

## Exploratory study on application of MALDI-TOF-MS to detect SARS-CoV-2 infection in human saliva

Monique Melo Costa<sup>1,2,3,#</sup>, Hugo Martin<sup>1,2,3,#</sup>, Bertrand Estellon<sup>4</sup>, François-Xavier Dupé<sup>4</sup>, Florian Saby<sup>1,2,3</sup>, Nicolas Benoit<sup>1,2,3,5</sup>, Hervé Tissot-Dupont<sup>3,6</sup>, Matthieu Million<sup>3,6</sup>, Bruno Pradines<sup>1,2,3,5</sup>, Samuel Granjeaud<sup>7</sup>, Lionel Almeras<sup>1,2,3\*</sup>

<sup>1</sup>Unité Parasitologie et Entomologie, Département Microbiologie et Maladies Infectieuses, Institut de Recherche Biomédicale des Armées, Marseille, France.

<sup>2</sup>Aix-Marseille Univ, IRD, SSA, AP-HM, VITROME, Marseille, France.

<sup>3</sup>IHU Méditerranée Infection, Marseille, France.

<sup>4</sup>Laboratoire d'Informatique et Systèmes, Aix-Marseille Univ, CNRS, Univ. de Toulon, Marseille, France.

<sup>5</sup>Centre national de référence du Paludisme, Marseille, France.

<sup>6</sup>Aix-Marseille Univ, IRD, AP-HM, MEPHI, Marseille, France.

<sup>7</sup>CRCM Integrative Bioinformatics platform, Centre de Recherche en Cancérologie de Marseille, INSERM, U1068, Institut Paoli-Calmettes, CNRS, UMR7258, Aix-Marseille Université UM 105, Marseille, France.

#These authors contributed equally to this work

\***Corresponding author:** Dr. Lionel ALMERAS. Unité Parasitologie et Entomologie (IRBA), Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 Boulevard Jean Moulin 13385 Marseille cedex 05, France. Phone: 33 (0) 4 91 32 43 75. Fax: 33 (0) 4 91 83 03 90. E-mail address: [almeras.lionel@gmail.com](mailto:almeras.lionel@gmail.com).

**Authors' emails:** MMC: [mcosta.monique@gmail.com](mailto:mcosta.monique@gmail.com); HM : [hugomartin18@orange.fr](mailto:hugomartin18@orange.fr); BE: [bertrand.estellon@univ-amu.fr](mailto:bertrand.estellon@univ-amu.fr); F-XD: [francois-xavier.dupe@univ-amu.fr](mailto:francois-xavier.dupe@univ-amu.fr); FS: [flo.saby80@laposte.net](mailto:flo.saby80@laposte.net); NB: [nicobenoit73@hotmail.com](mailto:nicobenoit73@hotmail.com); HTD : [herve.tissot-dupont@ap-hm.fr](mailto:herve.tissot-dupont@ap-hm.fr); MM : [Matthieu.MILLION@ap-hm.fr](mailto:Matthieu.MILLION@ap-hm.fr); BP: [bruno.pradines@gmail.com](mailto:bruno.pradines@gmail.com); SG : [samuel.granjeaud@inserm.fr](mailto:samuel.granjeaud@inserm.fr); LA: [almeras.lionel@gmail.com](mailto:almeras.lionel@gmail.com).

**Running title:** Covid-19 diagnosis by MALDI profiling.

## Abstract

SARS-CoV-2 caused a large outbreak since its emergence in December 2019. The COVID-19 diagnosis became a priority to isolate and treat infected individuals in order to break the contamination chain. Currently, the reference test for COVID-19 diagnosis is the molecular detection (RT-qPCR) of the virus from nasopharyngeal swab (NPS) samples. Although this sensitive and specific test remains the gold standard, it has several limitations, such as the invasive collection method, the relative high cost and the duration of the test. Moreover, the material shortage to perform tests due to the discrepancy between the high demand for tests and the production capacities puts additional constraints on RT-qPCR. Here, we propose a PCR-free method for diagnosing SARS-CoV-2 based on Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) profiling and machine learning (ML) models from salivary samples. Kinetic saliva samples were collected at enrollment and ten and thirty days later (D0, D10 and D30), to assess the classification performance of the ML models compared to the molecular tests performed on NPS specimens. Spectra were generated using an optimized protocol of saliva collection and successive quality control steps were developed to ensure the reliability of spectra. A total of 360 averaged spectra were included in the study. At D0, the comparison of MS spectra from SARS-CoV-2 positive patients (n=105) with healthy healthcare controls (n=51) revealed nine peaks that significantly distinguished the two groups. Among the five ML models tested, Support Vector Machine with Linear Kernel (SVM-LK) provided the best performance on the training dataset (accuracy = 85.2 %, sensitivity = 85.1 %, specificity = 85.3 %, F1-Score = 85.1 %). The application of the SVM-LK model on independent datasets confirmed its performances with 88.9% and 80.8% of correct classification for samples collected at D0 and D30, respectively. Conversely, at D10, the proportion of correct classification fallen to 64.3%. The analysis of saliva samples by MALDI-TOF MS and ML appears as an interesting supplementary tool for COVID-19 diagnosis, despite the mitigated results obtained for convalescent patients (D10).

Keywords: SARS-CoV-2 detection; Saliva; MALDI-TOF MS; Machine learning

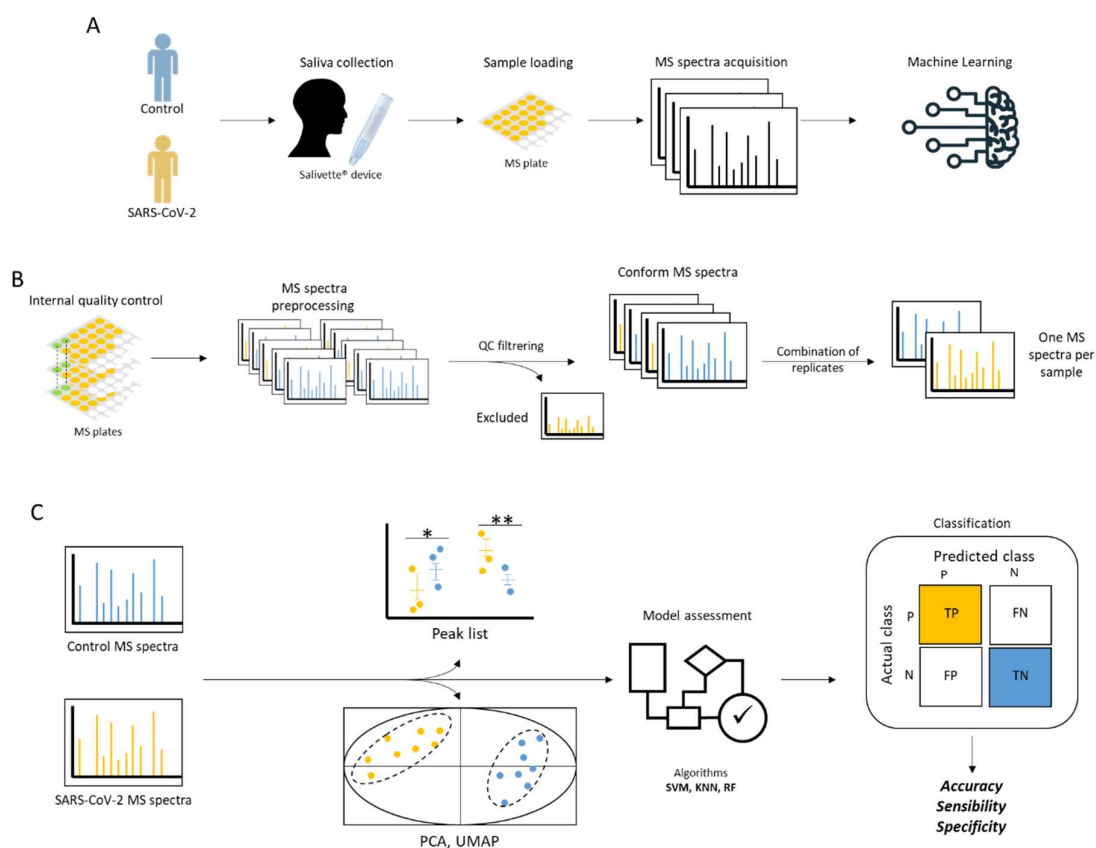
## Introduction

Since March 2020, the novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been considered as a pandemic [1]. Its high rate of spread and infection in the human population, the lack of effective and validated treatment, has led authorities in several countries to impose lockdowns to slow the spread of coronavirus disease 2019 (COVID-19) [2]. To manage this health crisis, screening individuals is essential to isolate infected cases. These screening tests are mainly carried out using nasopharyngeal swabs (NPSs) and reverse-transcription quantitative polymerase chain reaction (RT-qPCR) for detection of viral RNA [3]. NPSs are invasive and can cause from pain to nose bleeding for patients having a sensitive nose. This discomfort could induce sneezing by patients, exposing healthcare workers to a significant contamination risk. To avoid the constraints of NPS collection, others types of samples were proposed [4,5]. Among them, saliva appeared to be a relevant alternative and had the advantages of being minimally invasive and allowing self-collection [6]. Moreover, we recently demonstrated that the standardization of saliva collection with Salivette® device improved the molecular detection of SARS-CoV-2 compared to NPSs [7]. Although sensitive and specific, RT-qPCR tests remain relatively long (2-5 hours) and expensive, and the strong international demand for molecular biology consumables are factors limiting the implementation of general screening and creating tensions in reagent procurement [8].

