Body Volatilome Study Strategy for COVID-19 Biomarker Identification Considering Exogenous Parameters

Elsa Boudard, Nabil Moumane, José Dugay, Jérôme Vial, Didier Thiébaut

separations

MDPI

Article

Body Volatilome Study Strategy for COVID-19 Biomarker Identification Considering Exogenous Parameters

Elsa Boudard [1,2,*], Nabil Moumane [2], José Dugay [1], Jérôme Vial [1] and Didier Thiébaut [1]

[1] Unité Mixte de Recherche Chimie Biologie Innovation (UMR CBI), Laboratoire des Sciences Analytiques, Bioanalytiques et Miniaturisation, Ecole Supérieure de Physique et de Chimie Industrielle de la ville de Paris (ESPCI Paris), PSL Research University, 10 rue Vauquelin, 75005 Paris, France; jose.dugay@espci.fr (J.D.); jerome.vial@espci.fr (J.V.); didier.thiebaut@espci.fr (D.T.)

[2] SenseBiotek Health-Care, 21 Grande rue, 78240 Aigremont, France; an.moumane@sensedetecthealthcare.com

* Correspondence: elsa.boudard@espci.fr

Abstract: Since the 1950s, the screening of the body volatilome has proven to be a powerful tool for preventing diseases from spreading. Following the COVID-19 crisis, several studies began investigating the connection between viruses and the body volatilome, gradually identifying potential biomarkers, which varied depending on the study. To try to elucidate potential sources of inconsistency in biomarker findings, we decided to set up a study taking into consideration the factors often overlooked in previous studies. The VOCs constituting the body volatilomes of 40 COVID-19 patients and 13 healthy subjects were sampled by using PowerSorb® as the sorbent phase. Thermodesorption, followed by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry (TD-GC×GC/TOF MS), was utilized for the analysis. A non-targeted biomarker research methodology compared Covid(+) and Covid(−) chromatograms, assessing statistical significance and peak area changes. Out of the 25 compounds highlighted, 13 associated with cosmetic products were excluded, and 8 linked to air pollution in urban settings were also excluded. Finally, after having quantitatively evaluated the potential sources of the compounds (cosmetic or environmental), 4 compounds remained and their relevance was assessed using ROC curves. Among them, hexanoic acid, 2-ethyl- identification was confirmed with standard and led to an area-under-the-curve value of 92%. More in-depth studies are needed to investigate the specificity of the biomarker in relation to COVID-19, but the strategy of this study shows how to avoid obtaining data that are biased by exogenous factors.

Keywords: body odor; body volatilome; VOCs; GC×GC; COVID-19

Citation: Boudard, E.; Moumane, N.; Dugay, J.; Vial, J.; Thiébaut, D. Body Volatilome Study Strategy for COVID-19 Biomarker Identification Considering Exogenous Parameters. *Separations* **2024**, *11*, 336. https://doi.org/10.3390/separations11120336

Academic Editor: Victoria Samanidou

Received: 3 October 2024
Revised: 15 November 2024
Accepted: 18 November 2024
Published: 22 November 2024

1. Introduction

Improving people's lives has always been the primary goal of research. One way to achieve this is by preserving health through early intervention at the onset of disease, with diagnosis being the first critical step. Currently, there is still significant potential to enhance diagnostics by making them more accessible, earlier to perform, and less invasive. In this context, the study of the body volatilome has gained increasing attention since the 1950s. The body volatilome comprises a diverse array of volatile organic compounds (VOCs) with varying polarity and volatility. These primarily result from the interaction between compounds secreted by sweat glands (apocrine, eccrine, and sebaceous) and microorganisms present on the skin's surface (bacteria, fungi, viruses, etc.) [1]. Changes in metabolism and skin barrier disruption and microbiome alterations can lead to shifts in the composition of the body volatilome, potentially aiding in the early detection of disease [2]. Early instances of disease detection through body volatilome were often linked to the production of odoriferous compounds, which gave patients a distinct body odor. For example, metabolic disorders like phenylketonuria result in the emission of a musty odor due to the accumulation of phenylalanine in body fluids. This is caused by the abnormal

metabolism of phenylalanine into phenylpyruvic acid and phenylacetate [3]. Viral and infectious diseases have also been associated with specific body odors, even when the responsible molecules have not been identified: there is a foul odor for scarlet fever [4], a sweet and pungent smell for smallpox [5], a beer-like odor for scrofula [6], and an odor like a butcher's shop body for yellow fever [5]. To deepen the understanding of these phenomena, gas chromatography–mass spectrometry (GC-MS) has been increasingly used to study the body volatilome in the context of disease detection, including for melanoma [7–9], Parkinson's disease [10], schizophrenia [11], and certain cancers [12].

In December 2019, the SARS-CoV-2 coronavirus, the causative agent of COVID-19, emerged in Wuhan, China. By June 2023, this virus had infected over 760 million people and caused nearly 7 million deaths worldwide [13]. SARS-CoV-2 primarily spreads through the inhalation of viral particles and binds to cells expressing angiotensin-converting enzyme 2 (ACE2) receptors [14]. While it predominantly affects the respiratory system, it can also impact the urinary, nervous, endocrine, immune, and reproductive systems. Given the virus's virulence, it quickly became crucial to identify infected individuals to prevent further transmission. The gold standard for testing at the time was Real-Time Reverse Transcriptase Polymerase Chain Reaction (RT-PCR). However, alternative approaches were also developed, including sensor-based methods [15,16] and the analysis of volatile organic compounds (VOCs) in exhaled air, the latter of which has been studied a lot more than the body volatilome [17,18].

The earliest studies using the body volatilome to detect SARS-CoV-2 primarily utilized dogs' powerful sense of smell [19–22]. In 2023, the first analytical study on the body volatilome for COVID-19 detection was published [23]. In this context, the body volatilome matrix had the added advantage of containing no viral load [24]. In their study, a research team based in Thailand and led by Dr. Kulsing used cotton swabs to collect sweat samples from the armpits of 82 people affected by COVID-19 and 286 healthy subjects, analyzing them using static headspace–gas chromatography with a flame ionization detector (SHS-GC-FID), and headspace solid-phase microextraction followed by comprehensive two-dimensional gas chromatography coupled with mass spectrometry (HS-SPME GC×GC-MS) [23]. Their research identified p-cymene as a compound present in significantly higher amounts in samples from COVID-19 patients. The authors also demonstrated that the levels of this compound were not significantly influenced by factors such as gender, body mass index, age, the use of deodorants, or diet (e.g., orange juice consumption). However, it is worth noting that the study did not specify whether healthy and sick subjects were sampled in the same environment, potentially overlooking this factor.

