be an exogenous VOC and is ubiquitous in indoor air. However, studies have shown that it can be useful as a breath biomarker [35, 36]. The two terpenes, camphene and beta-cubebene, are also generally considered to be exogenous in nature, however, their differential metabolism within certain disease groups has shown promising relevance as a diagnostic aid [34, 37].

Two other published studies examined the diagnostic potential of volatile breath markers in COVID-19. Ruszkiewicz *et al* [13] were able to differentiate COVID-19 patients from other cardio-respiratory conditions using a compact GC coupled to ion mobility spectrometry (GC-IMS) in 31 COVID-19 patients from a total of 98 participants across two centres. Multivariate analysis identified 7 VOCs from a total of 80 features across all samples, including aldehydes (ethanal, octanal), ketones (acetone, butanone), and methanol as the main compounds

driving this separation. In an attempt to validate these compounds in our cohort, a logistic regression model was fitted with the dependant variable as PCR positive/negative and the independent variables as the identified compounds from the Ruszkiewicz paper. None of the identified compounds held any significant discriminatory value in our cohort. This was not unexpected given the inherent differences in recruitment strategy, sampling procedure, ionisation selectivity and sensitivity of the two techniques, without prior optimisation. Grassin-Delyle *et al* [38] used real-time, online, proton transfer reaction time-of-flight mass spectrometry (PTR-MS) to measure breath VOCs of 40 ventilated patients with severe COVID-19 or non-COVID-19 acute respiratory distress syndrome, diagnosed by PCR. The prominent VOCs observed in COVID-19 patients were assigned to methylpent-2-enal, 2,4-octadiene, 1-chloroheptane and nonanal, however, the lack of chromatographic separation in this technique makes definitive chemical speciation and subsequently, meaningful study comparison, difficult to achieve.

While RT-PCR-SARS-CoV-2 remains the most widely used COVID-19 diagnostic test, its limited sensitivity, particularly in the early course of the disease, illustrates its inadequacy as an isolated COVID-19 diagnostic test [39-41]; which highlights the importance of incorporating clinical probability as an essential component in making a definitive diagnosis. Nonetheless, the peculiarities of SARS-CoV-2 infection do not allow for the development of undisputed clinical criteria for a COVID-19 diagnosis, making our VOC biomarker results, based on clinical probability alone, difficult to interpret.

The diagnostic value of breath VOCs is expectedly limited in high prevalence cohorts with increased pre-test probability. However, once appropriately validated in a low COVID prevalence cohort, exhaled breath VOCs promise to be a useful tool in ruling out COVID-19 infection at the point of admission. This would facilitate extrication from isolation spaces, preserving infection control resources and preventing onward nosocomial transmission.

There are limitations associated with this study that must be taken into consideration when interpreting the findings. Firstly the sample size was relatively small: patient sampling during the first wave was largely dependent on clinical practicality and increasing the sample size was not possible owing to reduced hospital admissions in response to Public Health England recommendations. Secondly, the reference standard test used was RT-PCR: whilst PCR was the best available diagnostic test at the time of testing it will have still resulted in some false-negative cases. It is also worth noting that with relatively small-size and complex datasets, overfitting can lead to poor model performance however, to mitigate for this, feature selection was carried out by the application of LASSO regression to the peak table. External replication of these results is needed in similar populations with suspected severe COVID-19 infection to validate these findings.

Conclusions:

This proof-of-concept study demonstrated the potential role of breath testing in COVID-19 diagnostics. The next stage is a large-scale diagnostic accuracy study in a population with a more realistic prevalence of COVID-19 infection to determine the clinical applicability of such a test.

Table 1: SARS-CoV-2 infection diagnostic criteria

Major criteria	Minor criteria
Presence of new bilateral peripheral, middle and lower pulmonary infiltrates on chest X-ray	Low blood Lymphocyte count below the lower limit of normal (<1.0 $x10^9$/L)
New continuous cough, high temperature, and loss of sense of smell and or taste	Low blood Eosinophil count below the lower limit of normal (<0.04 $x10^9$/L)
	High C-reactive protein level (>50 mg/L)
	New oxygen requirements
	Contact with confirmed COVID-19 household

Table 2: clinical characteristics of adult patients recruited with suspected or confirmed COVID-19 infection.