To prevent the risks of screening shortage, an alternative strategy, PCR-free, based on the detection of specific protein signatures by Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) profiling has been proposed. MALDI-TOF MS profiling has been used for more than a decade in routine diagnostics by microbiology laboratories for the identification of microorganisms (e.g., bacteria, yeasts and fungi) [9]. Its performance in terms of sensitivity, specificity, speed, robustness, simplicity of sample preparation and low cost explain the rapid spread of this tool, which has revolutionized identification techniques in microbiology [10]. The principle of sample classification relies on matching of query MS spectra against a library of reference MS

spectra. This strategy of spectral matching was successfully applied for viral species identification [11,12].

A pioneering study reported the potential of MALDI-TOF MS profiling approach to differentiate COVID-19-positive individuals from negative based on analysis of MS spectra from NPSs [13]. Another work has shown that the combination of MALDI-TOF MS and machine learning (ML) analysis appears to be a suitable strategy for COVID-19 diagnosis using NPS specimens [14]. The application of this strategy to serum samples succeeded to classify COVID-19 patients according to the severity of clinical symptoms [15]. Among the up-regulated plasma proteins, amyloid A1 and A2 proteins appeared as the best biomarker candidates [15,16]. These data revealed the potential of MALDI-TOF MS profiling for SARS-CoV-2 diagnosis using either blood puncture or NPS samples. Until now, no work had explored the potential of saliva for the diagnosis of COVID-19 using MALDI-TOF MS profiling. Recently, we established and standardized a protocol for the preparation of saliva samples for MALDI-TOF MS analysis [17]. This work revealed heterogeneity in saliva MS profiles between healthy individuals, which suggested the need for sophisticated bio-statistical analyses, such as ML approaches, to highlight saliva protein signature for COVID-19 diagnosing. In order to address the growing need for screening, the present study investigated the performance of MALDI-TOF MS profiling combined with ML approaches to discriminate COVID-19 patients from healthy individuals applied to saliva MS spectra. Prior to the detection of potential biomarkers, a pipeline of quality control steps from the pre-processing stage was developed to filter out poor quality or non-conform MS spectra (Figure 1). A set of statistical analyses and five ML approaches were performed on the MS spectra, in order to search for features in spectra distinguishing individuals molecularly diagnosed positive and negative for SARS-CoV-2. Moreover, the ML approach with the best classification performance was applied to a kinetic saliva collection taken from the same individuals ten and thirty days after the primary collection. The corroboration of the classification prediction with the clinical results will be discussed.



**Figure 1. Schematic presentation of the experimental workflow. A.** Main steps of the study. Individuals SARS-CoV-2 positive and negative (controls) based on NPS RT-qPCR tests were enrolled. Saliva were collected with Salivette® devices, loaded onto MS plate and then submitted to MALDI-TOF MS acquisition. The resulting MS spectra were analyzed using machine learning (ML) methods. **B.** Pre-processing steps of the MS spectra included a quality control step, evaluated the homogeneity of the data among the Cov+ and Cov- groups, and determined peak detection conditions. **C.** Strategy used for the prediction of SARS-CoV-2 positive saliva using ML. Statistical analyses were performed to reveal relevant MS peaks before the assessment of ML models.

## Materials and Methods

*Ethical statement.* The study protocol was reviewed and approved by the Ile de France 1 ethical committee (N°2020-A01249-30 protocol, 8 June 2020). Demographics, clinical data and samples were collected uniquely after the understanding of the study protocol and consent acknowledgement by the participants. A questionnaire on the health status of each participant was carried out. All participant information and samples were anonymized before use. Sample handling was carried out under MSC-Advantage™ class II biological safety cabinets (Thermo Fischer Scientific, Villebon sur Yvette, France).

*Individual recruitment.* COVID group. During the period from 23 July 2020 to 21 September 2020, outpatients consulting at the Institut Hospitalo-Universitaire (IHU) Méditerranée Infection (Marseille, France) and tested molecularly positive within the last 5 days for SARS-CoV-2 by NPSs were invited to enroll the research study. Saliva samples were collected on the day of patient inclusion (D0), ten (D10) and thirty (D30) days later. In parallel, a new NPS was performed for all participants to check their COVID-19 status, on the D0 and D10. Individuals under 18 years old, non-French-speaking, pregnant women and individuals suffering of Gougerot-Sjögren Syndrome were excluded.

Healthcare workers group. Healthcare workers without fever or respiratory symptoms potentially exposed to patients or patient samples with COVID-19 were invited to enroll in the study. A NPS was performed to all participants to establish their COVID-19 status the same day of saliva collection. Saliva collection was performed on the day of individual inclusion (D0), ten (D10) and thirty (D30) days later. Saliva from one healthcare worker (male, 46 years old), was aliquoted (20µL per tubes), loaded onto each MS plate and was used as inter-plate control.

*NPS management.* A standard protocol was used for NPSs collection using nasal swabs with viral transport medium (Pacific Laboratory Products, Blackburn, Australia). Routine diagnosis protocol

was applied for SARS-CoV-2 detection on NPS samples by RT-qPCR [18,19]. The classification of individuals was realized according to molecular results of SARS-CoV-2 detection on NPS samples.

*Saliva collection.* The optimized protocol for saliva sampling and treatment, previously described, was applied [17]. Briefly, a bottle of spring water was given to each participant who performed a mouthwash prior to saliva sampling. Saliva was collected using Neutral Salivette® (SARSTEDT, Numbrecht, Germany), under the supervision of a laboratory technician. The samples were stored in ice, until their use. The storing time never exceeded 6 hours. All Salivette® were centrifuged at 1500×g for 2 min at 4 °C and were transferred to 1.5 mL Eppendorf tubes stored at 4°C until their use.

*Sample loading.* Twenty microliters of saliva sample were mixed with one volume (i.e., 20 µL) of the Mix buffer (mixture (50/50) of 70% (v/v) formic acid and 50% (v/v) acetonitrile (Sigma-Aldrich, Lyon, France)) used for MALDI-TOF MS profiling analysis [20]. One microliter of the mixed saliva sample was spotted in quadruplicate onto a 96 polished steel MALDI target plate (Bruker Daltonics, Wissembourg, France). After air-drying, 1 µL of matrix solution composed of saturated  $\alpha$ -cyano-4-hydroxycinnamic acid (Sigma-Aldrich), 50% (v/v) acetonitrile, 2.5% (v/v) trifluoroacetic acid, prepared with HPLC-grade water, was added. Matrix quality (i.e., absence of MS peaks due to matrix buffer impurities) and MALDI-TOF apparatus performance were controlled, respectively, by loading matrix solution in duplicate and saliva from the same healthy healthcare worker onto each MS plate. After air-drying at RT, the MS target plate was applied in the mass spectrometer.

*MALDI-TOF MS parameters.* Protein mass profiles were obtained using a Microflex LT MALDI-TOF Mass Spectrometer (Bruker Daltonics), with detection in the linear positive-ion mode at a laser frequency of 50 Hz within a mass range of 2-20 kDa. The setting parameters of the MALDI-TOF MS

apparatus were identical to those previously used [21]. Briefly, the acceleration voltage was 20 kV, and the extraction delay time was 200 ns. Each spectrum corresponds to ions obtained from 240 laser shots performed in six regions of the same spot and automatically acquired using the AutoXecute of the Flex Control v.3.0 software (Bruker Daltonics).

*MS spectra preprocessing.* Raw data of MALDI-TOF MS spectra (Bruker Daltonics) were preprocessed using MALDIquant [22] and MALDIrppa [23] R packages. The MS spectra were then processed by applying a square root transformation, a smoothing using the Savitzki-Golay-Filter algorithm [24], baseline correction with the SNIP algorithm [25], and normalization of intensity values using the total-ion-current (TIC). All these transformations were performed through the MS processing functions (`transformIntensity`, `smoothIntensity`, `removeBaseline`, `calibrateIntensity`) of the MALDIquant package using the default parameters. To detect and filter out poor quality (or no conform) MS spectra, the `screenSpectra` function of the MALDIrppa package was used. It computed an atypicality score for each spectrum. All spectra with a score outside the limit were excluded from the rest of the analysis. Then the non-excluded spectra from replicates of the same sample were averaged (`averageMassSpectra`, MALDIquant package), giving an average spectrum for each sample. The intensity of each average spectrum was normalized by adjusting the TIC (`calibrateIntensity`, MALDIquant package). The alignment of peaks and spectra were performed with functions of the MALDIquant package. Peaks were detected by adjusting the signal-to-noise ratio (SNR) value and binned with a tolerance value of 0.001. To determine the optimal SNR value, the number of detected peaks and its standard deviation were measured for each SNR value tested. The SNR value for which the number of detected peaks per group was homogenous and stabilized, was considered as the optimal trade-off. Only peaks with a frequency of occurrence greater than 50% per group were considered. At the end of the peak detection, the resulting matrix of peak intensities was used for statistical analyses and ML approaches.

*Statistical analyses.* After verifying that the values in each group did not assume a Gaussian distribution, Wilcoxon matched-pairs signed-rank tests were computed when appropriate with R. For multiple testing, a false discovery rate (FDR) corrections was applied using Benjamini–Hochberg correction [21]. The selection of MS peaks required a significant Wilcoxon rank sum test ( $p < 0.05$ ) and a relevant Benjamini-Hochberg correction ( $p < 0.1$ ).