In a subsequent study, the same team identified nonanal as a volatile biomarker for COVID-19 and used it to develop a fluorescence-based sensor, achieving a diagnostic accuracy of 94% [25]. In 2024, the Thai team published two additional studies related to COVID-19 biomarkers and the body volatilome. Using GC×GC, they achieved similar levels of accuracy, sensitivity, and selectivity as the sensor-based method, but this time used p-cymene, linalool, and 2,6,11-trimethyldodecane as biomarkers [26]. In another study, they used GC-MS to analyze a set of molecules, including pentadecane, nonanal, styrene, xylene, ethylbenzene, and 2-ethylhexyl acrylate, again achieving a diagnostic accuracy of 94% [27]. Lastly, they demonstrated changes in the body volatilome during the 6 h following COVID-19 vaccination, this time using gas chromatography–ion mobility spectrometry (GC-IMS) [28]. In this study, 55 features were highlighted as significantly different over time using a non-targeted method, but only 4 specific ones were investigated in detail, including methanol. On the basis of these features, differences were noted regarding the volatilome depending on the type of vaccine (Pfizer vs. Moderna). While these results are promising and indicate strong diagnostic potential for COVID-19, it is important to note that, to our knowledge, they are limited to a single research team. Furthermore, according to the papers published by this research team, the identified biomarkers were not always the same, thereby justifying our present investigation to elucidate potential sources of variability.

In this study, we analyzed samples from 40 COVID-19 patients and 13 healthy individuals to comparatively investigate their body volatilomes. Samples were collected using the PowerSorb® sorbent phase and analyzed to identify potential significant differences. Thermodesorption, followed by comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (TD-GC×GC/TOF MS), was used for this analysis. This method was selected due to the expected high number of analytes, as it offers increased peak capacity [29] and sensitivity [30]. The aim of the study was not only to identify potential biomarkers but also to develop a reliable research methodology incorporating chemometric tools, while carefully considering potential external biases such as the influence of cosmetics or environmental VOCs.

2. Materials and Methods

2.1. Clinical Study

Sample collection started in 2022 after obtaining an agreement from the French Personal Protection Committee of South-West and overseas (ID RCB: 2021-A01322-39—CPP 1-21-092/21.00252.000039). Overall, 40 patients were sampled and included in the Covid(+) group after confirming their health status with a positive PCR or antigenic test, respectively, carried out between 96 h and 48 h before inclusion. This allowed us to include patients who were at the height of their symptoms. The Covid(+) patients were sampled at Angers UHC, Paris St Antoine hospital, or Amiens UHC, where they were hospitalized because of COVID-19. Overall, 13 individuals were sampled and included in the Covid(−) group as they had negative antigenic test results less than 24 h prior. The Covid(−) individuals were sampled at the ESPCI (Paris, France) or in a partner canine training center involved in COVID-19 detection (French K9 Academy, Parthenay, France). They were not sampled in the hospital due to logistical constraints. No specific restrictions were given to patients, but samples were taken minimum 2 h after meals.

2.2. Sampling

Before sampling, the area of skin to be sampled was cleaned with an odorless cleansing wipe (Ront), removing potential cosmetic residues. Then, after allowing the area to air dry, 2 PowerSorb® (2 × 20 mm, Action Europe, Sausheim, France) were attached to the skin for 1 h. We placed a small box in polypropylene/acrylonitrile butadiene (PP/AB), isolating it from the outside environment, and performed dressing to fix it. PowerSorb® acted as a sorbent phase, capturing the VOCs emitted by the skin of the sampled individual for 1 h. Before sampling, PowerSorb® were individually cleaned using the tube conditioner TC-20 (Markes International, Bridgend, UK) for 4 h at 240 °C with a N2 flow at 50 mL/min. After each sampling, the sampling system was detached from the skin and the PowerSorb® were put in 2 mL glass vials. These were stored between 4 and 8 °C before being sent to the ESPCI for TD-GC×GC/TOF MS analysis. For each batch of 20 conditioned PowerSorb®, 2 PowerSorb® were kept for analyzed by TD-GC×GC/TOF MS as blanks.

The Covid(+) patients were sampled from the armpits, hands, and groin to investigate whether the associated odor of the infection extends across the entire body or is predominantly expressed through specific glands. Covid(−) individuals were only sampled from the armpits. As all individuals were sampled from both sides of the body (left and right), samples from one side were sent to the ESPCI for TD-GC×GC/TOF MS analysis (samples from the other side were sent to the French K9 Academy for dog training). This was made possible thanks to preliminary results, which indicated that the VOCs emanating from both sides of the body were analogous. These results were obtained by our team after comparing left/right hands, armpits, and forearm of 19 individuals [31].

As the sampling system used at the time emitted some VOCs, six blanks were run to assess these emissions and avoid confusing them with the compounds actually emitted by the sampled individuals. To do so, 2 conditioned PowerSorb® were put in the PP/AB box, which was then wrapped in bandage strip and placed in an oven at 40 °C for 1 h to simulate body heat. The PowerSorb® were then recovered and analyzed via TD-GC×GC/TOF MS.

2.3. TD-GC×GC/TOF MS Analysis

For the analysis, the 2 PowerSorb® used for each sample were put together in a stainless steel thermodesorption tube (6 × 89 mm). Thermodesorption was performed using a TD 100-xr obtained from Markes (Markes International, Bridgend, UK). Stainless steel tubes, containing the PowerSorb®, were desorbed at 220 °C for 20 min with a helium (Alphagaz 1, Air Liquide, Ile-de-France, France) flow rate of 50 mL/min into the built-in cold trap "General Purpose" C4/5-C30/32 UNITY2 at −10 °C containing graphitized carbon. The cold trap was finally heated for 5 min at 60–100 °C/s to 320 °C with a counter-current flow of 4.2 mL/min for desorption into the GC×GC for injection using a 3 mL/min split, which led to a split ratio of 3.5.

The system employed to perform the two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC/TOF MS) consisted of a LECO Pegasus BT4D instrument (LECO, Villepinte, France). Following the process of thermodesorption, the analytes were transferred to the GC×GC system by the carrier gas, helium, at a flow rate of 1.2 mL/min. In the primary oven, the first dimension (1D), separation was performed using a non-polar Rxi-5ms column, 5% diphenyl and 95% dimethyl polysiloxane, with dimensions of 30 m × 0.5 mm × 0.25 µm (Restek, Lisses, France). In the second dimension (2D), a medium polar DB1701 14% cyanopropyl-phenyl-methylpolysiloxane column with dimensions of 50 cm × 0.18 mm × 0.18 µm (Agilent) was utilized in the secondary oven. A temperature program was applied. This started at 35 °C. The temperature was held for 2 min, and ramped up at a rate of 3 °C/min until reaching 230 °C. The same temperature program was applied in the secondary oven with an offset of 15 °C. The modulation step was facilitated by a QuadJet™ nitrogen cryogenic modulator, employing a modulation period of 3 s, which comprised two cycles of 1.5 s each, which were divided into 0.90 s of hot jet and 0.6 s of cold jet. Subsequently, the analytes were transferred to the mass spectrometer through a 31 cm × 0.18 mm transfer line composed of deactivated silica, operating at a temperature of 250 °C.

The mass spectrometer was operated with an electron ionization source set to 70 eV and 250 °C, utilizing a scan range of 45–300 *m/z* at a scan frequency of 200 Hz. Data acquisition was performed using ChromaToF software version 5.51.50 (LECO).

