Clinica Chimica Acta 526 (2022) 6–13

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Breath analysis using electronic nose and gas chromatography-mass spectrometry: A pilot study on bronchial infections in bronchiectasis

Luciana Fontes de Oliveira [a], Celia Mallafré-Muro [a,b], Jordi Giner [c], Lidia Perea [d], Oriol Sibila [d], Antonio Pardo [b], Santiago Marco [a,b,*]

[a] Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028, Barcelona, Spain
[b] Department of Electronics and Biomedical Engineering, University of Barcelona, Marti I Franqués 1, 08028 Barcelona, Spain
[c] Department of Pneumology and Allergy. Hospital de la Sta. Creu I Sant Pau. Barcelona, Spain
[d] Respiratory Department, Hospital Clinic, IDIBAPS, Barcelona, Spain

ARTICLE INFO

Keywords:
Breath analysis
Bronchiectasis
Signal processing
E-nose
GC-MS

ABSTRACT

Background and aims: In this work, breath samples from clinically stable bronchiectasis patients with and without bronchial infections by *Pseudomonas Aeruginosa- PA*) were collected and chemically analysed to determine if they have clinical value in the monitoring of these patients.
Materials and methods: A cohort was recruited inviting bronchiectasis patients (25) and controls (9). Among the former group, 12 members were suffering PA infection. Breath samples were collected in Tedlar bags and analyzed by e-nose and Gas Chromatography-Mass Spectrometry (GC-MS). The obtained data were analyzed by chemometric methods to determine their discriminant power in regards to their health condition. Results were evaluated with blind samples.
Results: Breath analysis by electronic nose successfully separated the three groups with an overall classification rate of 84% for the three-class classification problem. The best discrimination was obtained between control and bronchiectasis with PA infection samples 100% ($CI_{95\%}$: 84–100%) on external validation and the results were confirmed by permutation tests. The discrimination analysis by GC-MS provided good results but did not reach proper statistical significance after a permutation test.
Conclusions: Breath sample analysis by electronic nose followed by proper predictive models successfully differentiated between control, Bronchiectasis and Bronchiectasis PA samples.

1. Introduction

The potential advantages of breath analysis for volatolomics studies, including the unlimited sample supply, the non-invasive way to collect samples, and the possibility deliver fast analysis results have been described in several previous works [1–7]. However, despite the abundant literature, the use of breath analysis for clinical applications is in its infancy [8,9] and the lack of standardization on sample collection/analysis [10–13] and the complexity of data analysis step leave space to further developments [14–18].

In breath analysis, there are different types of confounding factors and the most important are the clinical (gender, age, diet, medication) and instrumental ones (time of the measurements, time from collection until analysis). A good design is essential to handle confounding factors and methods as randomization, restriction, or matching [19] can be used. Appropriate control of the clinical and instrumental confounding factors on observational studies in breath analysis could improve and decrease biased results [20–22].

On the other hand, there are a variety of instrumental techniques for breath analysis that differ on usability, cost, and retrieved chemical information: namely, GCxGC-MS [23], chemical sensor systems [24], Proton Transfer Reaction-Mass Spectrometry (PTR-MS) [25], Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS) [26], Laser Spectroscopy [27] or Gas Chromatography -Ion Mobility Spectrometry (GC-IMS) [28].

Among the several analytical platforms available to analyse breath samples a review on cancer detection mentions that GC-MS and e-nose are the most commonly used platforms (47% and 26% of papers), while their simultaneous use on the same samples appears only in 8% of the

* Corresponding author.
E-mail addresses: smarco@ibecbarcelona.eu, santiago.marco@ub.edu (S. Marco).

https://doi.org/10.1016/j.cca.2021.12.019
Received 7 July 2021; Received in revised form 20 December 2021; Accepted 20 December 2021
Available online 23 December 2021

studies [29]. The use of two (or more) analytical distinct platforms [30,31] on the same samples or confront the results from breath analysis with another type of sample (as tissue/ sputum) [32–34] are an interesting avenue of research in breath analysis and can be used to confirm somehow the obtained results.

As said the raw data provided by different analytical platforms differ in information content but also on dimensionality and data processing needs. E-noses and GC-MS datasets, particularly, are represented as a vector and a matrix, for one sample, and as a matrix and a three-way array for the whole dataset, respectively. In general terms, GC-MS provides higher dimensional and richer information data, at the expense of requiring a more complex data processing pipeline [35].

Proper validation methodologies help to avoid overfitting and consequently reduce false discoveries [36–38]. Overfitting problems in GC/MS are aggravated by the curse of dimensionality [39], because these datasets are highly dimensional (10^2–10^3 detected analytes) and most studies have limited sample size (20–100 subjects). In this context, we advocate the use of resampling methods for model optimization and external validation for performance assessment [40,41]. We have to remind that for small datasets, all samples can be used for external validation using double cross-validation methodologies [42]. Even if external validation is unbiased, small sample datasets do still feature performance estimators with large variance and a permutation test should be used to confirm the obtained results [43].

In the last decade, a number of VOCs in breath have been found to be helpful in the diagnostics of several diseases including respiratory diseases and cancer [5,44–46] Among the several diseases that can be evaluated by breath analysis the development of bronchial infections on bronchiectasis patients has been described and evaluated by previous work. In this case, electronic nose has been used to identify airway bacterial colonization in Chronic Obstructive Pulmonary Disease (COPD) patients [47,48]. Furthermore, e-nose technology has been proven successful to identify *Pseudomonas aeruginosa* infection in bronchiectasis [49].

In the clinical stability phase the presence of potential pathogens bacteria in the airway of bronchiectasis patients are common (30–70%) being mainly *Pseudomonas aeruginosa* [50]. Furthermore, aggravations as faster lung function loss, pulmonary and systemic inflammation are serious concerns. Bronchial infection is the reason for 60–70% of these aggravations [51], that has a direct relation with mortality increase in Bronchiectasis [52] and the reasons why bronchiectasis patients are more susceptible to developing a bronchial infection are still unknown.

In this work, we build upon previous works to test the adequacy of breath sampling to monitor infections in bronchiectasis patients, particularly with *Pseudomonas aeruginosa*. Previous studies have reported success on e-nose applications using linear discriminant analysis and leave-one-out internal validation. In this new study, we amplified the finds collecting a new group of samples and using external validation methodologies and permutations test. Additionally, the same samples collected were also analyzed by GC-MS aiming to understand, if possible, the origin of the chemical discrimination already proven by e-noses and discuss the advantages and disadvantages of these distinct analytical platforms on breath analysis. In fact, prior studies using GC-MS have found VOCs related to the presence of *Pseudomonas Aeruginosa (PA)* in cystic fibrosis patients: methyl thiocyanate [53] and 2-aminoacetophenone [54] have been reported as putative biomarkers. In this work, a volatolomics untargeted approach is proposed to discover potential signatures of PA infection in Bronchiectasis patients.