*Dimension reduction algorithms.* Principal Component Analysis (PCA) [26] and Uniform Manifold Approximation and Projection (UMAP) [27] were used to visualize data on a two-dimension reduced space and to deduce the relations between spectra profiles of individuals. MS spectra from confirmed SARS-CoV-2 positive patients (n=105) and control healthcare workers (n=51) collected at D0 were used. Dimension reductions were carried out either on all peaks detected or on selected peaks. PCA was performed using the R FactoMineR [28] package. UMAP was computed using the umap R library [29].

*Machine Learning (ML) analyses.* MS spectra from 102 saliva samples, including confirmed SARS-CoV-2 positive patients (n=51) and control healthcare workers (n=51) collected at D0, were selected to establish the best algorithm classifying SARS-CoV-2 positives from control samples. Five different algorithms, namely, support vector machine with linear kernel (SVM-LK), support vector machine with radial basis function kernel (SVM-RK), random forest (RF) and k-nearest neighbors (K-NN) and linear discriminant analysis (LDA) were evaluated. ML models were implemented with the caret package [30], except for the SVM-RK which was performed using the e1071 package [31]. Contrary to caret, the e1071 library provides an implementation of SVM-RK models with the gamma hyperparameter, which allows gamma to be tuned to improve the performance of models. The training process was conducted by a threefold nested repeated tenfold cross-validation to ensure an unbiased

estimation of performance. The proportion of the training and test datasets were adjusted to 75% and 25% respectively to perform the repeated cross-validation. Model performances after the training step were estimated by computing sensitivity, specificity, accuracy, area under the curve (AUC) and average precision (AP) values. AUC and AP were computed on receiver operating characteristic (ROC) and precision-recall (PR) curves with the ROC R package [32].

## Results

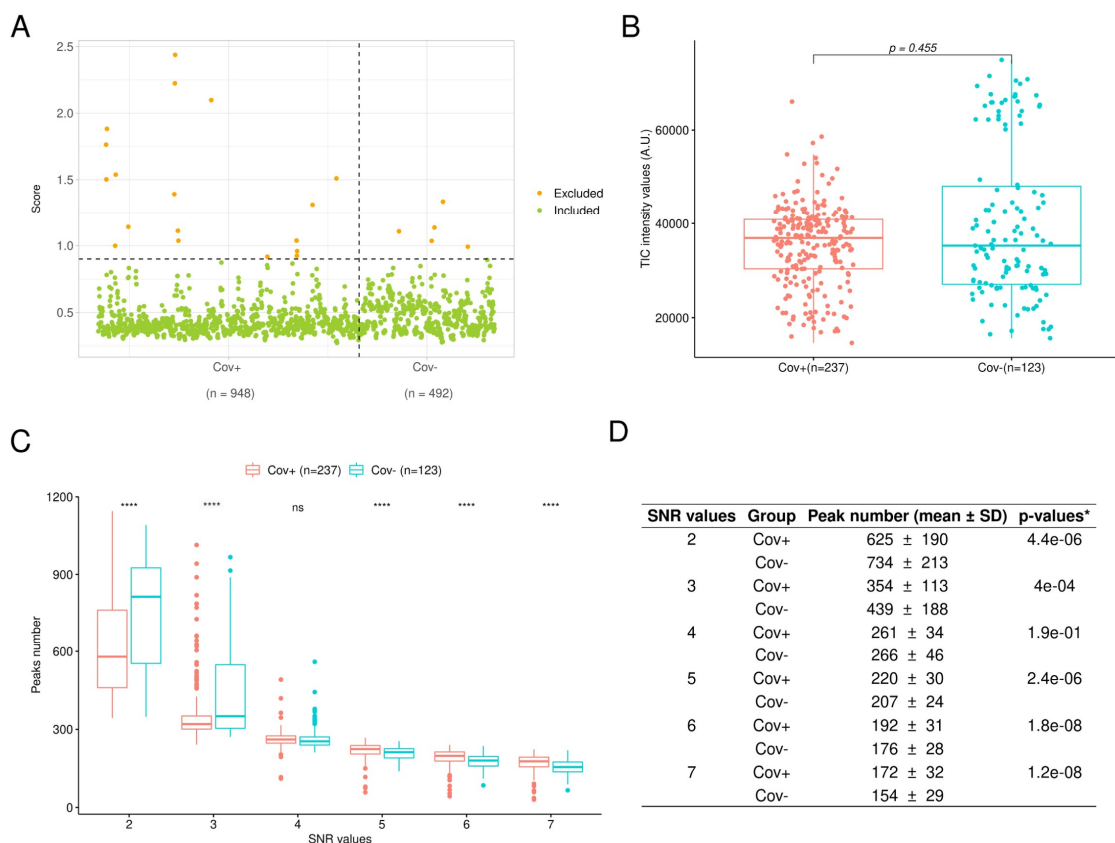
### *Characteristics of participants*

A total of 360 saliva samples, obtained from 138 patients diagnosed SARS-CoV-2 positive (Cov+ group) and 51 healthcare workers considered as healthy control (Cov- group), were included in the study. For both groups, serial saliva sampling was performed ten (D10) and thirty (D30) days after the first saliva collection (Table 1). Among patients from the Cov+ group, the SARS-CoV-2 positive status was confirmed by NPS tests for 76.1% (105/138) of them at the enrollment day (D0). At D10, the virus was still detectable for 16 out of 79 (20.3%) patients from the Cov+ group. At D30, the 20 individuals from the Cov+ group were considered as exempt of infections based on the delay since primary infection and in the absence of any clinical symptoms in the ten days before saliva sampling. SARS-CoV-2 detection performed on NPS samples from healthcare workers (Cov- group) were all negative at the enrollment day (D0, n=51) and ten days later (D10, n=40) (Table 1). The absence of any clinical symptoms in the ten days before saliva sampling at D30 for the 32 individuals from the Cov- group were considered as exempt from infections. Moreover, SARS-CoV-2 was detected in none of the individuals from Cov+ and Cov- groups molecularly tested on saliva at D30. No significant differences were noted for age ( $p=0.069$ , Kruskal-Wallis test) or gender ( $p=0.279$ ,  $df=5$ , Pearson's Chi-square test) among the groups considering collection time points. It is interesting to note that all individuals from the Cov+ group were tested positively, outside our institute, less than one week before to be included in the study.

### *Preprocessing steps of saliva MS spectra*

Prior to research saliva protein signature distinguishing Cov+ from Cov- individuals, MS spectra were submitted to successive quality control steps. As each sample was loaded in quadruplicate, a total of 1440 MS spectra were obtained. The screenSpectra function (MALDIrppa package) assigned an atypicality score (score threshold  $> 0.9$ , Figure 2A) to 23 MS spectra (Cov+, n=18; Cov-, n=5).

These spectra were rejected for the rest of the analysis. As no more than two replicate spectra were excluded for each sample, no saliva sample was removed after this step. After filtering atypical spectra and averaging spectra replicates, the comparison of the TIC between Cov+ (n=237) and Cov- (n=123) groups revealed no significant differences ( $p > 0.05$ , Wilcoxon rank sum test, Figure 2B), underlining a stable spectra acquisition throughout the study independently of the sample group origin.



**Figure 2. Pre-processing steps of MS spectra.** **A.** Detection of atypical spectra from Cov+ (n=948) and Cov- (n=564) groups using the screenSpectra function from the MALDIrppa package. A score was computed for each spectrum. Horizontal dotted line corresponds to atypical score threshold. **B.** Comparison of total-ion-current intensities after filtering atypical spectra and averaging spectra replicates between Cov+ (n=237) and Cov- (n=123) groups (two-tailed Wilcoxon rank sum test). **C.** Comparison of peak number distributions according to SNR values ranging from 2 to 7 between

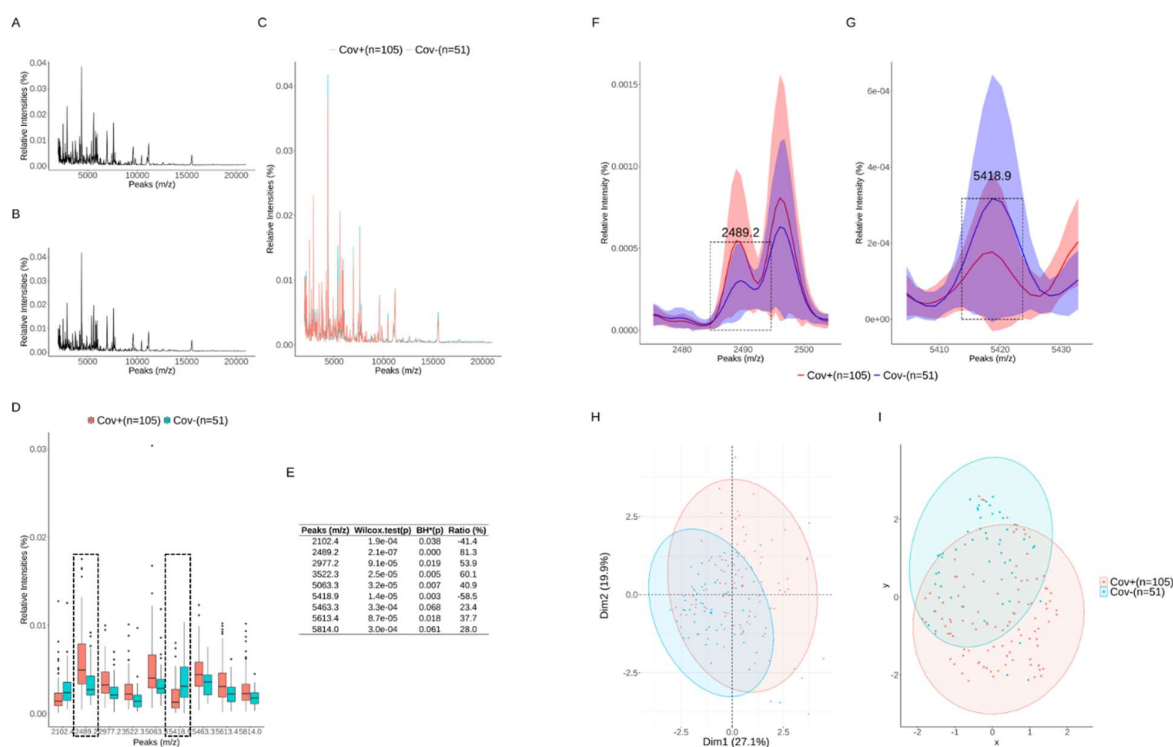
average MS spectra from Cov+ (n=237) and Cov- (n=123) groups. \*\*\*\*  $p < 0.001$  by two-tailed Wilcoxon rank sum test. **D.** Summary table of peak number distributions and comparisons for each SNR value between average MS spectra from Cov+ (n=237) and Cov- (n=123) groups. \*exact *p-values* of by two-tailed Wilcoxon rank sum test. A.U.: arbitrary units; Cov+/-: MS spectra from individuals enrolled in the COVID-19 and control groups, respectively; SD: standard deviation; SNR: signal-to-noise ratio.