2.4. Data Processing and Biomarker Research Methodology

When a consequential amount of data were obtained, these data were investigated in the first instance using the LECO software ChromaToF Tile (version 1.01). Data obtained from different sampling body areas (hands, armpits, groin) were combined to search for biomarkers that would be common. This software allows for untargeted research of compounds that are significantly different between the pre-defined groups (here, Covid(+) and Covid(−)) by examining the chromatograms zone by zone and using Fisher's statistics [32,33]. A zone was defined as a tile and its size was set at 3 × 29 (modulations × spectras). The signal-to-noise (S/N) threshold was set at 1000 and the significance threshold was set using a F-ratio corresponding to a *p*-value of 0.0001. This step allowed the initial sorting, returning a list of hits which corresponded to compounds seen as significantly different in the body volatilomes of healthy and ill individuals.

To provide an additional level of verification, each previously identified hit was manually checked on chromatograms using ChromaToF software (version 5.55.41). Each chromatogram had been previously processed by applying a Peak Finding method to highlight the peaks exceeding a S/N threshold of 1000. The purpose of this verification step was to remove any hits highlighted by ChromaToF Tile that were actually artifacts originating from column bleeding, baseline instability, etc.

In the third stage, the remaining hits were evaluated on a fold change basis (FC_1 and FC_1') using Formulas (1) and (2). The threshold was set to 10. This allowed us to keep only

compounds whose areas were 10 times higher in one of the groups (Covid(+) or Covid(−)), i.e., compounds that were either overexpressed or underexpressed for sick patients.

$$FC_1 = \frac{\overline{x_{Covid(+)}}}{\overline{x_{Covid(-)}}} (\text{eventual overexpressed compound in Covid}(+)\text{group}) \quad (1)$$

$$FC_1' = \frac{\overline{x_{Covid(-)}}}{\overline{x_{Covid(+)}}} (\text{eventual underexpressed compound in Covid}(+)\text{group}) \quad (2)$$

$\bar{x}$: mean area value for a given compound in the Covid(+) or Covid(−)group

Finally, to avoid any bias related to the sampling process, a second fold change (FC_2) analysis was performed in order to compare the area of a given compound, in the body volatilome samples of the group in which it was overexpressed, to its area in blank samples using Formula (3). The compound was finally kept if this second fold change exceeded a threshold of 2.

$$FC_2 = \frac{\bar{x}_{Covid\ group}}{\bar{x}_{blank}} \quad (3)$$

After these filtering steps, the remaining compounds were observed and sorted on the basis of attributions made by comparing their mass spectra with those seen in the NIST database. For example, if a compound such as linalool was observed, it was placed in the "potential cosmetic effect" category as it is a commonly used ingredient in cosmetics. This sorting is discussed in more detail in Sections 3.1.1 and 3.1.2. Finally, regarding the remaining compounds, which were not excluded as a safety measure in order to avoid exogenous influence, their receiver operating characteristic (ROC) curves were plotted and their *p*-values were re-calculated using the Wilcoxon test [34,35]. This test was used here as it is a non-parametric test and as the obtained *p*-value is directly linked to the area under the curve (AUC) of ROC curves.

2.5. Quantification

For the 5 compounds kept as potential volatile biomarkers of COVID-19, two attributions were confirmed using standards of nonanoic acid and 2-ethyl hexanoic acid (Sigma-Aldrich, Saint-Quentin-Fallavier, France). Calibration curves were produced to quantify the presence of these compounds in the samples. To achieve this, solutions were prepared in methanol at the following concentrations for each compound: 5, 10, 15, 30, and 40 mg/L. Then, 1 µL of each solution was spiked on 2 cleaned PowerSorb® using a calibration solution loading rig (CSLR) from Markes, meaning that the following masses were spiked: 5, 10, 15, 30, and 40 ng. Three replicates were produced per concentration. Spiked PowerSorb® were analyzed using the same TD-GC×GC/TOF MS method as described in Section 2.3. Calibration curves are displayed in Figure S1.

3. Results and Discussion

3.1. Search for Biomarkers

Overall, 130 samples related to the COVID-19 clinical study were analyzed using TD-GC×GC/TOF MS. Each chromatogram contained approximately 1000 peaks, which explains the impossibility of processing the data manually. Thus, the biomarker research methodology described in Section 2.4 was applied to the data. This method allowed us to filter the information at each stage in order to retain only the most relevant compounds. The process and its intermediate results are described in Figure 1. This shows that, from the whole dataset, 25 compounds were retained due to the discovery of significant differences between the Covid(+) and Covid(−) groups (see Tables 1–3), with an area at least 10 times greater on average in one group compared to the other and at least 2 times greater compared to that observed in the blank of the sampling system. Despite the effectiveness of the chemometric tools employed, a critical assessment was conducted to evaluate the relevance of the 25 compounds identified to the pathology. Based on their attribution, which was derived through a comparison of mass spectra with the NIST database (matching score > 800),

the potential cosmetic or environmental origins of the compounds were investigated and are detailed in Sections 3.1.1 and 3.1.2.

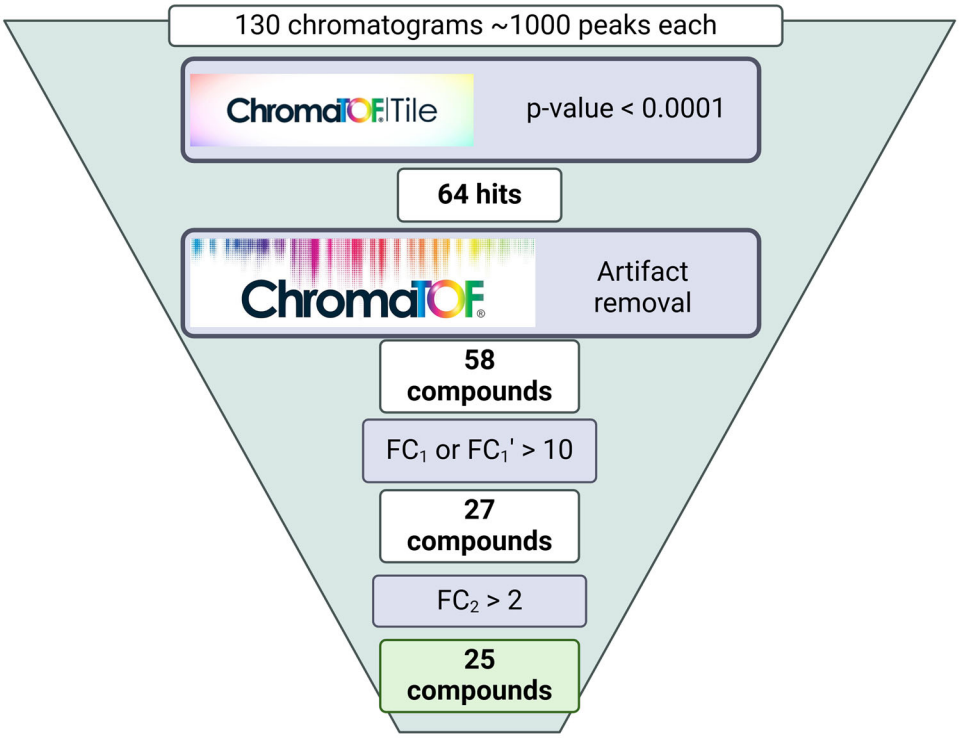

Figure 1. A schematic representation of the research methodology used for volatile biomarkers associated with COVID-19. The ChromaToF Tile was used as a non-targeted step to obtain a list of statistically significant hits, the fold changes FC_1 and FC_1' allowed us to keep only compounds 10 times over- or underexpressed for COVID-19 patients, and FC_2 allowed us to eliminate interfering compounds coming from the sampling device. (The figure was made with BioRender).