2. Methods

A schematic representation of the applied methodology can be observed in Fig. 1. This research features parallel analysis of breath samples by electronic nose and GC/MS. The following sections provide methodological information.

2.1. Cohort selection and experimental design

Observational studies are always suspect of bias. In order to block potential confounding factors, we carried out a proper experimental design. To prevent gender as a confounding factor a restriction strategy was applied. It is known that in non-cystic fibrosis bronchiectasis has more prevalence among females than among males. Additionally, females suffer more severe diseases and with worse prognoses in terms of poorer lung function and survival [55]. For these reasons, only females were included in the cohort.

Breath samples from woman subjects were collected, all of them were not currently smokers and the ones with prescribed drugs therapies were asked to stop medication 1 day before sample collection and food-drink intake at least 12 h before. All patients signed the informed consent form to participate in the study (Ethical approvement code: Institut d'Investigació Biomédica Sant Pau- IIBSP-PRI-2018-105). Diagnosis of Bronchiectasis was performed according to current European guidelines [56]. Bronchial infection was determined using a quantitative sputum culture prior to breath samples collection. PA infection was diagnosed using sputum culture that was not performed in healthy controls because PA only affects patients with pulmonary diseases who had chronic sputum production [50]. The procedure used to diagnose PA was well validated and previously described [57], besides that no other pathogens differently of PA were detected. Bacterial colonization was considered when patients had PA infection and clinical stability, defined by the absence of increased symptoms that required changes in baseline treatment during 4 weeks [57].

To block instrumental shifts often found in e-noses [37] and even in

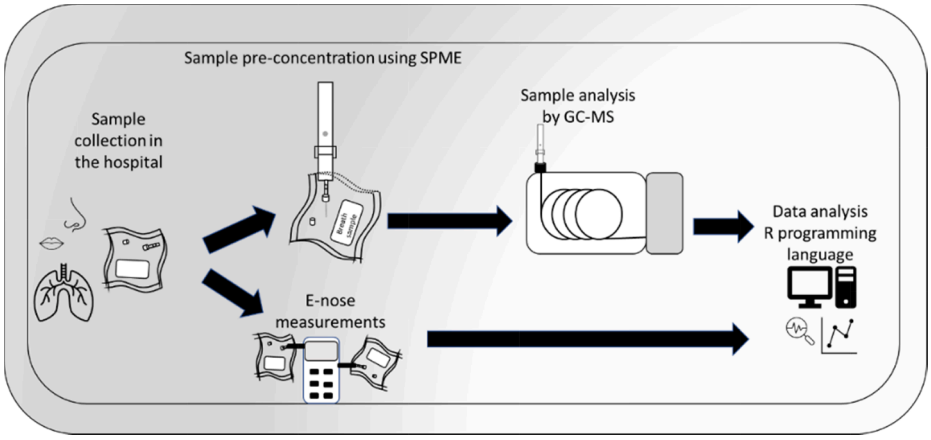

Fig. 1. Overview of sample collection and analysis.

GC-MS, we matched the collection and analysis of samples belonging to the different groups as depicted in Fig. 2. Supplementary Material 1 shows specific information for the subjects in the study.

2.2. Breath sample collection

Three liters Tedlar® bags were used to collect the total amount of exhaled air by the patients. Two-valve Tedlar® bags were cleaned before use by flushing with argon and baking at 45 °C during 15 min (repeated three times) [58]. All samples were collected in the same room after the patients breathing through a Hans-Rudolph valve during 3 min as described in previous work [49]. Also, a biological filter was used for each patient to avoid pathogens entering the bags and cross contamination between patients.

Samples were collected in the hospital and e-nose measurements were done a few minutes after the patients filled the bags. Then, the bags were carried out to the laboratory and analyzed on the same day of sample collection by GC-MS. For each day of sample collection, ambient air, controls, and cases samples were collected. The final sample set analyzed consisted of 8 ambient air samples, 9 **Controls** (healthy women), 13 bronchiectasis patients (**Bronch),** and 12 bronchiectasis subjects with bronchial infection by *Pseudomonas aeruginosa* (**Bro_PA).**

2.3. Sample analysis

2.3.1. E-nose

The e-nose device Cyranose 320® (Smith Detections, Pasadena, CA), that features a nanocomposite sensor array with 32 sensors, was connected to the breath Tedlar bag for 5 min and each measurement consisted of 5 replicates. Nitrogen was used as carrier gas and a constant flow rate of 120 mL/min was used during 60 s and 40 s for baseline recording and sample analysis, respectively, followed by an increase of the flow for 180 mL/min for sample line purging and air intake. A Tedlar bag with ambient air collected in the day of sample analysis was analysed in parallel every day as background measurement. Supplementary Materials shows actual pictures of the breath sampling process (Figure 2 in supplementary materials.)

2.3.2. Gas chromatography

Solid phase micro extraction (SPME) sample preconcentration was carried out using a 75 μm carboxen®/ Polydimethylsiloxane (CAR/PDMS) fiber [59]. The fiber was exposed inside the bags for 30 min at ambient temperature and immediately after it was desorbed into the GC injector. The chromatographic column used was type DB-624 (60 m × 0.320 mmID × 1.8μm – Agilent). The temperature of the column was maintained at 40 °C for 2 min and then subjected to a temperature ramp of 10 °C/min till 250 °C and stayed at this temperature for 5 additional minutes. The carrier gas used was helium in a constant flow of 1.7 mL min^{-1}. The temperatures of the injector and the transfer line were set to 250 °C and 230 °C, respectively. Ion source temperature was set to 200 °C and the mass scan range was from 40 to 400 *m/z*.

2.4. Data analysis

2.4.1. E-nose

A non-linear transformation (arctangent transformation) was used to improve data gaussianity [60]. Data normality was then confirmed with the Shapiro-Wilk test at the 5% risk with Benjamini-Hochberg multitest correction [61]. Variance of inter-replicates for each sample and robust PCA [62] was used for outlier detection. Specifically, the algorithm ROBPCA (available in the *rospca* package in R) was used and outliers were selected based on the robust score distance and the robust orthogonal distance. Proper cutoff values for those statistics are given by Hubert et al. [63]. After outlier removal, data was autoscaled and inspected by classical PCA.

Subject classification was based on the K-NN algorithm (available in the *class* package in R [64]) plus a majority voting over the replicate measurements. K-NN classifier optimization and performance assessment were based on double cross-validation [42] using leave one subject out (LOSO). By LOSO we mean that all the replicate measurements from the same subject are treated as a single indicator to decide the final label given to the subject. In the inner loop (internal validation) the number of neighbors was optimized, while performance assessment was carried out in the external loop (external validation). In both cases, the chosen figure of merit was classification accuracy (classification rate: CR). Final class assignment to each subject was based on the joint classification of all the replicates through a voting mechanism. To check that the obtained value cannot be obtained by random choice (null hypothesis) we calculated a permutations test [65] with 500 iterations. Supplementary Material 3 shows a block diagram of the e-nose data analysis (Figure 3 in supplementary materials.)