The comparison of peak numbers detected according to the SNR value was performed to select the SNR value allowing the inclusion of mainly true peaks and assessing peak detection between the Cov+ and Cov- groups. For SNR values lower than 4, the number of detected peaks was highly heterogeneous between the two groups and among the SNR values. Moreover, an important decrease of the number of detected peaks was noticed with the increment of SNR values (Figure 2C). Conversely, for SNR values of 4 and upper, the mean number of detected peaks and its standard deviation diminished slowly (Figure 2D). Interestingly, the peak number distributions between the Cov+ and Cov- groups were not significantly different only for the SNR value equal to 4 ( $p > 0.05$ , Wilcoxon rank sum test, Figure 2D). Then, for the remainder of the analysis, a SNR value of 4 was selected and an intensity matrix of 212 peaks was obtained.

#### *Comparison of saliva MS spectra and detection of potential biomarkers*

For these analyses, uniquely saliva samples from Cov+ group confirmed SARS-CoV-2 positive at D0 (n=105) were compared to the Cov- group from D0 (n=51). No visible difference was noticed when the mean spectra of the Cov+ (n=105, Figure 3A) and the Cov- (n=51, Figure 3B) groups were compared (Figure 3C), reflecting the complexity and the heterogeneity of spectra profiles. The application of a Wilcoxon rank sum test ( $p < 0.05$ ) identified 52 peaks distinguishing SARS-CoV-2 positive from negative samples (Additional table 1). The correction for multiple hypothesis testing

(Benjamini-Hochberg,  $p < 0.1$ ) allowed nine peaks to be selected (Figure 3D). Among these peaks, seven and two exhibited higher intensity in SARS-CoV-2 positive and negative groups, respectively (Figure 3E). The peaks with mass-to-charge ratio ( $m/z$ ) of 2489.2 and  $m/z$  of 5418.9 possessed the most significant intensity differences for the positive (Figure 3F) and negative (Figure 3G) groups, respectively. PCA analysis, carried out on these 9 selected peaks, allows a partial distinction between SARS-CoV-2 positive and negative groups (Figure 3H). Carrying out a UMAP on these selected peaks does not improve the separation of groups (Figure 3I). When PCA and UMAP were performed using all peaks (Figure S1A-B) or the 52 peaks selected with two-tailed Wilcoxon rank sum test (Figure S1C-D), no clear separation could be noticed.

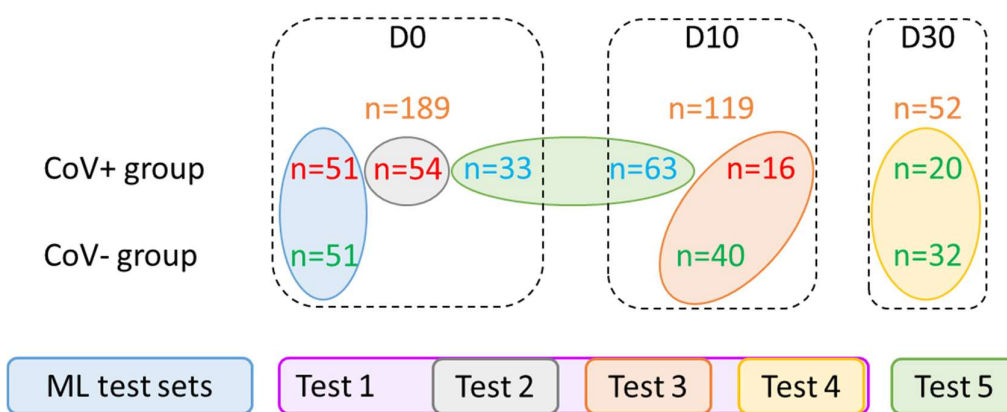


**Figure 3. Comparison of saliva MS spectra between SARS-CoV-2(+) (n=105) and control (n=51) groups sampled at D0.** Mean MS spectra profiles with a total-ion-current normalization applied to peak spectra intensities of Cov+ (A) and Cov- (B) individuals and their superposition (C). D. List of MS peaks with significant total-ion-current normalized intensity differences (two-tailed Wilcoxon rank sum test ( $p < 0.05$ ) with a Benjamini-Hochberg correction ( $p < 0.1$ ) for each peak) between

Cov+ and Cov- groups. Dashed box highlighted the most two relevant peaks. **E.** Summary table of Wilcoxon rank sum tests with a Benjamini-Hochberg (BH) correction for each peak. Ratio represent the proportion (%) of mean peak intensity variations between Cov+ and Cov- groups. Positive and negative values correspond, respectively, to upper and lower mean peak intensity variations in Cov+ compared to Cov- groups. Comparison of the two greatest peaks at m/z 2489.2 (**F**) and at m/z 5418.9 (**G**) differentiating Cov+ and Cov- groups. Lines and shadow represent mean values  $\pm$  interquartile range (IQR), respectively. Dashed box framed each relevant peak. PCA (**H**) and UMAP (**I**) performed on Cov+ (n=105) and Cov- (n=51) groups using the selected peaks (n=9) with two-tailed Wilcoxon rank sum test ( $p < 0.05$ ) and a Benjamini-Hochberg correction ( $p < 0.1$ ).

#### *Building and selection of machine learning models*

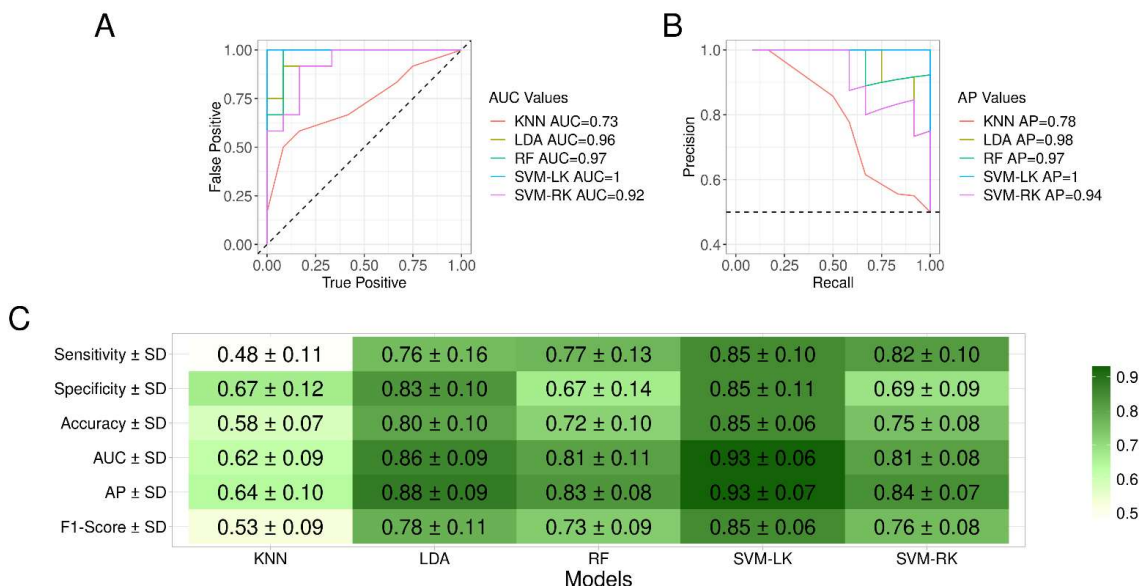
To assess the performance of ML approaches to separate SARS-CoV-2 positive from negative individuals, samples were split into different groups. A first group was used during the training phase of the ML models, which allowed the selection of the best model. Then, during the test phase, the performance of the selected model was evaluated on independent groups. For the training, a balanced group of samples (51 Cov+, 51 Cov-) was formed from individuals at D0 (Figure 4). The 51 Cov+ samples were randomly selected from the 105 positive samples at D0 with a confirmed infection status. The remaining samples were organized into five test groups (Test 1 to Test 5), according to collection time points (D0, D10 and D30) and NPS RT-qPCR results (Figure 4). A total of 96 samples from Cov+ group were classified as uncertain negative for SARS-CoV-2, because all these individuals were diagnosed SARS-CoV-2 positive either in the week preceding their enrollment (n=33 at D0) or ten days before this second NPS test (n=66 at D10). Considering SARS-CoV-2 false-negative, the detection rate can reach 30% [33], these 96 samples were classified as uncertain negatives and were tested separately (Test 5).