3.1.1. Cosmetic Influence

In this study, no restrictions were given to patients in terms of cosmetics, which is sometimes the case when studying the body volatilome. Even if the true impact these products can have on body volatilome has not yet been demonstrated, their use was sometimes prohibited in other studies for 24 h [36], 48 h [37], or 7 days [38] before sampling. Here, among the 25 compounds obtained as potential biomarkers, it was decided exclude 13 as their origin had a high probability of being exogenous. This was decided after being able to find these molecules on sites such as ScenTree [39], INCI beauty [40], or The Good Scents Company [41], which list the ingredients used in perfumery and cosmetics. Table 1 summarizes the 13 molecules associated with cosmetics that were observed in the body volatilome in the present study, giving their trade names when available. Moreover, these 13 compounds were all overexpressed for healthy individuals. It would therefore seem rational to think that patients who were hospitalized because they suffered from severe symptoms would be less concerned about certain aesthetic aspects and would use fewer cosmetic products at that time. Nevertheless, to avoid this type of influence in future studies, a restriction on the use of cosmetic products will be added before sampling.

Table 1. A list of the 13 compounds overexpressed in the Covid(−) group and attributed to a cosmetic influence.

Identified Compound	Trade Name	RT ^{1}D	RT ^{2}D	FC_1'
Octanal, 2-(phenylmethylene)-	Jasmonal A®	2834.8	2.13	16.5
Octanal, 2-(phenylmethylene)-	Jasmonal A®	2885.8	2.05	16.5
Benzenemethanol, α-(trichloromethyl)-, acetate	Rosacetol®	2417.8	2.08	14.3
Amberonne (isomer 1)	Amberonne®	2639.8	1.99	11.2
Cyclopenta[g]-2-benzopyran, 1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethyl-	Galaxolide®	3050.8	1.89	10.4
3-Buten-2-one,4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-	B-Ionone	2252.9	2.1	24.4
Cedrol	Cedrol	2519.8	2.02	12.0
α Isomethyl ionone	Isoraldeine 70®	2234.9	2.02	26.5
Benzoic acid, 2-hydroxy-, 2-methylbutyl ester	/	2453.8	1.99	17.5
Cyclopenta[g]-2-benzopyran, 1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethyl-	Galaxolide®	3158.8	1.93	29.2
Cyclopenta[g]-2-benzopyran, 1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethyl-	Galaxolide®	3095.8	1.93	64.7
4,7-Methano-1H-indenol, hexahydro-	Cyclacet	1712.9	2.27	11.3
Cyclopenta[g]-2-benzopyran, 1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethyl-	Galaxolide®	3137.8	1.92	12.5

Retention time for the first dimension (RT ^{1}D) and retention time for the second dimension (RT ^{2}D) are both given in seconds.

3.1.2. Environmental Influence

After setting aside the compounds associated with cosmetics, 7 other molecules (see Table 2) were excluded because they were suspected of being associated with VOCs in the ambient air. Indeed, these 7 molecules were alkanes or alkyl benzenes and the literature has already cited this type of molecule as being associated with air pollution [42,43]. This hypothesis was reinforced here by the fact that these molecules were overexpressed in healthy individuals, but only for those sampled in an urban environment (Paris). For benzene, (1-ethylnonyl)-, for example, the mean area value for healthy individuals sampled in urban environment was $510 \times 10^{6,}$ while for healthy individuals sampled in the countryside it was 3.5×10^6.

Table 2. A list of the 7 molecules overexpressed for the Covid(−) individuals sampled in urban environment (Paris).

Identified Compounds	RT ^{1}D	RT ^{2}D	FC_1'
Benzene, (1-ethylnonyl)-	2657.8	1.59	15.5
Benzene, (1-pentylheptyl)-	2795.8	1.56	16.7
Decane	977.9	1.38	12.6
Benzene, (1-pentylhexyl)-	2585.8	1.56	10.7
Benzene, (1-methyldecyl)-	2738.8	1.61	12.2
Benzene, (1-ethyldecyl)-	2873.8	1.59	13.9
Benzene, (1-methylnonyl)-	2513.8	1.61	12.7

However, as shown in Figure 2, a clear difference was observed between Covid(+) and Covid(−) chromatograms associated with the compound, which eluted at around 40 min in RT1D and 0.75 s in RT2D. This compound was identified as benzoic acid, 2,4-dichloro-, with matching scores above 900 in the different chromatograms. Although the difference in intensity regarding this compound was visually observable in the chromatograms, the compound had not been highlighted by the biomarker research methodology. As the "smell of the hospital" is commonly recognized as characteristic, the effect of VOCs present in the ambient air of this environment was suspected. To confirm this hypothesis, 2 healthy individuals previously sampled at the ESPCI laboratory were sent to one of the sampling centers (Paris St Antoine hospital) for sampling there. The hospital air was also sampled

at the same time by opening for 1 h a glass vial containing 2 PowerSorb®, which were subsequently analyzed by TD-GC×GC/TOF MS. Firstly, benzoic acid, 2,4-dichloro- was found in the 3 air samples. Secondly, looking at the arithmetic mean value of the peak area of this compound for the two healthy individuals sampled at the ESPCI and in the hospital, multiplicative factors up to 100 were obtained when individuals were at the hospital. Although more replicates on more individuals would be needed to fully confirm this hypothesis, it seems that ambient hospital air does have an effect on the body's volatilome. This finding was consistent with the results of previous hospital air studies, which had already identified the presence of chlorinated compounds [44,45]. According to those studies, most of the compounds found in hospital air were related to cleaning/disinfectant products and other types of chemical product used for different health activities. These results were also interesting in terms of the exposomic point of view highlighted in the review of Samon et al. [46]. As these tests were carried out over only half a day, even greater increases could be expected during longer hospital stays. To avoid this type of influence, future studies should ideally perform the sampling of the different groups in the same environment. However, if this is not possible, systematic ambient air sampling must be carried out simultaneously with volatilome sampling.