2.4.2. Gas chromatography

Features from the raw chromatograms were extracted using the XCMS package in R [66,67]. On XCMS matched filter algorithm was used for peak detection followed by peak clustering and alignment. Data

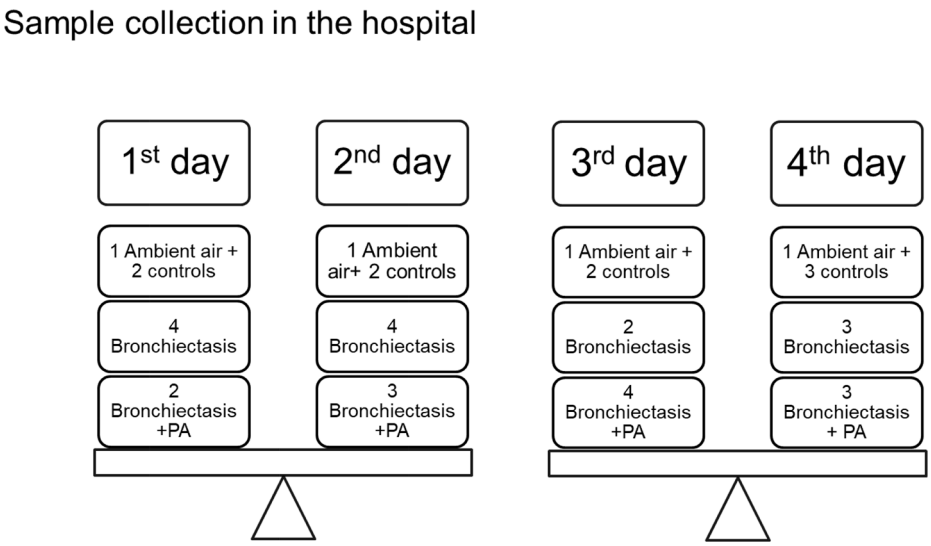

Fig. 2. Distribution of cases, controls, and quality controls samples between the days of collection.

imputation was used to fill missing values based on the integration of the peak position. Robust Principal component analysis (RobPCA) was used to explore the data and verify the presence of outliers in the same manner described above.

The extracted features were then corrected using log transformation and PQN normalization [68]. The AlpsNMR package [69] was used to create Partial Least Squares – Discriminant Analysis (PLS-DA) [70] classification models followed by permutations test. Furthermore, a second strategy was used applying variable selection based on Wilcoxon and binary problems were built using the same strategy applied to the e-nose data (double cross-validation [42] using leave one subject out (LOSO)), and PLS-DA and KNN models were built. All data analysis for e-nose and GC-MS was done in RStudio 4.0.3.

3. Results

3.1. E-nose analysis

E-nose analysis aim is to record a breath-print of a wide range of gases and vapours (mixture of compounds) present in each breath sample not focusing on a specific compound or class of compounds, in the case of Cyranose 320® this is done by 32 nanocomposite sensor arrays. Each patient breath measurement is represented as a matrix, having 32 columns, one per sensor, and as many rows as replicates have been measured (typically five). Supplementary Material 4 shows a heatmap where on rows is showed the different replicates of each patient and in the columns the sensors. The colors correspond to the value of the sensor's response after preprocessing. The rows are arranged according to the class of the patient and the columns are ordered according to hierarchical clustering of the sensor responses. After the non-linear transformation, all sensors except numbers 6 and 28 were normally distributed and we could not reject the null hypothesis of a normal distribution using the Shapiro-Wilk hypothesis test with Benjamini-Hochberg multitest correction. Three entire samples (including all replicates) were considered outliers (see methods) and were removed from the dataset before the construction of the models. After preprocessing, data were visually inspected by PCA and the score plot (PC1xPC2) is shown in Fig. 3. No clear data separation was observed at this point.

The best number of neighbors k (optimized in the internal validation loop), the values of classification rates (in external validation), and the p-value after permutation tests for all constructed models are showed in Table 1 (including three class and two class models). Furthermore, Supplementary Material 5 shows the confusion matrix for the three class K-NN models.

3.2. Gas chromatographic analysis

The application of XCMS to the raw data provided a feature table

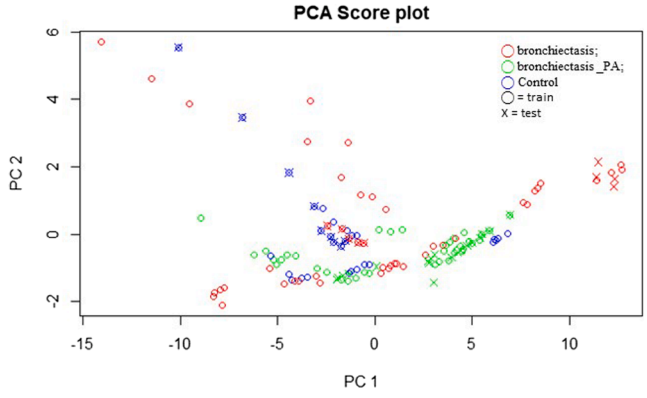

Fig. 3. Score plot for the Principal Component Analysis (PC1xPC2) containing all samples from the e-nose measurements.

Table 1
Summary of the KNN models performance in external validation for the e-nose dataset (confidence limits 95% in brackets, calculated according to the binomial distribution).

		All replicates			
Models	**best k**	**Sensitivity**	**Specificity**	**CR(%)**	**p-value**
Control vs Bronch vs Bro_PA	7	–	–	78	0.002
Control vs Bronch	7	0.86 (0.73,0.92)	0.9 (0.76,0.96)	89(84,96)	0.002
Control vs Bro_PA	5	0.94 (0.82,0.97)	0.9 (0.76,0.96)	92(83,95)	0.004
Bronch vs Bro_PA	5	0.86 (0.71,0.92)	0.92 (0.83,0.97)	89(81,93)	0.002
	Majority vote				
Models	**best k**	**Sensitivity**	**Specificity**	**CR (%)**	**p-value**
Control vs Bronch vs Bro_PA	7	–	–	84	0.002
Control vs Bronch	7	0.92 (0.64,1.00)	1(0.66,1.00)	95 (77,100)	0.002
Control vs Bro_PA	5	1(0.75,1.00)	1(0.66,1.00)	100 (84,100)	0.004
Bronch vs Bro_PA	5	0.75 (0.43–0.95)	1(0.74,1.00)	87(69,97)	0.002

with dimensions 42 samples × 409 features. The outlier detection step did not show any anomalous sample and all subjects were kept in the data set. The Total Ion Chromatograms (TICs) in log scale for the GC-MS analysis for all collected samples and the outlier detection step plot are shown on Supplementary Material 6(i) and 6(ii).