	All patients	PCR positive	PCR negative
	(N=81)	(N=52)	(N=29)
Baseline demographics			
Age (years) - mean ± (SD)	56.5 (15.1)	53.6 (14.5)	61.7 (14.9)
Male – number/total (%)	48/81 (59%)	29/52 (55%)	19/29 (65%)
Ethnicity – Caucasian – number/total (%)	42/81 (51%)	19 /52(36%)	23/29 (79%)
Current smoker – number (%)	7/81 (8.6%)	4/52 (7%)	3/29 (10%)
Height (meters) - mean (SD)	1.68 (0.08)	1.66 (0.08)	1.72 (0.07)
Weight (kilograms) - mean (SD)	88.6 (25.8)	87.7 (24.9)	90.2 (27.7)
BMI - mean (SD)	31.0 (8.1)	31.6 ± (7.7)	30.1 (8.7)
Co-morbidities			
Diabetes – number (%)	21/81 (25.9%)	14/52 (27%)	7/29 (24%)
Chronic pulmonary disease – number (%)	13/81 (16.0%)	5/52 (9%)	8/29 (27%)
Cardiac disease – number (%)	6/81 (7.4%)	5/52 (9%)	1/29 (3%)
Admission observations			
Temperature (>37.8c) no. /total no. (%)	25/81 (30.8%)	15/52(28%)	10/29 (34%)
Respiratory rate (>25/min) no. /total no. (%)	13/81 (16.0%)	9/52 (17%)	4/29 (13%)
Requiring supportive oxygen no. /total no. (%)	38/81 (46.9%)	30/52 (57%)	8/29 (27%)
Blood biomarkers			
Serum lymphocyte count (10^9/L) – median (Q1-Q3)	0.98 (0.7 - 1.3)	1.0 (0.7-1.3)	0.8 (0.7-1.2)
Serum eosinophil count (10^9/L) median (Q1-Q3)	0.02 (0.01 – 0.08)	0.02 (0.01-0.06)	0.06 (0.02-0.1)
C-reactive protein (mg/dl) median (Q1-Q3)	71.0 (41.5 – 152.5)	68.5 (47-119)	77.0 (32-228)
Imaging			
Presence of bilateral consolidation - no. /total no. (%)	67/81 (82%)	47/52 (90%)	20/29 (69%)
Length of hospital stay mean (SD)	5.9 (5.4)	6.5 (6.3)	5.0 ± (3.7)
30 days mortality - number (%)	9 (11%)	7 (8%)	2 (2.4%)

Figure legends:

Figure 1:

1(a): Partial Least Squares Discriminant Analysis (PLS-DA) on the seven extracted breath features with the response variable set as PCR status.

1(b): Principal Component Analysis (PCA) on the seven extracted breath features with the response variable set as PCR status.

1(c): Wilcoxon test applied to the first discriminant function**. 1(d):** Wilcoxon test on the first PCA scores

1(e): Receiver operating characteristic (ROC) curve for the discriminant score. The area under the curve (AUC) for the first discriminant function score: 0.83 (95% CI: 0.74-0.92), sensitivity: 0.68 (95% CI 0.55-0.80), specificity: 0.85 (95% CI 0.72-0.98), positive predictive value (PPV): 0.89 (95% CI 0.79-0.99), and negative predictive value (NPV): 0.60 (95% CI 0.44-0.75).

1(f): ROC curves for first principal component score (PC1). The AUC for the first PCA: 0.79 (95% CI: 0.69-0.90), Sensitivity: 0.70 (95% CI 0.57-0.82), Specificity: 0.78 (95% CI 0.63-0.93), PPV: 0.85 (95% CI 0.74-0.96), and NPV: 0.59 (95% CI 0.43-0.75).

Figure 2:

2(a): Partial Least Squares Discriminant Analysis (PLS-DA) on the eleven extracted breath features with the response variable set as clinical probability.

2(b): Principal Component Analysis (PCA) on the eleven extracted breath features with the response variable set as clinical probability.

2(c): Wilcoxon test applied to the first discriminant function**. 2(d):** Wilcoxon test on the first PCA scores