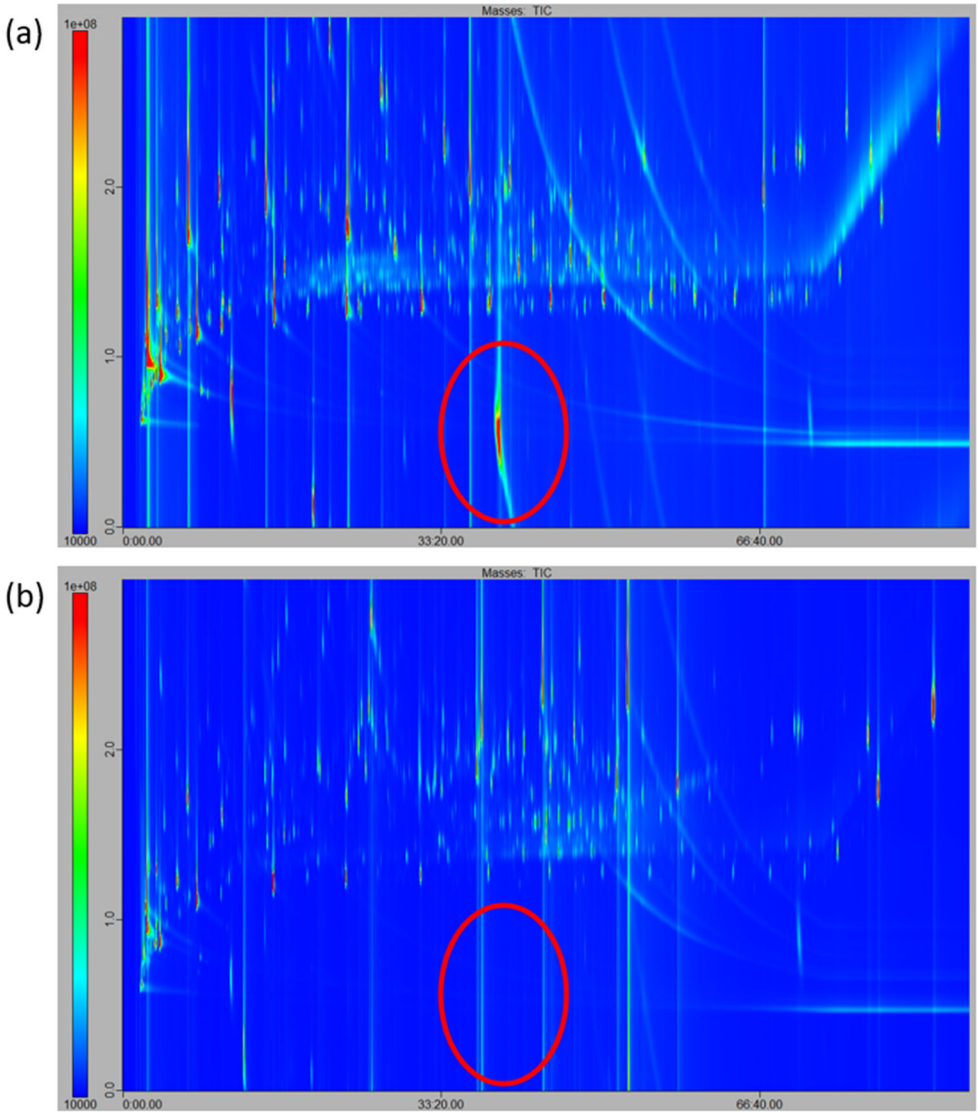

Figure 2. TD-GC×GC/TOF MS chromatograms of body volatilome sample for (**a**) a Covid(+) individual sampled in the hospital and (**b**) a healthy subject sampled in the ESPCI. The compound present in the area circled in red corresponds to the benzoic acid, 2,4-dichloro-.

3.2. COVID-19 Volatile Biomarkers

Using the previously described research methodology, and after the exclusions based on compound sources, we found that 5 molecules remained as potential COVID-19 volatile biomarkers: hexanoic acid, 2-ethyl-; nonanoic acid; oxime-, methoxy-phenyl-_ (B3); 1,4-methanobenzocyclodecene, 1,2,3,4,4a,5,8,9,12,12a-decahydro- (B4); naphthalene, 6,7-diethyl-1,2,3,4-tetrahydro-1,1,4,4-tetramethyl- (B5). The attribution of the two first compounds (hexanoic acid, 2-ethyl- and nonanoic acid) was confirmed by injecting standards. Among those compounds, 4 were overexpressed for Covid(+) and only B5 was underexpressed. The ROC curves of the 5 molecules are displayed in Figure 3. Good predictive results were obtained for the majority of the molecules with an AUC > 90%, except for B5, which had more chances to lead to misdiagnosis. Thanks to the method described in Section 2.5, hexanoic acid, 2-ethyl- and nonanoic acid were quantified after calculating the mean area value in each group using the arithmetic mean. This quantification represents the quantity of compound trapped in the PowerSorb® during the sampling period. The values obtained are presented in Table 3 with other information obtained for the 5 potential biomarkers.

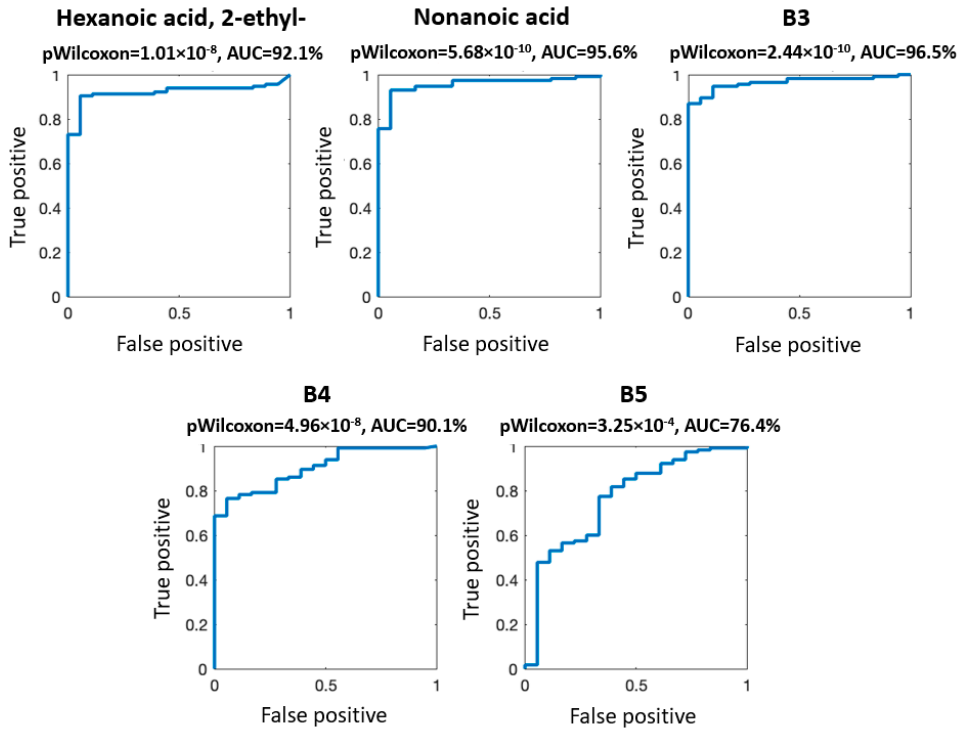

Figure 3. ROC curves of the 5 potential volatile biomarkers of COVID-19. *p*-values are obtained using Wilcoxon's test.