Fig. 4 (i) shows the score plots of a PLS-DA model for a binary classification problem. However, it is well known that scoreplots are overoptimistic. In this spirit, we gave better credit to the evaluation of the classifiers in external validation. The best results were obtained for the discrimination between Control versus Bronchiectasis_PA. PLS-DA models presented good classification rates above 0.75 on external validation. However, permutation tests were applied for all binary PLS-DA models, and it was not possible to reject the null hypothesis (see Fig. 4 (ii)).

Table 2 shows a summary of the obtained results for the GC-MS after feature selection applying Wilcoxon test (binary models) and then applying the same strategy used to the e-nose (double cross-validation using leave one subject out).

The last step on an untargeted approach is the compound identification and although several important compounds were already identified and described as potential to be related with specific diseases on breath samples [71,72] the untargeted methodology used here was not able to reach the annotation step, in other words, even though some PLS-DA presented good results neither model was able to overcome permutations tests for statistical significance.

4. Discussion

4.1. Sample collection and analysis

Breath samples can be collected and analyzed online and offline, the main reason for choosing one or another method will depend on the final aim of the work. Although, analyzing breath directly and the use of cartridges are preferential for e-nose and GC-MS, respectively, Tedlar® bags fits very well when the objective is to analyze the same sample with two or more analytical platform [58]. Furthermore, when a patient has a pulmonary disability, identifying and collecting the end-tidal breath it is not a simple task and for this reason and, to follow the same protocol for all involved subjects in the study, whole breath samples were collected

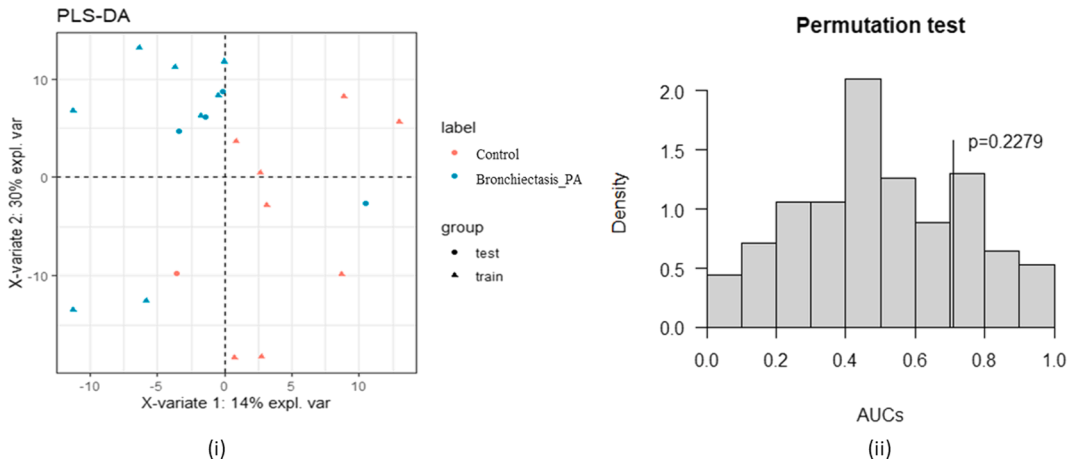

Fig. 4. (i) Scoreplot PLS-DA model obtained from alpsNMR using as class Control versus Bronchiectasis_PA (ii) permutation test for the predictive model.

Table 2
Summary of results obtained for the GC-MS after feature selection applying Wilcoxon test (binary models).

	Number of selected features	PLS-DA CR (%)	KNN CR (%)
Control vs Bronch	39	40	62
Control vs Bro_PA	42	62	52
Bronch vs Bro_PA	13	48	58

and the methodology was previous validated [49].

Furthermore, the CAR/PDMS fiber and the column selected possess intrinsic characteristics that will allow the pre-concentration and analysis of a delineated group of compounds presented in breath samples analyzed by GC-MS. However, the preconcentration step is mandatory since many compounds will be present in very low concentrations inside the bags mainly when whole breath samples are collected as in this work. Breath sampling methodologies, advantages, and disadvantages are described in the literature [73].

4.2. E-nose analysis

Initial visual inspection of the sensor response distribution indicated a strong lack of normality. The histogram presented long right-side tails but also the presence of negative values. This is known to have a negative effect on data analysis techniques based on the analysis of variance. Data normality was greatly improved using a non-linear transformation. To be able to deal with negative values we selected the arctangent transformation instead of the most common logarithmic transform.

It is possible to observe that samples do not appear linearly separable in an unsupervised exploration based on the PCA scoreplot (Fig. 3). K-NN and PLS-DA classifiers were evaluated. PLS-DA is one of the most common classifiers in metabolomics, but it provides only linear partitions of the input space. To test a more flexible input space partition K-NN models were chosen as a simple an model-free alternative. All K-NN models constructed for the e-nose dataset presented better results than the PLS-DA ones and for this reason, only the KNN results are shown.

The classification rates were calculated using two different approaches. First, considering each individual replicates and second using the most voted class using all replicates from the same individual. This last procedure does not represent in fact additional costs since all the replicates are just consecutive analyses from the same bag as explained in Section 2.4.1. Using all the replicates approach significantly improved the performance of the classifier, except in the case of the three-class problem.

When trying to classify individual measurements, K-NN models presented very good classification rates on external validation, and the values varied between 78% and 92%. Permutation tests were used for all models and in all cases, the classification rates were considered statistically significant (risk level 0.05) compared with the distribution of the null hypothesis. The three-class problem resulted in a smaller CR (78%) but still statistically significant. For all the other binary problems the CR ranged between 89 and 92%, but those differences were not statistically significant due to the limited cardinality of the different groups.

Results improved significantly when we used the majority vote mechanism to classify a subject using the five consecutive replicates. In this case, the CR for the three-class problem improved up to 84%, while we got perfect classification (100)% for the Control vs Bronchiectasis with PA infection. The next model in terms of good performance was the discrimination between Controls and Bronchiectasis, while the presence of PA infection in Bronchiectasis performed a bit lower but still with an excellent 87% classification rate. In general, models presented better specificity than sensitivity, however, the latter still ranged from 92% to 100% (see Table 1).

The current study indicates that the e-nose was able to classify the breath samples not only in internal validation as previously described but also in external validation. Furthermore, the class separation is not linear requiring non-linear decision functions to obtain good results. While these results are encouraging, they should be further validated with more subjects (due to the risk of over adjustment related with the small sample conditions), during a longer study, and eventually in a multicenter study. It should be independently tested with additional e-nose units. Another direction of study is to investigate if this very good separation is specific to the sensing technology used with the presently used device or if they can be replicated with electronic noses of different technologies.