**Figure 4. Training and testing sets for machine learning models.** The individuals enrolled were divided in several groups to establish the best training set and to assess the performance of the selected ML model on different groups. The total number of samples, in the Cov+ (n=237) and Cov- (n=123), at the inclusion (D0), ten (D10) and thirty (D30) days later were indicated (orange numbers). Green, red and blue numbers correspond to saliva samples classified as SARS-CoV-2 positive, negative or uncertain negative, respectively, based on RT-PCR results. The samples selected for training set selection (n=102) and assessment of the best ML model, organized into five groups (Test 1 to Test 5) are indicated by circles of the same color code. Test 2 to Test 4 were included in Test 1. Cov+/-: MS spectra from individuals enrolled in the COVID-19 and control groups, respectively; ML: machine learning.

Five different ML models (K-NN, LDA, RF, SVM-LK and SVM-RK) were studied. The best ROC and PR curves obtained for each ML model are presented in Figure 5A and 5B, respectively. With the exception of the K-NN model, the models exhibited high performance with the best area under the curve (AUC) values greater than 0.92 and 0.94 for ROC and PR curves, respectively. To select the model with the best compromise of sensitivity and specificity, the F1-score was computed (Figure 5C). The SVM-LK model obtained the best F1-score, about 0.85, achieving a sensibility of 85.1% and a specificity of 85.3%. The SVM-LK model reached the greatest average area under the curve (AUC) values of 0.93 for ROC and 93 for PR curves, supporting the existence of a protein pattern

associated with saliva mass spectra. The SVM-LK model was then applied to test groups, to determine its prediction performance.

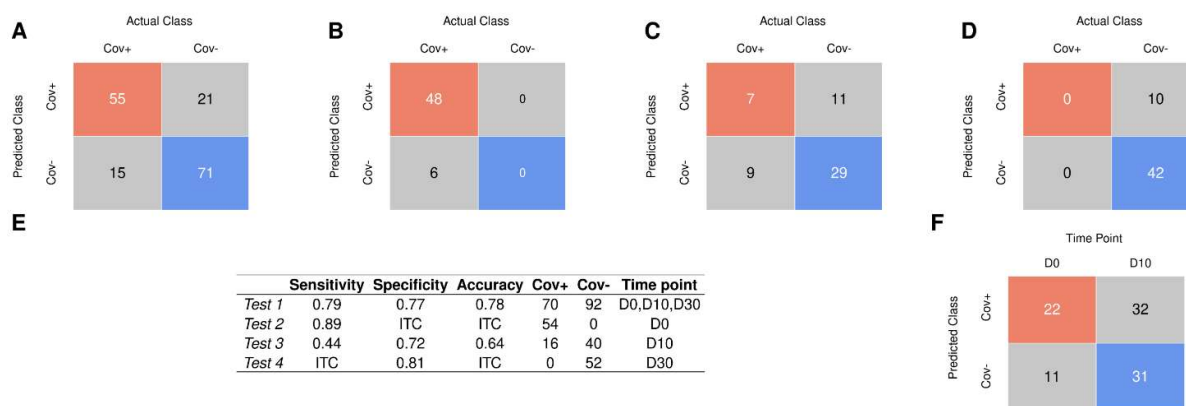


**Figure 5. Performances curves of the five machine learning models.** Best AUC values for ROC (A) and best AP values for Precision Recall (B) curves associated to each model trained on samples from Cov+ (n=51) and Cov- (n=51) groups collected at D0, are presented. C. Summary of performances obtained per ML model. F1-Score was used as a performance measure to select the most suitable developed model. AP: average precision; AUC: area under the curve; Cov+/-: MS spectra from individuals enrolled in the COVID-19 and control groups, respectively; K-NN: K-nearest neighbors; LDA: linear discriminant analysis; ML: machine learning; RF: random forest; SVM-LK: support vector machine with linear kernel; SVM-RK: support vector machine with radial kernel.

#### *Performance of the SVM-LK model*

To evaluate the performance of the SVM-LK model, all saliva MS spectra were included except those used for model training. The first test set (Test 1, n=162), included MS spectra collected at D0, D10 and D30 from Cov+ (n=90) and Cov- (n=72) groups, with the exception of samples classified as

uncertain negatives. Saliva samples collected at D30 from both groups (Cov+ and Cov-) were classified as SARS-CoV-2 negatives (n=52). Then, among Test 1, 70 and 92 samples were classified as positive and negative for SARS-CoV-2 respectively, based on NPS RT-qPCR results (Figure 4). On the Test 1 set, the SVM-LK model distinguished SARS-CoV-2 positive and negative samples with accuracy, sensibility and specificity of 77.8%, 78.6% and 77.2%, respectively (Figure 6A and 6E). The model performance on this independent dataset was lower than those obtained during model training.



**Figure 6. Confusion matrix for test results using SVM linear model.** The metrics of the prediction using samples from the Test 1 (A), Test 2 (B), Test 3 (C), Test 4 (D). E. Summary of performances obtained per group tested (from Test 1 to 4) applied to SVM-LK model. ITC: Impossible to Compute. F. The metrics of the prediction using samples from the Test 5 applied to SVM-LK model. SVM-LK: support vector machine with linear kernel.

The model, trained only on samples collected at D0, could be less efficient to classify samples collected at D10 and D30. To assess this hypothesis, the SVM-LK model was applied to MS spectra from samples collected at D0 (Test 2, n=54), D10 (Test 3, n=56) and D30 (Test 4, n=52) separately (Figure 6B, 6C and 6D). The proportions of correct classification were of 88.9% for Test 2, 64.3% for Test 3 and 80.8% for Test 4 (Figure 6E). Interestingly, the higher performance was obtained for

Test 2, that includes samples collected at D0 as those used to train the model. Finally, we applied the model to MS spectra from samples classified as uncertain negative samples (Test 5, n=96). More than half of the samples (n=54/96) remained predicted as SARS-CoV-2 positive samples (Figure 6F): two third (n=22/33) at D0 and half (n=32/63) at D10.

## Discussion

To overcome the constraint of NPS sampling and the limitations of molecular diagnosis of COVID-19, the use of saliva specimens combined with MALDI-TOF MS and ML approach appeared as a promising alternative strategy. The present work described the application of MALDI-TOF MS for the identification of a specific salivary protein pattern to distinguish SARS-CoV-2 positive from negative individuals. For this purpose, the entire analytical chain must be of high quality.

The quality of spectra is directly linked to the mode of sample collection, storing and preparation before submission to MS [34,35]. Although passive drooling is the most simple and common method for saliva collection, the use of specific devices, such as Salivette®, allows the specimen collection to be standardized [7,36]. Moreover, the saliva is composed of numerous enzymatic proteins that could alter MS spectra by auto-digestion of the sample [37]. To limit these phenomena, an optimized protocol for saliva collection and management applied for MS analyses has recently been established and was used in the present work [17]. Although this protocol enables excellent intra-individual reproducibility, the salivary profile between individuals is heterogeneous. This heterogeneity requires the acquisition of high-quality MS spectra to succeed in a correct classification. In this way, pre-processing steps evaluating MS spectra quality and their homogeneity between groups are crucial. They were carefully defined and performed in this work.

Filtering of atypical MS spectra with MALDIrppa package was performed on each replicate. The rejection of spectra was operated before averaging spectra per sample to prevent alteration of protein profiles. No significant difference in TIC distribution between the Cov+ and Cov- groups was determined, which ensured homogeneity of the spectra intensities. Quality control steps excluding outlier MS spectra improved the comparison of MS spectra. Although the filtering step is frequently used for the pre-processing of MS spectra [13], the TIC distribution is seldom considered. Despite TIC results being linked to the sample composition, important variations of TIC could occur between samples of the same type. These TIC variations generally reflect experimental or technical variations

such as sample preparation (e.g., treatment or amount) and apparatus performances (e.g., laser power or detector sensitivity) [38,39], which could hamper the detection of pertinent biological information. The application of a normalization to TIC values allows all samples to be equalized.

The establishment of peak detection threshold is another decisive parameter for the detection of biological markers [40]. Peak detection depends on the determination of the optimal SNR value. Applying an insufficiently strict SNR value increases the risk of including peaks corresponding to background noise, which could distort the statistical analysis and lead to an inappropriate interpretation of the results. Inversely, choosing too high SNR value may lead to the miss-detection of true peaks, and this lack of information could be deleterious for further analysis steps. Here, the analysis of the number of peaks detected and its variation among samples according to the SNR value, allowed the selection of a threshold of 4. Indeed, for SNR values lower than 4, the high number of peaks detected and its high dispersion among samples were likely related to peaks detected in the background of spectra. Moreover, at a SNR of 4, we controlled that the number of peaks detected between the Cov<sup>+</sup> and Cov<sup>-</sup> groups were comparable, limiting any bias in the analysis. The recent studies evaluating the performance of MALDI-TOF MS to discriminate SARS-CoV-2 positive from negative individuals, using either NPS or serum samples, did not clearly detail the criteria for selecting the SNR value [13,15,41].