However, given the above observations, the 4 compounds overexpressed for Covid(+) were searched for in ambient air samples taken at one of the sampling center hospitals (Table 3). The results call into question the legitimacy of using nonanoic acid as a COVID-19 biomarker as this compound was found at an equivalent level in hospital ambient air. As healthy and sick people were sampled in different environments, this compound may be another marker of exposure to the hospital environment rather than a marker of the pathology studied. To avoid this type of undesirable influence, nonanoic acid was set aside, along with the 7 other molecules associated with the parasitic effects of the ambient environment. However, for hexanoic acid, 2-ethyl- and B4 the results were satisfactory as they were observed in more than 25 times greater quantities in Covid(+) patients than in the ambient air in the hospital. For B3, suspicions persist as it was present in quantities only twice larger in samples from patients than in hospital air. With no established threshold to

distinguish contamination from an endogenous compound, it was not possible to confirm the exact origin of the compound. Taking all these elements into consideration, hexanoic acid, 2-ethyl- seemed to be a very interesting biomarker that allowed us to accurately distinguish between Covid(+) and Covid(−) groups. More advanced studies are needed to determine whether it is a specific marker of COVID-19 or, more generally, a biomarker of a viral/respiratory infection. However, endogenous carboxylic acids have already been associated with some reactions, such as the breakdown or peroxidation of lipids present on the skin's surface or reactions caused by bacterial activity [8]. Moreover, this specific compound (hexanoic acid, 2-ethyl-) had already been detected and identified in sweat using GC/TOF MS [47].

Finally, as p-cymene and nonanal, previously highlighted as biomarkers of COVID-19 in the literature [23,25], were visible in our samples, it was decided to study them specifically, even though our research methodology had not highlighted them. The comparison of areas, which were collected manually using ChromaToF software (version 5.55.41), led to *p*-values lower than 0.0001 for both molecules when using Wilcoxon's test. However, FC_1 values were only 2.5 for both molecules. More importantly, the FC_2 values were 1.4 and 1.0, respectively, for p-cymene and nonanal. This indicates that, in our case, the molecules, although significantly different between groups, were in fact the result of a contamination of the sampling system. This was particularly true for p-cymene, for which the significance was probably be due to the different levels of expertise of the individuals performing the sampling: Covid(−) samples were collected by experienced individuals accustomed to the sampling system, whereas in the hospital, the nursing staff were less experienced and hospital constraints resulted in less time being available. This type of constraint, inducing potential false positives, had already been noted in the COVID-19 study of Phusrisom et al. [28]. For nonanal, the source of significance was unclear, as the compound was found in equivalent quantities in samples from Covid(+) individuals, in emissions from the sampling system, and in the hospital's ambient air.

Table 3. Data on the potential COVID-19 biomarkers: retention times in the first and second dimensions (RT ^{1}D, RT ^{2}D); the quantification of the mean compound mass trapped during the sampling for Covid(+) and Covid(−) groups; and the mean area value in the overexpressed group and mean area value in hospital air samples. Biomarkers overexpressed for Covid(+) patients are indicated by "^" and those underexpressed are given by "∨". Mean values were calculated using the arithmetic mean.

Biomarker	RT ^{1}D; RT ^{2}D (s)	Quantification in Covid(+) (ng)	Quantification in Covid(−) (ng)	FC_1	FC_1'	Mean Area in Overexpressed Group (10^6)	Mean Area in Hospital Air (10^6)
Hexanoic acid, 2-ethyl- ^	1295.9; 2.54	16.4	6.5	19.4	/	104	3.81
Nonanoic acid ^*	1712.9; 2.56	29.2	10.0			193	186
B3 ^	687; 0.79	/	/	10.1	/	5140	1890
B4 ^	2573.8; 1.78	/	/	11.9	/	181	5.4
B5 ∨	2969.8; 1.85	/	/	/	32.5	118	/

* excluded as it was linked to the hospital environment.

4. Conclusions

The TD-GC×GC/TOF MS analysis method has proven to be efficient and suitable for the study of this type of matrix. A biomarker research methodology has been implemented for this study, using chemometric tools such as Fischer ratio, employed through the software ChromaToF Tile, and fold change to verify the absence of contamination among the significant compounds. The measurement of statistical significance with an evaluation of the magnitude of the change was very useful for this type of data, where the variation in response is not evenly distributed. However, despite the use of these high-performance tools, it was necessary to remain vigilant and critical with regard to the results in order to eliminate compounds coming from exogenous sources. Indeed, no restrictions were

given on the use of cosmetics and therefore 13 compounds were found to be associated with that parameter. Thus, a restriction on the use of cosmetics, a minimum of 24 h before the sampling, will now be added for future studies. However, regarding the effect of the surrounding air, the same kind of restriction cannot be applied, and the collection of air samples seemed to be the only way to avoid bias related to that parameter. Here, 7 molecules were found to be linked to air pollution and, on the final list of 5 potential COVID-19 biomarkers, nonanoic acid was finally excluded after being found at the same level in hospital air as in Covid(+) samples. Finally, when looking at two molecules highlighted by the literature as COVID-19 biomarkers, we found ourselves unable to confirm these results because, although they were significantly different between groups, p-cymene and nonanal seemed to come mainly from the sampling system's VOCs. Nevertheless, 4 molecules were found to be potential volatile biomarkers of COVID-19 after excluding all the confounding factors mentioned. One of these molecules was the hexanoic acid, 2-ethyl-, which was quantified and found in Covid(+) samples at 16.4 ng after being trapped on the sorbent phase during 1 h of sampling. To sum up, in this study, we quantitatively evaluated the effect of environmental and cosmetic influences on the body's volatilome before finally proposing a biomarker. This biomarker has been identified in samples taken from individuals with COVID-19, but further studies are needed to determine whether it can be more generally associated with lung infections, for example, or whether it is pathology-specific. One way of achieving this would be to use, as a control, a group of asymptomatic patients who have been in contact with COVID-19 patients, and/or a group with symptoms who have not tested positive for COVID-19.

For future studies, individuals from positive and negative groups should be sampled in the same environment to minimize its impact. If not possible, environmental air should be sampled and an additional criterion based on a fold change should be implemented to avoid environmental bias. Improving the sampling system to reduce emissions and make it easier to use would also be beneficial, limiting false positives and avoiding the exclusion of certain molecules through FC_2. Finally, the use of balanced groups should be encouraged in order to reinforce the reliability of statistical tests.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/separations11120336/s1. Figure S1: Calibration curves for hexanoic acid, 2-ethyl- and nonanoic acid.

Author Contributions: Conceptualization, E.B., N.M., J.V., and D.T.; methodology, E.B.; software, E.B.; validation, E.B., N.M., J.V., and D.T.; formal analysis, E.B.; investigation, E.B.; resources, N.M., J.V., and D.T.; data curation, E.B.; writing—original draft preparation, E.B.; writing—review and editing, E.B., J.V., and D.T.; visualization, E.B.; supervision, J.D., J.V., and D.T.; project administration, N.M., J.V., and D.T.; funding acquisition, N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the ANRT (Association Nationale de la Recherche et de la Technologie) through PhD grant ("Bourse Cifre" #2021/0838).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank Isabelle Rivals, who has sadly passed away recently, but whose critical eye on the data and editing of the text contributed greatly to the work presented in this article.