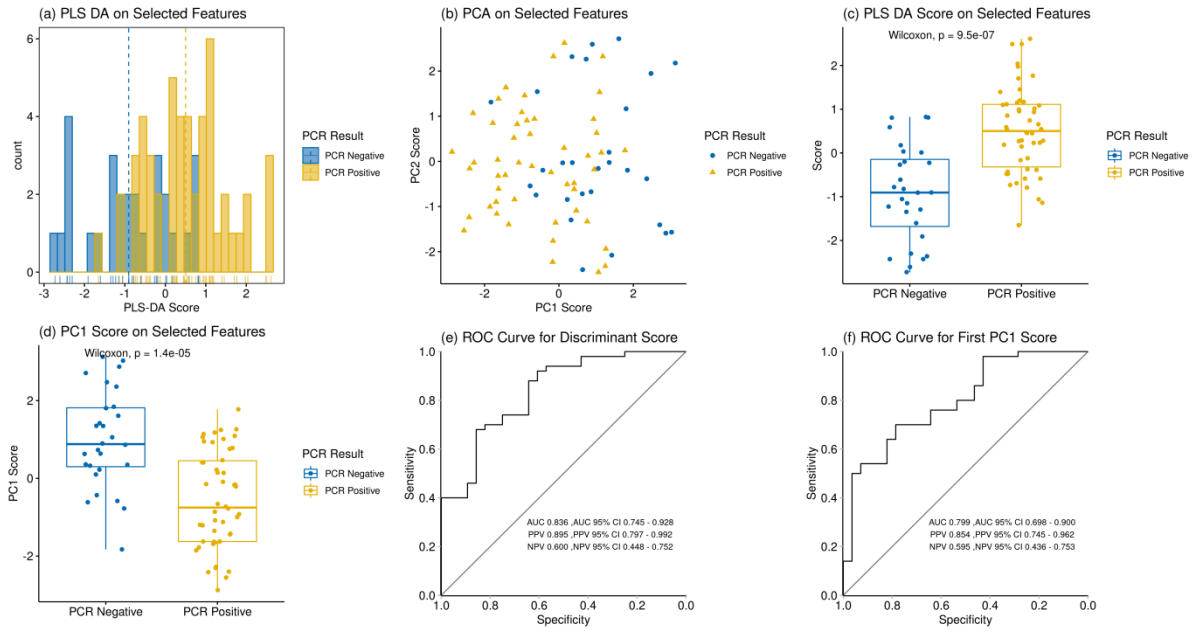
2(e): The AUC for the first discriminant function score, based on clinical probability: 0.65 (95% CI: 0.52-0.78), PPV: 0.82 (95% CI 0.70-0.95), and NPV: 0.41 (95% CI 0.27-0.56).

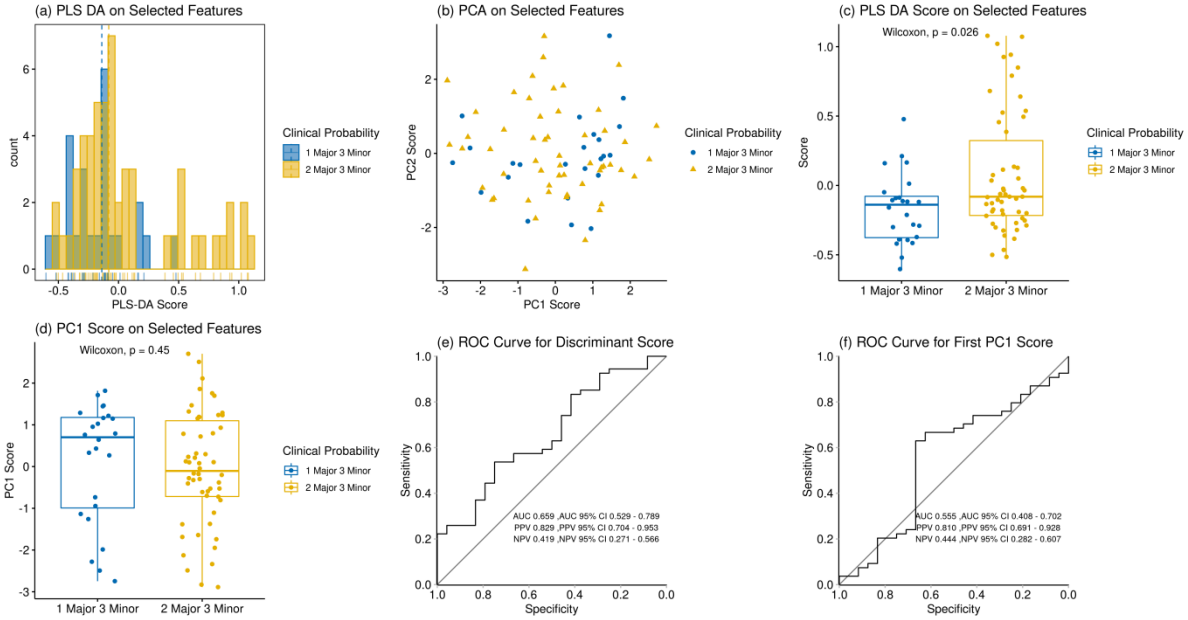
2(f): The AUC for the first PCA was 0.55 (95% CI: 0.69-0.90), PPV was 0.81 (95% CI 0.69-0.92) and NPV was 0.44 (95% CI 0.28-0.60)

References:

1. WHO, *WHO coronavirus disease 2019 (covid-19) situation report-101.* 2020.
2. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.* Euro Surveill, 2020. **25**(3).
3. Kucirka, L.M., et al., *Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure.* Ann Intern Med, 2020.
4. Broughton, J.P., et al., *CRISPR–Cas12-based detection of SARS-CoV-2.* Nature Biotechnology, 2020. **38**(7): p. 870-874.
5. Yamayoshi, S., et al., *Comparison of Rapid Antigen Tests for COVID-19.* Viruses, 2020. **12**(12).
6. Hashoul, D. and H. Haick, *Sensors for detecting pulmonary diseases from exhaled breath.* European Respiratory Review, 2019. **28**(152): p. 190011.
7. van de Kant, K.D.G., et al., *Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review.* Respiratory Research, 2012. **13**(1): p. 117.
8. Neerincx, A.H., et al., *Identification of Pseudomonas aeruginosa and Aspergillus fumigatus mono- and co-cultures based on volatile biomarker combinations.* J Breath Res, 2016. **10**(1): p. 016002.
9. Janssens, E., et al., *Breath analysis by ion mobility spectrometry allows to discriminate COPD from lung cancer patients.* European Respiratory Journal, 2018. **52**.
10. Schleich, F.N., et al., *Exhaled Volatile Organic Compounds are Able to Discriminate between Neutrophilic and Eosinophilic Asthma.* Am J Respir Crit Care Med, 2019.
11. Schleich, F.N., et al., *Exhaled Volatile Organic Compounds are Able to Discriminate between Neutrophilic and Eosinophilic Asthma.* American journal of respiratory and critical care medicine, 2019.
12. Ibrahim, W., et al., *Breathomics for the clinician: the use of volatile organic compounds in respiratory diseases.* Thorax, 2021: p. thoraxjnl-2020-215667.
13. Ruszkiewicz, D.M., et al., *Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry - a feasibility study.* EClinicalMedicine, 2020: p. 100609.
14. Grassin-Delyle, S., et al., *Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study.* EBioMedicine, 2021. **63**: p. 103154.
15. Durrani, M., et al., *Chest X-rays findings in COVID 19 patients at a University Teaching Hospital - A descriptive study.* Pakistan journal of medical sciences, 2020. **36**(COVID19-S4): p. S22-S26.
16. Cadegiani, F.A., et al., *The AndroCoV Clinical Scoring for COVID-19 Diagnosis: A Prompt, Feasible, Costless, and Highly Sensitive Diagnostic Tool for COVID-19 Based on a 1757-Patient Cohort.* Cureus, 2021. **13**(1): p. e12565-e12565.
17. Liu, Z., et al., *Lymphocyte subset (CD4+, CD8+) counts reflect the severity of infection and predict the clinical outcomes in patients with COVID-19.* J Infect, 2020. **81**(2): p. 318-356.
18. Lindsley, A.W., J.T. Schwartz, and M.E. Rothenberg, *Eosinophil responses during COVID-19 infections and coronavirus vaccination.* The Journal of allergy and clinical immunology, 2020. **146**(1): p. 1-7.
19. Luo, X., et al., *Prognostic Value of C-Reactive Protein in Patients With Coronavirus 2019.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2020. **71**(16): p. 2174-2179.