For serum samples, the selection of a SNR value of 5 appeared sufficient to separate COVID-19 patients into 3 levels of severity by applying ML models [41]. Nevertheless, the stringency of peak detection might miss some specific peaks that could improve this classification. Conversely, Nachtigall *et al.* chose a SNR value of 2 for peak detection of NPS from individuals diagnosed as positive or not for COVID-19 [13]. Based on this SNR value, they determined seven significantly different peaks. These discriminant peaks had a very low intensity and could be likely not detected if a higher SNR value was selected. Generally, the SNR value is chosen high enough to prevent detection of peaks in background noise. However, it is likely that a too high SNR value will filter out

decisive peaks for classification. Therefore, establishing an accurate strategy to optimize the SNR value becomes compulsory to avoid miss- or over-detection of peaks.

Using an SNR value of 4 at the peaks detection step, we obtained an intensity matrix of 212 peaks. Among them, nine peaks with a relevant different intensity distribution were identified between positive samples for SARS-CoV-2 and healthy individuals. A significantly higher intensity distribution for the Cov+ group compared to the Cov- group was observed for 7 of the 9 peaks. Natchigall et al. [13] reported in nasal swab samples that five of the seven peaks, distinguishing Cov+ and Cov- groups, have a higher intensity distribution in healthy individuals. Similarly, in another study using swab samples, all the discriminant peaks possessed a higher intensity distribution for uninfected individuals than for SARS-CoV-2 contaminated patients [14]. Conversely, here, in saliva samples, discriminant MS peaks correspond mainly to up-regulation protein expression subsequent to SARS-CoV-2 infection.

In parallel to the differential analysis of peak intensity, none of the dimension reduction algorithms (i.e., PCA or UMAP), applied to the intensity matrix of the 9 selected peaks, were able to show a clear separation between SARS-CoV-2 infected and healthy individuals. Similar observations were noticed by other studies examining MS spectra from Covid-19 patients [13,16]. Effectively, only partial distinctions could be identified between infected and uninfected individuals. Based on the above observations, we can assume that dimension reduction algorithms struggle to separate groups because the inter-individual heterogeneity of spectra profiles is large and stronger than the differences between infected and non-infected groups. These intergroup differences are due to a few biomarkers which differential expression contributes little to the overall heterogeneity.

As the differences between the groups are masked by the overall heterogeneity, five ML models were implemented to specifically identify and combine protein profiles in order to increase their effectiveness in separating the groups. The training process of the models allowed us to determine that the SVM model with linear kernel was the best model, based on the F1-score [8,41]. Saliva

spectra were classified with a high accuracy (i.e., 85%). These results are comparable or better to the performance of ML models already described for COVID-19 diagnosis [8,14]. Several studies established that the SVM model performed the best predictions [8,13,15].

To confirm the performance and the robustness of the selected classifier, the SVM-LK model was applied to independent datasets (Test 1 to 5). First, an independent set was constructed by including all saliva spectra, excluding the spectra used in the training set and those classified as uncertain negatives. The performance on this set was a little lower than on the training set, with an accuracy of 77.8%. The spectra were then split according to the collection time (i.e., D0, D10, and D30). The separate evaluation on these independent datasets revealed good predictions at D0 (89%) and D30 (81%). In contrast, at D10, a low accuracy was obtained, notably, due to incorrect classification of SARS-CoV-2 positive samples by the ML model (sensitivity lower than 45%). The weak performance of the ML model on this test set may be due to the gradual recovery of infected patients. It could be hypothesized that a progressive disappearance of the protein pattern linked to the spectra classification in the Cov+ group occurred during the recovery situation. A recent study demonstrated that the concordance of correct classification between saliva and NPS samples by RT-qPCR was lower for convalescent patients [7] and was corroborated by others [42]. The mitigated results of classification were attributed to an intermediate status of patients, whom are no longer at peak disease but not yet fully recovered.

Test 5 consists of patients defined as uncertain negative samples for SARS-CoV-2. The discrepancy in classification between the molecular results and the model prediction can be explained by several factors. Firstly, as all individuals from Cov+ group were diagnosed molecularly as positive for SARS-CoV-2 a few days before enrollment, it is likely that several of them remained positive at the enrollment day (D0) and that the incorrect prediction could be attributed to a molecular miss-detection of the virus on NPSs. Indeed, the sensitivity of the molecular detection of the virus using NPSs was estimated around 70% [33,43]. In our cohort, active SARS-CoV-2 infection was confirmed in 76.1%

(n=105/138) of the patients in the Cov+ group at D0. Despite a negative RT-qPCR result, persistence of virus molecules or infection-related proteins would likely occur, and sustainability of serological inflammatory markers several days or weeks after patients recovering was reported [44,45]. These long-term protein dysregulations could perturb ML classification. The second hypothesis could be that the virus was recently cleared from these individuals, at least in the nasopharyngeal part, and that they are therefore truly negative for SARS-CoV-2. Nevertheless, despite the disappearance of the virus, it is conceivable that traces of the infection could persist for a longer period of time leading to incongruent results between the ML model and the molecular analysis. The correct classification of more than 80% of the samples at D30 by SVM-LK model comforts this claim. We noted that the majority of studies did not evaluate performances of ML models on an independent dataset [13,15], but established the performance metrics on the training set. Two studies tested the ML models on an independent dataset [8,14]. The performance of the ML models was noticeably lower than that obtained on the training set, suggesting an overfitting of the model to the training data. To avoid this problem and to obtain an unbiased measure of the performance of the model, it is essential to use an independent test set [8].

The main originality of our study was to use saliva as specimen. Although, previous works reported the potential of MALDI-TOF MS to monitor SARS-CoV-2 proteins and host immune response against the virus using water gargle samples as saliva collection [46,47], these specimen collection would be unappropriated for young children or elderly. Unlike NPSs, saliva collection is a non-invasive technique which can be performed by the patients themselves and prevents the risk of contamination for healthcare workers [48,49]. Previous works reported the potential of MALDI-TOF MS to monitor SARS-CoV-2 proteins and host immune response against the virus using water gargle samples as saliva collection [46,47]. The introduction of the Salivette® device has standardized the collection and improved the molecular detection of SARS-CoV-2 compared to NPSs [7]. However, it has recently been reported that mouth washing before collection impaired the detection of the virus

in saliva [50], that is why here the molecular results of NPSs were used as a reference for sample classification. A mouth washing was systematically proposed to all individuals enrolled in the present study prior to saliva collection, as numerous patients ate and/or drank while waiting for their turn at the hospital. The record attendance occurred during the second wave of COVID-19 outbreak in the south of France, resulting in a long wait to be diagnosed. It is possible that the mouth washing could also have reduced the protein signatures associated with SARS-CoV-2 infections and thus altered the performance of ML models for sample classification. To assess the effect of mouthwash on model performance, complementary experiments are required in which individuals whom ingested drink or food in the hour prior salivation collection will be excluded [51].

The performance of ML models to predict clinical symptoms in SARS-CoV-2 positive patients will also be evaluated using saliva as specimen. The characterization of such protein markers will be used to predict the favorable or unfavorable clinical course of a patient in order to guide the physicians on the actions to be taken. A recent study reported that the application of ML models on serum samples succeeded in predicting the clinical evolution of SARS-CoV-2 positive patients [15]. The main limitation of the present work was the lack of saliva samples from individuals infected by another virus such as influenza. The potential of ML models to distinguish SARS-CoV-2 from other respiratory viruses, including influenza, has already been demonstrated based on the analysis of MS profiles from NPS [14].

In conclusion, MALDI-TOF MS with collection of salivary samples offers real new perspectives of SARS-CoV-2 clinical diagnostics. The non-necessity of RT-qPCR test and the affordable cost of the MALDI-TOF MS analysis per sample could allow a massive test of the population in order to distinguish infected from uninfected individuals, thus bringing potential improvements to sanitary conditions.

## List of abbreviations

AP: average precision; AUC: area under the curve; COVID-19: Coronavirus Disease 2019; K-NN: k-nearest neighbors; LDA: Linear Discriminant Analysis; MALDI-TOF MS: Matrix-assisted laser desorption ionization time-of-flight mass spectrometry; ML: machine learning; NPS: nasopharyngeal swab; PCA: principal component analysis; PCR: polymerase chain reaction; PR: precision-recall; RF: random forest; ROC: receiver operating characteristic; RT-qPCR: reverse transcriptase quantitative PCR; SARS-CoV-2: severe acute respiratory syndrome Coronavirus 2; SNR: signal-to-noise ratio; SVM-LK: support vector machine with linear kernel; SVM-RK: support vector machine with radial kernel; TIC: total-ion-current; UMAP: uniform manifold approximation and projection;

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work has been supported by the French Defence Innovation Agency – Agence de l'innovation de défense (AID, CoviDiagMS Project, Grant no 2020-COVID19-15).

## Authors' contributions

Conceived and designed the experiments: LA, SG, MMC. Performed the experiments: MMC, HM, NB, LA. Analyzed the data: LA, SG, MH, BE, F-XD, FS, MMC. Contributed reagents/materials/analysis tools: MMC, SG, BE, F-XD, FS, HTD, MM, BP. Sample collections: LA, MMC, NB, HTD, MM, BP. Drafted the paper: LA, SG, HM. Revised critically the paper: all the authors.

## Acknowledgments

We would like to acknowledge IHU Mediterranean Infection nurse staff for their reception in the routine test rooms and all the participants who accepted to donate saliva. We also acknowledge Catherine Verret and Carine Malle (DFRI, SSA, Paris) for their help in the redaction of the ethical statement folder.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References.