Conflicts of Interest: The authors declare no conflicts of interest. Research reported in this publication was supported by SenseDetect Health-Care and the ANRT (Association Nationale de la Recherche et de la Technologie). The content is solely the responsibility of the authors, and the support provided did not influence the nature of the results presented in this paper.

References

1. Grice, E.A.; Segre, J.A. The skin microbiome. *Nat. Rev. Microbiol.* **2011**, *9*, 244–253. [CrossRef] [PubMed]
2. Pandey, S.K. Human body-odor components and their determination. *Trends Anal. Chem.* **2011**, *30*, 13. [CrossRef]
3. Shirasu, M.; Touhara, K. The scent of disease: Volatile organic compounds of the human body related to disease and disorder. *J. Biochem.* **2011**, *150*, 257–266. [CrossRef]
4. Honig, P.J.; Frieden, I.J.; Kim, H.J.; Yan, A.C. Streptococcal intertrigo: An underrecognized condition in children. *Pediatrics* **2003**, *112*, 1427–1429. [CrossRef]
5. Liddell, K. Smell as a diagnostic marker. *Postgrad. Med. J.* **1976**, *52*, 136–138. [CrossRef] [PubMed]
6. Stitt, W.Z.; Goldsmith, A. Scratch and sniff. The dynamic duo. *Arch. Dermatol.* **1995**, *131*, 997–999. [CrossRef]
7. Abaffy, T.; Duncan, R.; Riemer, D.D.; Tietje, O.; Elgart, G.; Milikowski, C.; DeFazio, R.A. Differential Volatile Signatures from Skin, Naevi and Melanoma: A Novel Approach to Detect a Pathological Process. *PLoS ONE* **2010**, *5*, e13813. [CrossRef]
8. Abaffy, T.; Möller, M.G.; Riemer, D.D.; Milikowski, C.; De Fazio, R.A. Comparative analysis of volatile metabolomics signals from melanoma and benign skin: A pilot study. *Metabolomics* **2013**, *9*, 998–1008. [CrossRef]
9. Kwak, J.; Gallagher, M.; Ozdener, M.H.; Wysocki, C.J.; Goldsmith, B.R.; Isamah, A.; Faranda, A.; Fakharzadeh, S.S.; Herlyn, M.; Johnson, A.T.C.; et al. Volatile biomarkers from human melanoma cells. *J. Chromatogr. B* **2013**, *931*, 90–96. [CrossRef]
10. Trivedi, D.K.; Sinclair, E.; Xu, Y.; Sarkar, D.; Walton-Doyle, C.; Liscio, C.; Banks, P.; Milne, J.; Silverdale, M.; Kunath, T.; et al. Discovery of Volatile Biomarkers of Parkinson's Disease from Sebum. *ACS Cent. Sci.* **2019**, *5*, 599–606. [CrossRef]
11. Di Natale, C.; Paolesse, R.; D'Arcangelo, G.; Comandini, P.; Pennazza, G.; Martinelli, E.; Rullo, S.; Roscioni, M.C.; Roscioni, C.; Finazzi-Agrò, A.; et al. Identification of schizophrenic patients by examination of body odor using gas chromatography-mass spectrometry and a cross-selective gas sensor array. *Med. Sci. Monit.* **2005**, *11*, CR366–CR375. [PubMed]
12. Monedeiro, F.; Reis, R.B.D.; Peria, F.M.; Sares, C.T.G.; Martinis, B.S.D. Investigation of sweat VOC profiles in assessment of cancer biomarkers using HS-GC-MS. *J. Breath Res.* **2020**, *14*, 026009. [CrossRef] [PubMed]
13. Coronavirus: Chiffres Clés et Evolution de la COVID-19 en France et dans le Monde. Available online: https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolution-de-la-covid-19-en-france-et-dans-le-monde (accessed on 23 August 2024).
14. Rabaan, A.A.; Smajlović, S.; Tombuloglu, H.; Ćordić, S.; Hajdarević, A.; Kudić, N.; Al Mutai, A.; Turkistani, S.A.; Al-Ahmed, S.H.; Al-Zaki, N.A.; et al. SARS-CoV-2 infection and multi-organ system damage: A review. *Biomol. Biomed.* **2023**, *23*, 37–52. [CrossRef] [PubMed]
15. Ates, H.C.; Yetisen, A.K.; Güder, F.; Dincer, C. Wearable devices for the detection of COVID-19. *Nat. Electron.* **2021**, *4*, 13–14. [CrossRef]
16. Snitz, K.; Andelman-Gur, M.; Pinchover, L.; Weissgross, R.; Weissbrod, A.; Mishor, E.; Zoller, R.; Linetsky, V.; Medhanie, A.; Shushan, S.; et al. Proof of concept for real-time detection of SARS CoV-2 infection with an electronic nose. *PLoS ONE* **2021**, *16*, e0252121. [CrossRef]
17. Shekhawat, J.K.; Banerjee, M. Role of Breath Biopsy in COVID-19. *J. Appl. Lab. Med.* **2022**, *7*, 1175–1188. [CrossRef]
18. Subali, A.D.; Wiyono, L.; Yusuf, M.; Zaky, M.F.A. The potential of volatile organic compounds-based breath analysis for COVID-19 screening: A systematic review & meta-analysis. *Diagn. Microbiol. Infect. Dis.* **2022**, *102*, 115589. [CrossRef]
19. Grandjean, D.; Sarkis, R.; Lecoq-Julien, C.; Benard, A.; Roger, V.; Levesque, E.; Bernes-Luciani, E.; Maestracci, B.; Morvan, P.; Gully, E.; et al. Can the detection dog alert on COVID-19 positive persons by sniffing axillary sweat samples? A proof-of-concept study. *PLoS ONE* **2020**, *15*, e0243122. [CrossRef]
20. Sakr, R.; Ghsoub, C.; Rbeiz, C.; Lattouf, V.; Riachy, R.; Haddad, C.; Zoghbi, M. COVID-19 detection by dogs: From physiology to field application—A review article. *Postgrad. Med. J.* **2021**, *98*, 212–218. [CrossRef]
21. Devillier, P.; Gallet, C.; Salvator, H.; Lecoq-Julien, C.; Naline, E.; Roisse, D.; Levert, C.; Breton, E.; Galtat, A.; Decourtray, S.; et al. Biomedical detection dogs for the identification of SARS-CoV-2 infections from axillary sweat and breath samples. *J. Breath Res.* **2022**, *16*, 037101. [CrossRef]
22. Guest, C.; Dewhirst, S.Y.; Lindsay, S.W.; Allen, D.J.; Aziz, S.; Baerenbold, O.; Bradley, J.; Chabildas, U.; Chen-Hussey, V.; Clifford, S.; et al. Using trained dogs and organic semi-conducting sensors to identify asymptomatic and mild SARS-CoV-2 infections: An observational study. *J. Travel Med.* **2022**, *29*, taac043. [CrossRef] [PubMed]
23. Tungkijanansin, N.; Phusrisom, S.; Chatdarong, K.; Torvorapanit, P.; Sirinara, P.; Nhujak, T.; Kulsing, C. Gas chromatography-flame ionization detector for sweat based COVID-19 screening. *Anal. Chim. Acta* **2023**, *1280*, 341878. [CrossRef] [PubMed]
24. Fathizadeh, H.; Taghizadeh, S.; Safari, R.; Khiabani, S.S.; Babak, B.; Hamzavi, F.; Ganbarov, K.; Esposito, S.; Zeinalzadeh, E.; Dao, S.; et al. Study presence of COVID-19 (SARS-CoV-2) in the sweat of patients infected with Covid-19. *Microb. Pathog.* **2020**, *149*, 104556. [CrossRef] [PubMed]
25. Thaveesangsakulthai, I.; Jongkhumkrong, J.; Chatdarong, K.; Torvorapanit, P.; Sukbangnop, W.; Sooksimuang, T.; Kulsing, C.; Tomapatanaget, B. A fluorescence-based sweat test sensor in a proof-of-concept clinical study for COVID-19 screening diagnosis. *Analyst* **2023**, *148*, 2956–2964. [CrossRef] [PubMed]
26. Tungkijanansin, N.; Giebelhaus, R.T.; Schmidt, S.A.; Nhujak, T.; Chatdarong, K.; Torvorapanit, P.; Harynuk, J.J.; Kulsing, C. Identification of coronavirus disease marker compounds in sweat with comprehensive two dimensional gas chromatography using multiloop splitter-based non-cryogenic artificial trapping modulation system. *J. Chromatogr. Open* **2024**, *5*, 100113. [CrossRef]