4.3. Gas chromatographic analysis

While the score plot shows a good separation between the two studied classes, we have to take into account that PLS-DA score plots are easily overoptimistic [36,74]. Additionally, the apparent good result in classification rate is unable to overcome the additional permutation test due to the large variance of the CR estimator probably linked to the small number of samples compared to the input data dimensionality.

Regarding the second strategy applied (Table 2), these results agree with the obtained results for the e-nose in the sense that KNN models presented a better performance than PLS-DA (exception for the Control vs Bro_PA). However, in this work, all the predictive models constructed for the GC-MS data the classification rates were not good enough to distinguish between the classes, and consequently it was not possible to

discover the compounds that are important to class separation.

It is interesting to confront the successful results of the predictive models built with the e-nose measurements in opposition to the failure obtained using GC-MS data. We can point several underlying reasons behind these results. First, the e-nose measurements have replicates (5 per sample) while a single GC/MS analysis is carried out per bag. Secondly and as expected, the dimensionality of the e-nose is much smaller than the GC-MS leading to curse of dimensionality problems. This is more important when facing binary problems (for the GC-MS) since the sample count is even smaller.

Furthermore, the signal processing pipeline for GC-MS is more complex than for e-nose data. The large number of peaks, sometimes with strong coelution, baseline instabilities, and slight shifts in retention time leading to alignment problems, makes the whole data processing workflow a real challenge, particularly if in addition we have a limited supply of examples for the machine learning step. This is in agreement with previous research that combined GC-MS and e-nose analysis on the same samples for cancer screening [75,76]. The GC-MS results obtained in this study sign that, even though the use of experimental design and good analytical chemistry practices are essential, good validations techniques in the development of the models are key to avoiding false discoveries in complex data.

5. Conclusions

This study showed that e-noses were able to differentiate bronchiectasis and bronchiectasis with bronchial infections, produced by *pseudomonas aeruginosa,* patients from controls with good results in external validation and the results were confirmed by permutation tests.

We would like to highlight a number of methodological factors that support the results and the validity of the conclusions. First, the proper experimental design to block the most important confounding factors. Second, the evaluation of the predictive models in external validation using double leave one subject out and the additional permutation tests to explore if the obtained results can just be obtained due to the large variance of performance estimators in small sample conditions. Results for e-nose improved significantly after non-linear signal transformation, and the use of majority voting over measurement replicates.

The use of GC-MS to explore the important compounds for the class differentiation was not successful. The main reasons for that were the small sample counting, the lack of replicates and the complexity of the obtained signals. We consider that more strict validation methodologies should be in use to avoid false discoveries in breath analysis.

Despite the good results obtained by electronic nose, the fact that this approach does not allow to identify condition specific compounds is a clear limitation of this approach since it does not bring additional information for the understanding of the underlying mechanisms below the observed discrimination.

The obtained results should be considered as a positive indication supporting the validity of the proposed methodology. However, studies with larger cohorts, from different geographical areas and recruitment hospitals are needed to give additional support to the findings reported in this work.

CRediT authorship contribution statement

Luciana Fontes de Oliveira: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Celia Mallafré-Muro:** Software, Data curation, Writing – review & editing, Visualization. **Jordi Giner:** Validation, Investigation, Resources, Writing – review & editing. **Lidia Perea:** Validation, Investigation, Writing – review & editing. **Oriol Sibila:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Antonio Pardo:** Formal analysis, Resources, Writing – review & editing, Supervision. **Santiago Marco:** Conceptualization, Methodology, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is part of the BEST Postdoctoral Program, funded by the European Commission under Horizon 2020 Marie Skłodowska-Curie Actions COFUND scheme (Grant Agreement no. 712754) and by the Severo Ochoa program of the Spanish Ministry of Science and Competitiveness (Grant SEV-2014-0425 (2015-2019)). We would like to acknowledge, the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (expedient 2017 SGR 1721); the Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya; and the European Social Fund (ESF). Additional financial support has been provided by the Institut de Bioenginyeria de Catalunya (IBEC), Spain. IBEC is a member of the CERCA Programme/Generalitat de Catalunya, Spain. This work has been additionally funded by Spanish MINECO Project TENSOMICS (RTI2018-098577-B-C22). This work was supported also by Sociedad Española de Neumología y Cirugía Torácica (SEPAR), Societat Catalana de Pneumologia (SOCAP), Fundació Catalana de Pneumologia (FUCAP) and Instituto de SaludCarlos III - Fondos FEDER (PI18/00311). Finally, we want to acknowledge Prof. Francisco Javier Santos, from the Department of Chemical Engineering and Analytical Chemistry at the University of Barcelona, for his help in regards the GC-MS analysis of breath samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2021.12.019.

References

[1] S. Das, M. Pal, Review—non-invasive monitoring of human health by exhaled breath analysis: a comprehensive review, J. Electrochem. Soc. 167 (3) (2020) 037562, https://doi.org/10.1149/1945-7111/ab67a6.
[2] M.D. Davis, S.J. Fowler, A.J. Montpetit, Exhaled breath testing – a tool for the clinician and researcher, Paediatr. Respir. Rev. 29 (2019) 37–41, https://doi.org/10.1016/j.prrv.2018.05.002.
[3] J.D. Beauchamp, C.E. Davis, J. Pleil (Eds.), Breathborne Biomarkers and the Human Volatilome, Elsevier, Amsterdam, 2020, https://doi.org/10.1016/c2018-0-04980-4.
[4] W. Miekisch, J.K. Schubert, G.F.E. Noeldge-Schomburg, Diagnostic potential of breath analysis—focus on volatile organic compounds, Clin. Chim. Acta 347 (1-2) (2004) 25–39.
[5] Y. Saalberg, M. Wolff, VOC breath biomarkers in lung cancer, Clin. Chim. Acta 459 (2016) 5–9, https://doi.org/10.1016/j.cca.2016.05.013.
[6] M. Rodríguez-Aguilar, L. Díaz de León-Martínez, P. Gorocica-Rosete, R.P. Padilla, I. Thirión-Romero, O. Ornelas-Rebolledo, R. Flores-Ramírez, Identification of breath-prints for the COPD detection associated with smoking and household air pollution by electronic nose, Respir. Med. 163 (2020) 105901, https://doi.org/10.1016/j.rmed.2020.105901.
[7] L. Díaz de León-Martínez, M. Rodríguez-Aguilar, P. Gorocica-Rosete, C. A. Domínguez-Reyes, V. Martínez-Bustos, J.A. Tenorio-Torres, O. Ornelas-Rebolledo, J.A. Cruz-Ramos, B. Balderas-Segura, R. Flores-Ramírez, Identification of profiles of volatile organic compounds in exhaled breath by means of an electronic nose as a proposal for a screening method for breast cancer: a case-control study, J. Breath Res. 14 (4) (2020) 046009, https://doi.org/10.1088/1752-7163/aba83f.
[8] A. Bikov, Z. Lázár, I. Horvath, Established methodological issues in electronic nose research: how far are we from using these instruments in clinical settings of breath analysis? J. Breath Res. 9 (3) (2015) 034001, https://doi.org/10.1088/1752-7155/9/3/034001.
[9] S.F. Solga, L.A. Spacek, T.H. Risby, Challenges in clinical breath research development, in: J.D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne

Biomarkers Hum. Volatilome, Elsevier B.V., Amsterdam, 2020, pp. 601–613, https://doi.org/10.1016/b978-0-12-819967-1.00036-0.
[10] A. Amann, W. Miekisch, J. Pleil, T. Risby, J. Schubert, Methodological issues of sample collection and analysis of exhaled breath, in: I. Horvath, J.C. de Jongste (Eds.), Exhaled Biomarkers, European Respiratory Society, 2010, pp. 96–114, https://doi.org/10.1183/1025448x.00018509.
[11] P. Sukul, P. Trefz, J.K. Schubert, W. Miekisch, Immediate effects of breath holding maneuvers onto composition of exhaled breath, J. Breath Res. 8 (3) (2014) 037102, https://doi.org/10.1088/1752-7155/8/3/037102.
[12] J.D. Beauchamp, W. Miekisch, Breath sampling and standardization, in: J. D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier B.V., Amsterdam, 2020, pp. 23–41, https://doi.org/10.1016/b978-0-12-819967-1.00002-5.
[13] Z.-C. Yuan, W. Li, L. Wu, D. Huang, M. Wu, B. Hu, Solid-phase microextraction fiber in face mask for in vivo sampling and direct mass spectrometry analysis of exhaled breath aerosol, Anal. Chem. 92 (17) (2020) 11543–11547, https://doi.org/10.1021/acs.analchem.0c0211810.1021/acs.analchem.0c02118.s001.
[14] X. Li, D. Huang, J. Zeng, C.K. Chan, Z. Zhou, Positive matrix factorization: A data preprocessing strategy for direct mass spectrometry-based breath analysis, Talanta 192 (2019) 32–39, https://doi.org/10.1016/j.talanta.2018.09.020.
[15] K. Rosenthal, D.M. Ruszkiewicz, H. Allen, M.R. Lindley, M.A. Turner, E. Hunsicker, Breath selection methods for compact mass spectrometry breath analysis, J. Breath Res. 13 (4) (2019) 046013, https://doi.org/10.1088/1752-7163/ab34d4.
[16] A. Smolinska, A.-C. Hauschild, R.R.R. Fijten, J.W. Dallinga, J. Baumbach, F.J. van Schooten, Current breathomics - A review on data pre-processing techniques and machine learning in metabolomics breath analysis, J. Breath Res. 8 (2) (2014) 027105, https://doi.org/10.1088/1752-7155/8/2/027105.
[17] G. Stavropoulos, D. Salman, Y. Alkhalifah, F.-J. van Schooten, A. Smolinska, Preprocessing and analysis of volatilome data, in: J.D. Beauchamp, C.E. Davis, J. D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier B.V., Amsterdam, 2020, pp. 633–647, https://doi.org/10.1016/b978-0-12-819967-1.00038-4.
[18] J.D. Pleil, J.R. Sobus, Mathematical and Statistical Approaches for Interpreting Biomarker Compounds in Exhaled Human Breath, in: A. Amman, D. Smith (Eds.), Volatile Biomarkers Non-Invasive Diagnosis Physiol. Med., Elsevier, Oxford, 2013, pp. 3–18.
[19] W.B. Dunn, I.D. Wilson, A.W. Nicholls, D. Broadhurst, The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans, Bioanalysis 4 (18) (2012) 2249–2264, https://doi.org/10.4155/bio.12.204.
[20] R. Rodríguez-Pérez, R. Cortés, A. Guamán, A. Pardo, Y. Torralba, F. Gómez, J. Roca, J.A. Barberà, M. Cascante, S. Marco, Instrumental drift removal in GC-MS data for breath analysis: the short-term and long-term temporal validation of putative biomarkers for COPD, J. Breath Res. 12 (3) (2018) 036007, https://doi.org/10.1088/1752-7163/aaa492.
[21] L. Blanchet, A. Smolinska, A. Baranska, E. Tigchelaar, M. Swertz, A. Zhernakova, J. W. Dallinga, C. Wijmenga, F.J. van Schooten, Factors that influence the volatile organic compound content in human breath, J. Breath Res. 11 (1) (2017) 016013, https://doi.org/10.1088/1752-7163/aa5cc5.
[22] A.C. Hauschild, T. Frisch, J.I. Baumbach, J. Baumbach, Carotta: Revealing hidden confounder markers in metabolic breath profiles, Metabolites 5 (2015) 344–363, https://doi.org/10.3390/metabo5020344.
[23] D. Zanella, J.-F. Focant, J.E. Hill, P.-H. Stefanuto, Comprehensive gas chromatography-mass spectrometry, in: J.D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier, Amsterdam, 2020, pp. 239–251, https://doi.org/10.1016/b978-0-12-819967-1.00015-3.
[24] H. Haick, R. Vishinkin, C. Di Natale, S. Marco, Sensor systems, in: J.D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier B. V., Amsterdam, 2020, pp. 201–220, https://doi.org/10.1016/b978-0-12-819967-1.00013-x.
[25] C.A. Mayhew, J. Herbig, J.D. Beauchamp, Proton transfer reaction–mass spectrometry, in: J.D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier B.V., Amsterdam, 2020, pp. 155–170, https://doi.org/10.1016/b978-0-12-819967-1.00010-4.
[26] D. Smith, P. Španěl, G.B. Hanna, R. Dweik, Selected ion flow tube mass spectrometry, in: J.D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier, Amsterdam, 2020, pp. 137–153, https://doi.org/10.1016/b978-0-12-819967-1.00009-8.
[27] C. Wang, P. Sahay, Breath analysis using laser spectroscopic techniques: Breath biomarkers, spectral fingerprints, and detection limits, Sensors 9 (2009) 8230–8262, https://doi.org/10.3390/s91008230.
[28] D. Salman, G.A. Eiceman, D. Ruszkiewicz, V. Ruzsanyi, E. Brodrick, C.L.P. Thomas, Ion mobility spectrometry, in: J.D. Beauchamp, C.E. Davis, J.D. Pleil (Eds.), Breathborne Biomarkers Hum. Volatilome, Elsevier B.V., Amsterdam, 2020, pp. 171–183, https://doi.org/10.1016/b978-0-12-819967-1.00011-6.
[29] A. Krilaviciute, J.A. Heiss, M. Leja, J. Kupcinskas, H. Haick, H. Brenner, Detection of cancer through exhaled breath: a systematic review, Oncotarget 6 (36) (2015) 38643–38657, https://doi.org/10.18632/oncotarget.5938.
[30] P.H. Stefanuto, D. Zanella, J. Vercammen, M. Henket, F. Schleich, R. Louis, J. F. Focant, Multimodal combination of GC × GC – HRTOFMS and SIFT – MS for asthma phenotyping using exhaled breath, Sci. Rep. (2020) 1–12, https://doi.org/10.1038/s41598-020-73408-2.
[31] T. Saidi, M. Moufid, K. de Jesus Beleño-Saenz, T.G. Welearegay, N. El Bari, A. Lisset Jaimes-Mogollon, R. Ionescu, J.E. Bourkadi, J. Benamor, M. El Ftouh, B. Bouchikhi, Non-invasive prediction of lung cancer histological types through exhaled breath analysis by UV-irradiated electronic nose and GC/QTOF/MS, Sensors Actuators B Chem. 311 (2020) 127932, https://doi.org/10.1016/j.snb.2020.127932.