1. Thompson R. Pandemic potential of 2019-nCoV. *Lancet Infect Dis.* **2020**; 20(3):280.
2. Umakanthan S, Chattu VK, Ranade AV, Das D, Basavarajegowda A, Bukelo M. A rapid review of recent advances in diagnosis, treatment and vaccination for COVID-19. *AIMS Public Health.* **2021**; 8(1):137–153.
3. Kudo E, Israelow B, Vogels CBF, et al. Detection of SARS-CoV-2 RNA by multiplex RT-qPCR. *PLoS Biol.* **2020**; 18(10):e3000867.
4. Wang W-K, Chen S-Y, Liu I-J, et al. Detection of SARS-associated coronavirus in throat wash and saliva in early diagnosis. *Emerg Infect Dis.* **2004**; 10(7):1213–1219.
5. Lai CKC, Lam W. Laboratory testing for the diagnosis of COVID-19. *Biochem Biophys Res Commun.* **2021**; 538:226–230.
6. Tutuncu EE, Ozgur D, Karamese M. Saliva samples for detection of SARS-CoV-2 in mildly symptomatic and asymptomatic patients. *J Med Virol.* **2021**; 93(5):2932–2937.
7. Melo Costa M, Benoit N, Dormoi J, et al. Salivette, a relevant saliva sampling device for SARS-CoV-2 detection. *J Oral Microbiol.* **2021**; 13(1):1920226.
8. Deulofeu M, García-Cuesta E, Peña-Méndez EM, et al. Detection of SARS-CoV-2 Infection in Human Nasopharyngeal Samples by Combining MALDI-TOF MS and Artificial Intelligence. *Front Med.* **2021**; 8:661358.
9. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol Rev.* **2012**; 36(2):380–407.
10. Bizzini A, Durussel C, Bille J, Greub G, Prod'hom G. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *J Clin Microbiol.* **2010**; 48(5):1549–1554.
11. Cobo F. Application of maldi-tof mass spectrometry in clinical virology: a review. *Open Virol J.* **2013**; 7:84–90.
12. Calderaro A, Arcangeletti MC, Rodighiero I, et al. Identification of different respiratory viruses, after a cell culture step, by matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS). *Sci Rep.* **2016**; 6:36082.
13. Nachtigall FM, Pereira A, Trofymchuk OS, Santos LS. Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. *Nat Biotechnol.* **2020**; 38(10):1168–1173.
14. Rocca MF, Zintgraff JC, Dattero ME, et al. A combined approach of MALDI-TOF mass spectrometry and multivariate analysis as a potential tool for the detection of SARS-CoV-2 virus in nasopharyngeal swabs. *J Virol Methods.* **2020**; 286:113991.
15. Gomila RM, Martorell G, Fraile-Ribot PA, et al. Use of Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry Analysis of Serum Peptidome to Classify and Predict Coronavirus Disease 2019 Severity. *Open Forum Infect Dis.* **2021**; 8(6):ofab222.

16. Lazari LC, Ghilardi FDR, Rosa-Fernandes L, et al. Prognostic accuracy of MALDI-TOF mass spectrometric analysis of plasma in COVID-19. *Life Sci Alliance*. **2021**; 4(8):e202000946.
17. Costa MM, Benoit N, Saby F, Pradines B, Granjeaud S, Almeras L. Optimization and Standardization of Human Saliva Collection for MALDI-TOF MS. *Diagnostics* [Internet]. **2021**; 11(8). Available from: <https://www.mdpi.com/2075-4418/11/8/1304>
18. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull*. **2020**; 25(3).
19. Amrane S, Tissot-Dupont H, Doudier B, et al. Rapid viral diagnosis and ambulatory management of suspected COVID-19 cases presenting at the infectious diseases referral hospital in Marseille, France, - January 31st to March 1st, 2020: A respiratory virus snapshot. *Travel Med Infect Dis*. **2020**; 36:101632.
20. Nebbak A, Willcox AC, Bitam I, Raoult D, Parola P, Almeras L. Standardization of sample homogenization for mosquito identification using an innovative proteomic tool based on protein profiling. *Proteomics*. **2016**; 16(24):3148–3160.
21. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. [Royal Statistical Society, Wiley]; **1995**; 57(1):289–300.
22. Gibb S, Strimmer K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinforma Oxf Engl*. **2012**; 28(17):2270–2271.
23. Palarea-Albaladejo J, Mclean K, Wright F, Smith DGE. MALDIrppa: quality control and robust analysis for mass spectrometry data. *Bioinforma Oxf Engl*. **2018**; 34(3):522–523.
24. Savitzky Abraham, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem*. American Chemical Society; **1964**; 36(8):1627–1639.
25. Ryan CG, Clayton E, Griffin WL, Sie SH, Cousens DR. SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater At*. **1988**; 34(3):396–402.
26. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Transact A Math Phys Eng Sci*. **2016**; 374(2065):20150202.
27. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv180203426 Cs Stat* [Internet]. **2020** [cited 2021 Jun 1]; . Available from: <http://arxiv.org/abs/1802.03426>
28. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *J Stat Softw* [Internet]. Foundation for Open Access Statistics; **2008** [cited 2021 Jun 1]; 025(i01). Available from: <https://ideas.repec.org/a/jss/jstsof/v025i01.html>
29. Konopka T. tkonopka/umap [Internet]. 2021 [cited 2021 Jun 1]. Available from: <https://github.com/tkonopka/umap>
30. Kuhn M. Classification and Regression Training [R package caret version 6.0-88] [Internet]. Comprehensive R Archive Network (CRAN); 2021 [cited 2021 Jun 1]. Available from: <https://CRAN.R-project.org/package=caret>

31. Kumar V, editor. Introduction to parallel computing: design and analysis of algorithms. Redwood City, Calif: Benjamin/Cummings Pub. Co; 1994.
32. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinforma Oxf Engl*. **2005**; 21(20):3940–3941.
33. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, et al. False-negative results of initial RT-PCR assays for COVID-19: A systematic review. *PLoS One*. **2020**; 15(12):e0242958.
34. Cuénod A, Foucault F, Pflüger V, Egli A. Factors Associated With MALDI-TOF Mass Spectral Quality of Species Identification in Clinical Routine Diagnostics. *Front Cell Infect Microbiol*. **2021**; 11:646648.
35. Drevinek M, Dresler J, Klimentova J, Pisa L, Hubalek M. Evaluation of sample preparation methods for MALDI-TOF MS identification of highly dangerous bacteria. *Lett Appl Microbiol*. **2012**; 55(1):40–46.
36. Aita A, Basso D, Cattelan AM, et al. SARS-CoV-2 identification and IgA antibodies in saliva: One sample two tests approach for diagnosis. *Clin Chim Acta Int J Clin Chem*. **2020**; 510:717–722.
37. Esser D, Alvarez-Llamas G, Vries MP de, Weening D, Vonk RJ, Roelofsen H. Sample Stability and Protein Composition of Saliva: Implications for Its Use as a Diagnostic Fluid. *Biomark Insights*. **2008**; 3:25–27.
38. Dijkstra M, Vonk RJ, Jansen RC. SELDI-TOF mass spectra: a view on sources of variation. *J Chromatogr B Analyt Technol Biomed Life Sci*. **2007**; 847(1):12–23.
39. Cairns DA, Thompson D, Perkins DN, Stanley AJ, Selby PJ, Banks RE. Proteomic profiling using mass spectrometry--does normalising by total ion current potentially mask some biological differences? *Proteomics*. **2008**; 8(1):21–27.
40. Bauer C, Cramer R, Schuchhardt J. Evaluation of peak-picking algorithms for protein mass spectrometry. *Methods Mol Biol Clifton NJ*. **2011**; 696:341–352.
41. Tran NK, Howard T, Walsh R, et al. Novel application of automated machine learning with MALDI-TOF-MS for rapid high-throughput screening of COVID-19: a proof of concept. *Sci Rep*. **2021**; 11(1):8219.
42. Iwasaki S, Fujisawa S, Nakakubo S, et al. Comparison of SARS-CoV-2 detection in nasopharyngeal swab and saliva. *J Infect*. **2020**; 81(2):e145–e147.
43. Lippi G, Mattiuzzi C, Bovo C, Plebani M. Current laboratory diagnostics of coronavirus disease 2019 (COVID-19). *Acta Bio-Medica Atenei Parm*. **2020**; 91(2):137–145.
44. Chen Y, Yao H, Zhang N, et al. Proteomic Analysis Identifies Prolonged Disturbances in Pathways Related to Cholesterol Metabolism and Myocardium Function in the COVID-19 Recovery Stage. *J Proteome Res*. **2021**; 20(7):3463–3474.
45. Ong SWX, Fong S-W, Young BE, et al. Persistent Symptoms and Association With Inflammatory Cytokine Signatures in Recovered Coronavirus Disease 2019 Patients. *Open Forum Infect Dis*. **2021**; 8(6):ofab156.
46. Iles RK, Zmuidinaite R, Iles JK, Carnell G, Sampson A, Heeney JL. Development of a Clinical MALDI-ToF Mass Spectrometry Assay for SARS-CoV-2: Rational Design and Multi-Disciplinary Team Work. *Diagn Basel Switz*. **2020**; 10(10):E746.