27. Chuachaina, S.; Thaveesangsakulthai, I.; Sinsukudomchai, P.; Somboon, P.; Traipattanakul, J.; Torvorapanit, P.; Chatdarong, K.; Kulsing, C.; Nhujak, T. Identification of Volatile Markers in Sweat for COVID-19 Screening by Gas Chromatography-Mass Spectrometry. *ChemistrySelect* **2024**, *9*, e202304388. [CrossRef]
28. Phusrisom, S.; Tungkijananansin, N.; Sirinara, P.; Kulsing, C. Using Gas Chromatography-Ion Mobility Spectrometry to Investigate Volatile Compound Profiles in Human Sweat during COVID-19 Vaccination. *ChemistrySelect* **2024**, *9*, e202303353. [CrossRef]
29. Hilaire, F.; Basset, E.; Bayard, R.; Gallardo, M.; Thiebaut, D.; Vial, J. Comprehensive two-dimensional gas chromatography for biogas and biomethane analysis. *J. Chromatogr. A* **2017**, *1524*, 222–232. [CrossRef]
30. Franchina, F.A.; Zanella, D.; Dubois, L.M.; Focant, J. The role of sample preparation in multidimensional gas chromatographic separations for non-targeted analysis with the focus on recent biomedical, food, and plant applications. *J. Sep. Sci.* **2021**, *44*, 188–210. [CrossRef]
31. Boudard, E. Etude des conditions de prélèvement du volatolome corporel cutané pour son analyse discriminante par chromatographie en phase gazeuse intégralement bidimensionnelle en vue de l'identification de biomarqueurs. 2024. Université Paris Sciences et Lettres.
32. Parsons, B.A.; Marney, L.C.; Siegler, W.C.; Hoggard, J.C.; Wright, B.W.; Synovec, R.E. Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach. *Anal. Chem.* **2015**, *87*, 3812–3819. [CrossRef]
33. Schöneich, S.; Ochoa, G.S.; Monzón, C.M.; Synovec, R.E. Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome. *J. Chromatogr. A* **2022**, *1667*, 462868. [CrossRef] [PubMed]
34. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
35. Hart, A. Mann-Whitney test is not just a test of medians: Differences in spread can be important. *BMJ* **2001**, *323*, 391–393. [CrossRef]
36. Rankin-Turner, S.; McMeniman, C.J. A headspace collection chamber for whole body volatilomics. *Analyst* **2022**, *147*, 5210–5222. [CrossRef]
37. Haertl, T.; Owsienko, D.; Schwinn, L.; Hirsch, C.; Eskofier, B.M.; Lang, R.; Wirtz, S.; Loos, H.M. Exploring the interrelationship between the skin microbiome and skin volatiles: A pilot study. *Front. Ecol. Evol.* **2023**, *11*, 1107463. [CrossRef]
38. Penn, D.J.; Oberzaucher, E.; Grammer, K.; Fischer, G.; Soini, H.A.; Wiesler, D.; Novotny, M.V.; Dixon, S.J.; Xu, Y.; Brereton, R.G. Individual and gender fingerprints in human body odour. *J. R. Soc. Interface* **2007**, *4*, 331–340. [CrossRef]
39. ScenTree—La Nouvelle Classification Collaborative des Ingrédients de la Parfumerie. Available online: https://www.scentree.co/fr/map/G%C3%A9n%C3%A9ral (accessed on 7 August 2024).
40. INCI Beauty—Analysez la Composition de Vos Cosmétiques. Available online: https://incibeauty.com/ (accessed on 7 August 2024).
41. The Good Scents Company Information System. Available online: https://www.thegoodscentscompany.com/ (accessed on 7 August 2024).
42. Lucchi, G.; Crépin, M.; Chambaron, S.; Peltier, C.; Gilbert, L.; Guéré, C.; Vié, K. Effects of psychological stress on the emission of volatile organic compounds from the skin. *Sci. Rep.* **2024**, *14*, 7238. [CrossRef] [PubMed]
43. de Lacy Costello, B.; Amann, A.; Al-Kateb, H.; Flynn, C.; Filipiak, W.; Khalid, T.; Osborne, D.; Ratcliffe, N.M. A review of the volatiles from the healthy human body. *J. Breath Res.* **2014**, *8*, 014001. [CrossRef] [PubMed]
44. Bessonneau, V.; Thomas, O. VOC Contamination in Hospital, from Stationary Sampling of a Large Panel of Compounds, in View of Healthcare Workers and Patients Exposure Assessment. *PLoS ONE* **2013**, *8*, e55535. [CrossRef]
45. LeBouf, R.F.; Virji, M.A.; Saito, R.; Henneberger, P.K.; Simcox, N.; Stefaniak, A.B. Exposure to volatile organic compounds in healthcare settings. *Occup. Environ. Med.* **2014**, *71*, 642–650. [CrossRef]
46. Samon, S.M.; Hammel, S.C.; Stapleton, H.M.; Anderson, K.A. Silicone wristbands as personal passive sampling devices: Current knowledge, recommendations for use, and future directions. *Environ. Int.* **2022**, *169*, 107339. [CrossRef] [PubMed]
47. Delgado-Povedano, M.M.; Calderón-Santiago, M.; Priego-Capote, F.; de Castro, M.D.L. Development of a method for enhancing metabolomics coverage of human sweat by gas chromatography–mass spectrometry in high resolution mode. *Anal. Chim. Acta* **2016**, *905*, 115–125. [CrossRef] [PubMed]