20. Wang, L., *C-reactive protein levels in the early stage of COVID-19.* Médecine et Maladies Infectieuses, 2020. **50**(4): p. 332-334.
21. Jing, Q.L., et al., *Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study.* Lancet Infect Dis, 2020. **20**(10): p. 1141-1150.
22. Team, R.C., *R: A Language and Environment for Statistical Computing.* 2013.
23. Friedman, J.H., T. Hastie, and R. Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent.* Journal of Statistical Software, Articles, 2010. **33**: p. 1-22.
24. Rohart, F., et al., *mixOmics: An R package for 'omics feature selection and multiple data integration.* PLOS Computational Biology, 2017. **13**(11): p. e1005752.
25. Thévenot, E.A., et al., *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): p. 3322-35.
26. Sumner, L.W., et al., *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics : Official journal of the Metabolomic Society, 2007. **3**(3): p. 211-221.
27. Schurink, B., et al., *Viral presence and immunopathology in patients with lethal COVID-19: a prospective autopsy cohort study.* Lancet Microbe, 2020. **1**(7): p. e290-e299.
28. Walker, H.J. and M.M. Burrell, *Could breath analysis by MS could be a solution to rapid, non-invasive testing for COVID-19?* Bioanalysis, 2020. **12**(17): p. 1213-1217.
29. van Oort, P., et al., *Exhaled breath metabolomics for the diagnosis of pneumonia in intubated and mechanically ventilated ICU–patients.* European Respiratory Journal, 2017. **50**(suppl 61): p. OA4653.
30. Ulanowska, A., et al., *The application of statistical methods using VOCs to identify patients with lung cancer.* J Breath Res, 2011. **5**(4): p. 046008.
31. Ligor, M., et al., *Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry.* Clin Chem Lab Med, 2009. **47**(5): p. 550-60.
32. Phillips, M., J. Greenberg, and R.N. Cataneo, *Effect of age on the profile of alkanes in normal human breath.* Free Radic Res, 2000. **33**(1): p. 57-63.
33. Ibrahim, B., et al., *Non-invasive phenotyping using exhaled volatile organic compounds in asthma.* Thorax, 2011. **66**(9): p. 804-9.
34. Cazzola, M., et al., *Analysis of exhaled breath fingerprints and volatile organic compounds in COPD.* COPD Research and Practice, 2015. **1**(1): p. 7.
35. Bouza, M., et al., *Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients.* J Breath Res, 2017. **11**(1): p. 016015.
36. Bajtarevic, A., et al., *Noninvasive detection of lung cancer by analysis of exhaled breath.* BMC Cancer, 2009. **9**: p. 348.
37. Gaude, E., et al., *Targeted breath analysis: Exogenous volatile organic compounds (EVOC) as metabolic pathway-specific probes.* Journal of Breath Research, 2019. **13**(3).
38. Grassin-Delyle, S., et al., *Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study.* EBioMedicine, 2021. **63**: p. 103154.
39. Woloshin, S., N. Patel, and A.S. Kesselheim, *False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications.* New England Journal of Medicine, 2020. **383**(6): p. e38.
40. Zhou, Y., et al., *Sensitivity evaluation of 2019 novel coronavirus (SARS-CoV-2) RT-PCR detection kits and strategy to reduce false negative.* PLoS One, 2020. **15**(11): p. e0241469.
41. Holborow, A., et al., *The clinical sensitivity of a single SARS-CoV-2 upper respiratory tract RT-PCR test for diagnosing COVID-19 using convalescent antibody as a comparator.* Clinical Medicine, 2020. **20**(6): p. e209.


(a) PLS DA on Selected Features
count
PLS-DA Score
PCR Result
PCR Negative
PCR Positive
(b) PCA on Selected Features
PC2 Score
PC1 Score
PCR Result
PCR Negative
PCR Positive
(c) PLS DA Score on Selected Features
Wilcoxon, p = 9.5e-07
Score
PCR Negative
PCR Positive
PCR Result
PCR Negative
PCR Positive
(d) PC1 Score on Selected Features
Wilcoxon, p = 1.4e-05
PC1 Score
PCR Negative
PCR Positive
PCR Result
PCR Negative
PCR Positive
(e) ROC Curve for Discriminant Score
Sensitivity
Specificity
AUC 0.836 ,AUC 95% CI 0.745 - 0.928
PPV 0.895 ,PPV 95% CI 0.797 - 0.992
NPV 0.600 ,NPV 95% CI 0.448 - 0.752
(f) ROC Curve for First PC1 Score
Sensitivity
Specificity
AUC 0.799 ,AUC 95% CI 0.698 - 0.900
PPV 0.854 ,PPV 95% CI 0.745 - 0.962
NPV 0.595 ,NPV 95% CI 0.436 - 0.753


(a) PLS DA on Selected Features
count
PLS-DA Score
Clinical Probability
1 Major 3 Minor
2 Major 3 Minor
(b) PCA on Selected Features
PC2 Score
PC1 Score
Clinical Probability
1 Major 3 Minor
2 Major 3 Minor
(c) PLS DA Score on Selected Features
Wilcoxon, p = 0.026
Score
1 Major 3 Minor
2 Major 3 Minor
Clinical Probability
1 Major 3 Minor
2 Major 3 Minor
(d) PC1 Score on Selected Features
Wilcoxon, p = 0.45
PC1 Score
1 Major 3 Minor
2 Major 3 Minor
Clinical Probability
1 Major 3 Minor
2 Major 3 Minor
(e) ROC Curve for Discriminant Score
Sensitivity
Specificity
AUC 0.659 ,AUC 95% CI 0.529 - 0.789
PPV 0.829 ,PPV 95% CI 0.704 - 0.953
NPV 0.419 ,NPV 95% CI 0.271 - 0.566
(f) ROC Curve for First PC1 Score
Sensitivity
Specificity
AUC 0.555 ,AUC 95% CI 0.408 - 0.702
PPV 0.810 ,PPV 95% CI 0.691 - 0.928
NPV 0.444 ,NPV 95% CI 0.282 - 0.607