47. Chivte P, LaCasse Z, Seethi VDR, et al. MALDI-ToF Protein Profiling as Potential Rapid Diagnostic Platform for COVID-19. *medRxiv*. **2021**; :2021.05.26.21257798.
48. Vaz SN, Santana DS de, Netto EM, et al. Saliva is a reliable, non-invasive specimen for SARS-CoV-2 detection. *Braz J Infect Dis Off Publ Braz Soc Infect Dis*. **2020**; 24(5):422–427.
49. Azzi L, Carcano G, Gianfagna F, et al. Saliva is a reliable tool to detect SARS-CoV-2. *J Infect*. **2020**; 81(1):e45–e50.
50. Costa MM, Benoit N, Tissot-Dupont H, et al. Mouth Washing Impaired SARS-CoV-2 Detection in Saliva. *Diagnostics [Internet]*. **2021**; 11(8). Available from: <https://www.mdpi.com/2075-4418/11/8/1509>
51. Bhattarai KR, Kim H-R, Chae H-J. Compliance with Saliva Collection Protocol in Healthy Volunteers: Strategies for Managing Risk and Errors. *Int J Med Sci*. **2018**; 15(8):823–831.

## Supporting information files

**Additional Table S1.** List of peaks distinguishing SARS-CoV-2 positive from negative samples after the application of Wilcoxon rank sum test ( $p < 0.05$ ).

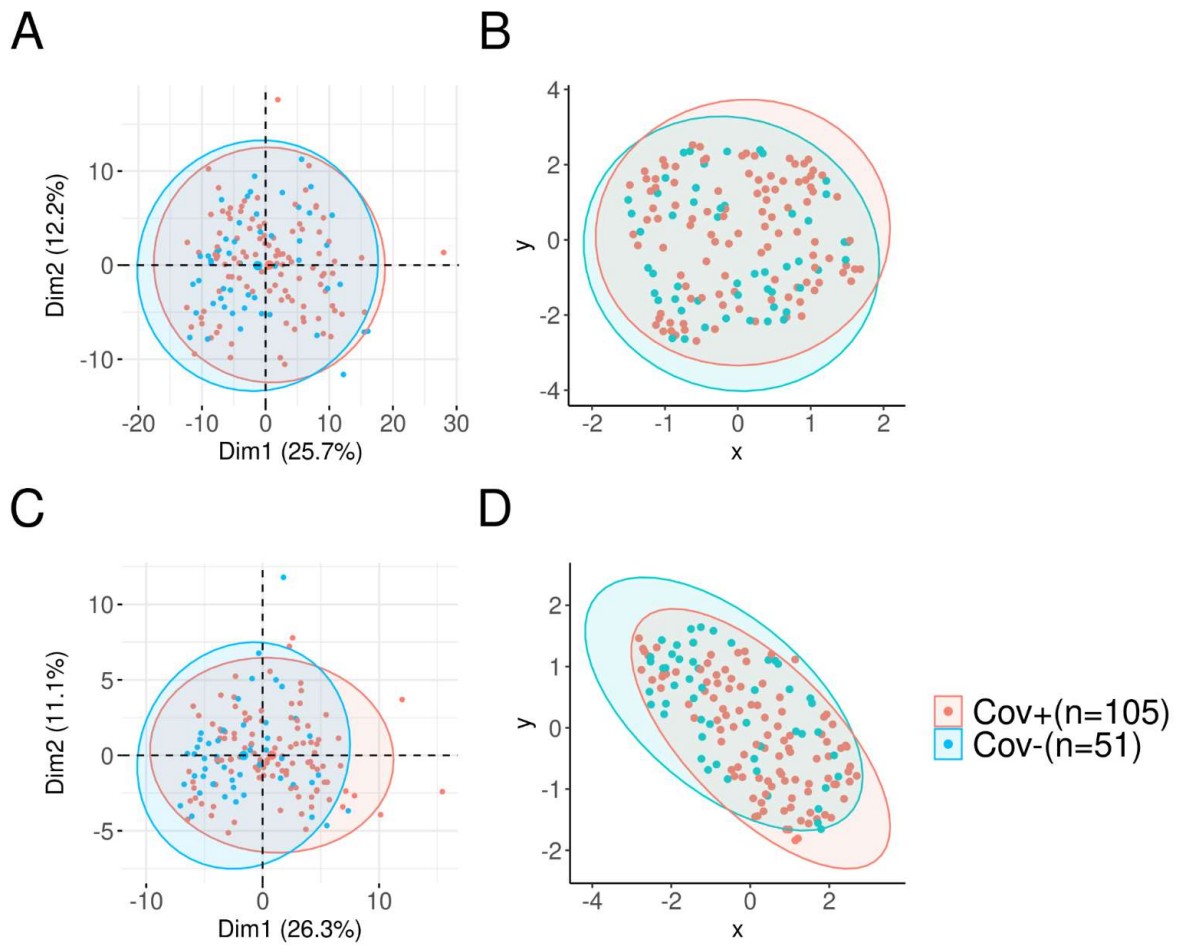
**Additional Table S1.** List of peaks distinguishing SARS-CoV-2 positive from negative samples after the application of Wilcoxon rank sum test ( $p < 0.05$ ).

Peaks (m/z)	p-value	**Ratio
2047.9	7.9e-04	31.1
*2102.4	1.9e-04	-41.4
2166.6	4.5e-02	16.0
2202.9	1.5e-03	-34.1
*2489.2	2.1e-07	81.3
2520.3	1.8e-02	14.9
2580.6	1.7e-02	14.0
2597	3.9e-02	-33.3
2687.7	4.7e-02	-27.7
2938.5	2.7e-03	52.7
*2977.2	9.1e-05	53.9
3203.5	4.8e-02	32.1
3220.3	1.0e-02	24.2
3494.4	5.6e-03	17.0
*3522.3	2.5e-05	60.1
3980.6	1.3e-03	35.6
4033.5	1.2e-03	80.9
4062.2	2.7e-02	50.8
4129.9	1.4e-02	-35.1
4191	3.2e-02	33.0
4394.1	8.1e-03	30.5
4449.2	3.5e-03	-45.1
4612.1	1.5e-02	30.4
4658.9	2.5e-02	16.9
4964.9	3.1e-02	41.9
5034.6	3.8e-03	92.7
*5063.3	3.2e-05	40.9
5099.3	3.7e-02	17.9
5134.5	2.5e-02	20.4
5153.9	1.1e-03	22.2
5233.9	9.9e-03	57.2
5380.4	1.3e-03	-38.3
*5418.9	1.4e-05	-58.5
*5463.3	3.3e-04	23.4
5503	4.9e-02	16.1
*5613.4	8.7e-05	37.7
5652.5	3.8e-03	8.8
5734.4	1.2e-02	22.3
*5814	3.0e-04	28.0
5867.2	1.3e-03	39.0
6305.7	1.3e-02	18.6
7376	1.4e-02	44.1
7647.3	4.5e-02	-37.9
7684.5	2.9e-03	-42.1
7853.8	3.0e-02	28.2
8145.2	5.9e-04	28.2
8300.7	1.3e-02	31.3
8741.5	3.1e-02	-30.1
9059.1	2.7e-02	35.4
9798.3	8.0e-03	60.3
9959.3	7.7e-03	-30.9
11004.3	2.7e-02	24.1

\*Peaks identified with a two-tailed Wilcoxon rank sum test ( $p < 0.05$ ) with a Benjamini-Hochberg correction ( $p < 0.1$ ).

\*\*Ratio computed from normalised mean intensities peaks between Cov- and Cov+ groups for the D0 time point.

**Additional Figure S1.** Reduction of dimension analysis performed from several sets of peaks for SARS-CoV-2(+) (n=105) and control (n=51). PCA (**A**) and UMAP (**B**) performed on all detected peaks. Ellipses define the boundary areas for the categorical variables projections PCA (**C**) and UMAP (**D**) performed on the 52 peaks selected with two-tailed Wilcoxon rank sum test ( $p < 0.05$ ). PCA: Principal Component Analysis; UMAP: Uniform Manifold Approximation and Projection.



**Table 1.** Characteristics of participants

	COVID-19 group <sup>a</sup>			Healthcare worker group		
	D0	D0+10	D0+30	D0	D0+10	D0+30
Collection time point <sup>b</sup>						
Participants, n (COVID+ <sup>c</sup> , n)	138 (105)	79 (16)	20 (0)	51	40	32
Age (years), median (IQR)	37.4 (23-52)	37.7 (24-52)	48 (39.5-57.3)	36.1 (27-45.5)	38.4 (28.8-46.8)	37.7 (27.8-46.8)
Male, n (%)	68 (49.3%)	42 (53.2%)	10 (50.0%)	22 (43.1%)	18 (45.0%)	16 (48.5%)
Onset of symptoms before D0 test (days), median (IQR)	2.2 (1-3)			/		
Symptoms at presentation, n (%)	89 (64.5%)			0 (0.0%)		
Headache, n (%)	38 (27.5%)			/		
Tiredness, n (%)	26 (18.8%)			/		
Cough, n (%)	24 (17.4%)			/		
Fever, n (%)	21 (15.2%)			/		
Myalgia, n (%)	20 (14.5%)			/		
Breathing difficulties, n (%)	12 (8.7%)			/		
Anosmia/Ageusia, n (%)	9 (6.5%)			/		
Sore throat, n (%)	7 (5.1%)			/		
Diarrhea, n (%)	7 (5.1%)			/		

Others, n (%)      4 (2.9%)

/

---

<sup>a</sup>Tested positively for SARS-CoV-2 by RT-qPCR on NPSs less than five day before enrollment. <sup>b</sup>Saliva sampled ten (D0+10) and thirty (D0+30) days after the first collection (D0). <sup>c</sup>Tested positively for SARS-CoV-2 by RT-qPCR on NPSs the sampling day (D0). Abbreviations: IQR, interquartile range; NPS, nasopharyngeal swab; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.