[32] Y. Sun, Y. Chen, C. Sun, H. Liu, Y. Wang, X. Jiang, Analysis of volatile organic compounds from patients and cell lines for the validation of lung cancer biomarkers by proton-transfer-reaction mass spectrometry, Anal. Methods. 11 (25) (2019) 3188–3197.
[33] E. Gashimova, A. Temerdashev, V. Porkhanov, I. Polyakov, D. Perunov, A. Azaryan, E. Dmitrieva, Investigation of different approaches for exhaled breath and tumor tissue analyses to identify lung cancer biomarkers, Heliyon 6 (6) (2020) e04224, https://doi.org/10.1016/j.heliyon.2020.e04224.
[34] D. Grove, G. Miller-Atkins, C. Melillo, F. Rieder, S. Kurada, D.M. Rotroff, A. R. Tonelli, A.R. Tonelli, R.A. Dweik, R.A. Dweik, Comparison of volatile organic compound profiles in exhaled breath versus plasma headspace in different diseases, J. Breath Res. 14 (2020), https://doi.org/10.1088/1752-7163/ab8866.
[35] A. Smolinska, A.-C. Hauschild, R.R.R. Fijten, J.W. Dallinga, J. Baumbach, F.J. Van Schooten, Current breathomics – a review on data pre-processing techniques and machine learning in metabolomics breath analysis, J. Breath Res. 8 (2) (2014) 027105, https://doi.org/10.1088/1752-7155/8/2/027105.
[36] R. Rodríguez-Pérez, L. Fernández, S. Marco, Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study, Anal. Bioanal. Chem. 410 (23) (2018) 5981–5992, https://doi.org/10.1007/s00216-018-1217-1.
[37] S. Marco, The need for external validation in machine olfaction: emphasis on health-related applications, Anal. Bioanal. Chem. 406 (16) (2014) 3941–3956, https://doi.org/10.1007/s00216-014-7807-7.
[38] R. Rodríguez-Pérez, M. Padilla, S. Marco, The need of external validation for metabolomics predictive models, in: R. Cumeras, X. Correig (Eds.), Volatile Org. Compd. Anal. Biomed. Diagnosis Appl, first ed.,, Apple Academic Press, Boca Raton, 2018, pp. 197–223, https://doi.org/10.1201/9780429433580-8.
[39] D.L. Donoho, The Curses and Blessings of Dimensionality, in: Am. Math. Soc. Lect. Challenges 21st Century, Los Angeles, 2000: pp. 1–33. https://www.dl.icdst.org/pdfs/files/236e636d7629c1a53e6ed4cce1019b6e.pdf (accessed November 8, 2017).
[40] N. Fens, A.C. Roldaan, M.P. van der Schee, R.J. Boksem, A.H. Zwinderman, E. H. Bel, P.J. Sterk, External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease, Clin. Exp. Allergy 41 (10) (2011) 1371–1378, https://doi.org/10.1111/j.1365-2222.2011.03800.x.
[41] M. Schumacher, N. Holländer, W. Sauerbrei, Resampling and cross-validation techniques: a tool to reduce bias caused by model building? Stat. Med. 16 (1997) 2813–2827, https://doi.org/10.1002/(SICI)1097-0258(19971230)16:24<2813::AID-SIM701>3.0.CO;2-Z.
[42] P. Filzmoser, B. Liebmann, K. Varmuza, Repeated double cross validation, J. Chemom. (2009) 160–171, https://doi.org/10.1002/cem.1225.
[43] F. Lindgren, B. Hansen, W. Karcher, M. Sjöström, L. Eriksson, Model validation by permutation tests: applications to variable selection, J. Chemom. 10 (1996) 521–532, https://doi.org/10.1002/(SICI)1099-128X(199609)10:5/6<521::AID-CEM448>3.0.CO;2-J.
[44] N.M. Zetola, C. Modongo, O. Matsiri, T. Tamuhla, B. Mbongwe, K. Matlhagela, E. Sepako, A. Catini, G. Sirugo, E. Martinelli, R. Paolesse, C. Di Natale, Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples, J. Infect. 74 (4) (2017) 367–376, https://doi.org/10.1016/j.jinf.2016.12.006.
[45] C. Lourenço, C. Turner, Breath analysis in disease diagnosis: methodological considerations and applications, Metabolites 4 (2014) 465–498, https://doi.org/10.3390/metabo4020465.
[46] M. Rodríguez-Aguilar, L. Díaz de León-Martínez, P. Gorocica-Rosete, R. Pérez-Padilla, C.A. Domínguez-Reyes, J.A. Tenorio-Torres, O. Ornelas-Rebolledo, G. Mehta, B.N. Zamora-Mendoza, R. Flores-Ramírez, Application of chemoresistive gas sensors and chemometric analysis to differentiate the fingerprints of global volatile organic compounds from diseases. Preliminary results of COPD, lung cancer and breast cancer, Clin. Chim. Acta 518 (2021) 83–92, https://doi.org/10.1016/J.CCA.2021.03.016.
[47] O. Sibila, L. Garcia-Bellmunt, J. Giner, J.L. Merino, G. Suarez-Cuartin, A. Torrego, I. Solanes, D. Castillo, J.L. Valera, B.G. Cosio, V. Plaza, A. Agusti, Identification of airway bacterial colonization by an electronic nose in Chronic Obstructive Pulmonary Disease, Respir. Med. 108 (11) (2014) 1608–1614.
[48] H. Shafiek, F. Fiorentino, J.L. Merino, C. López, A. Oliver, J. Segura, I. de Paul, O. Sibila, A. Agustí, B.G. Cosío, K. Kostikas, Using the electronic nose to identify airway infection during COPD exacerbations, PLoS One. 10 (9) (2015) e0135199, https://doi.org/10.1371/journal.pone.013519910.1371/journal.pone.0135199.g00110.1371/journal.pone.0135199.g00210.1371/journal.pone.0135199.g00310.1371/journal.pone.0135199.g00410.1371/journal.pone.0135199.t00110.1371/journal.pone.0135199.t00210.1371/journal.pone.0135199.t00310.1371/journal.pone.0135199.t00410.1371/journal.pone.0135199.s001.
[49] G. Suarez-Cuartin, J. Giner, J.L. Merino, A. Rodrigo-Troyano, A. Feliu, L. Perea, F. Sanchez-Reus, D. Castillo, V. Plaza, J.D. Chalmers, O. Sibila, Identification of Pseudomonas aeruginosa and airway bacterial colonization by an electronic nose in bronchiectasis, Respir. Med. 136 (2018) 111–117, https://doi.org/10.1016/j.rmed.2018.02.008.
[50] A. Rodrigo-Troyano, O. Sibila, The respiratory threat posed by multidrug resistant Gram-negative bacteria, Respirology 22 (7) (2017) 1288–1299.
[51] J.A. Wedzicha, T.A. Seemungal, COPD exacerbations: defining their cause and prevention, Lancet 370 (2007) 786–796.
[52] J.D. Chalmers, A. Filonenko, M. Shteinberg, P.C. Goeminne, M.J. McDonnell, S. Aliberti, T.C. Fardon, D. Obradovic, E. Operschall, C. Gerlinger, G. Sotgiu,

K. Dimakou, E. Polverino, A. De Soyza, A. Hill, Characterisation of the frequent exacerbator phenotype in bronchiectasis: Data from the friends cohort, Am. J. Respir. Crit. Care Med. 195 (2017) rccm.201711-2202OC. http://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2017.195.1_MeetingAbstracts.A7305%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed18&NEWS=N&AN=617704054.

[53] V. Shestivska, A. Nemec, P. Dřevínek, K. Sovová, K. Dryahina, P. Španěl, Quantification of methyl thiocyanate in the headspace of Pseudomonas aeruginosa cultures and in the breath of cystic fibrosis patients by selected ion flow tube mass spectrometry, Rapid Commun. Mass Spectrom. 25 (17) (2011) 2459–2467, https://doi.org/10.1002/rcm.5146.

[54] A.J. Scott-Thomas, M. Syhre, P.K. Pattemore, M. Epton, R. Laing, J. Pearson, S. T. Chambers, 2-Aminoacetophenone as a potential breath biomarker for Pseudomonas aeruginosa in the cystic fibrosis lung, BMC Pulm. Med. 10 (2010) 56, https://doi.org/10.1186/1471-2466-10-56.

[55] C. Vidaillac, V.F.L. Yong, T.K. Jaggi, -.-M. Soh, S.H. Chotirmall, Gender differences in bronchiectasis: a real issue? Breathe 14 (2) (2018) 108–121, https://doi.org/10.1183/20734735.000218.

[56] E. Polverino, P.C. Goeminne, M.J. McDonnell, S. Aliberti, S.E. Marshall, M. R. Loebinger, M. Murris, R. Cantón, A. Torres, K. Dimakou, A. De Soyza, A.T. Hill, C.S. Haworth, M. Vendrell, F.C. Ringshausen, D. Subotic, R. Wilson, J. Vilaró, B. Stallberg, T. Welte, G. Rohde, F. Blasi, S. Elborn, M. Almagro, A. Timothy, T. Ruddy, T. Tonia, D. Rigau, J. Chalmers, European Respiratory Society guidelines for the management of adult bronchiectasis, Eur. Respir. J. 50 (3) (2017) 1700629, https://doi.org/10.1183/13993003.00629-201710.1183/13993003.00629-2017.Supp110.1183/13993003.00629-2017.Supp2.

[57] O. Sibila, G. Suarez-Cuartin, A. Rodrigo-Troyano, T.C. Fardon, S. Finch, E. F. Mateus, L. Garcia-Bellmunt, D. Castillo, S. Vidal, F. Sanchez-Reus, M.I. Restrepo, J.D. Chalmers, Secreted mucins and airway bacterial colonization in non-CF bronchiectasis, Respirology 20 (7) (2015) 1082–1088, https://doi.org/10.1111/resp.12595.

[58] J. Beauchamp, J. Herbig, R. Gutmann, A. Hansel, On the use of Tedlar® bags for breath-gas sampling and analysis, J. Breath Res. 2 (4) (2008) 046001, https://doi.org/10.1088/1752-7155/2/4/046001.

[59] B. Buszewski, T. Ligor, J. Rudnicka, Clinical application of SPME: analysis of VOCs in exhaled breath as cancer biomarkers, Isocyanate Sampl. (2012) 17–18. http://dialnet.unirioja.es/servlet/articulo?codigo=4192463&orden=388902&info=link#page=17.

[60] R.A. van den Berg, H.CJ. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genom. 7 (1) (2006), https://doi.org/10.1186/1471-2164-7-142.

[61] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B. 57 (1) (1995) 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[62] T.-N. Yang, S.-D. Wang, Robust algorithms for principal component analysis, Pattern Recognit. Lett. 20 (9) (1999) 927–933, https://doi.org/10.1016/S0167-8655(99)00060-4.

[63] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: A new approach to robust principal component analysis, Technometrics 47 (2005) 64–79, https://doi.org/10.1198/004017004000000563.

[64] B.D.B.D. Ripley, W.N.W.N. Venables, in: Modern Applied Statistics with S, Springer-Verlag, New York, 2002, https://doi.org/10.1198/tech.2003.s33.

[65] M. Ojala, Permutation tests for studying classifier performance, J. Mach. Learn. Res. 11 (2009) 1833–1863, https://doi.org/10.1109/ICDM.2009.108.

[66] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779–787.

[67] N.G. Mahieu, J.L. Genenbacher, G.J. Patti, A roadmap for the XCMS family of software solutions in metabolomics, Curr. Opin. Chem. Biol. 30 (2016) 87–93, https://doi.org/10.1016/j.cbpa.2015.11.009.

[68] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics, Anal. Chem. 78 (13) (2006) 4281–4290, https://doi.org/10.1021/ac051632c.

[69] F. Madrid-Gambin, S. Oller-Moreno, L. Fernandez, L. Fernandez, S. Bartova, M. P. Giner, C. Joyce, F. Ferraro, I. Montoliu, S. Moco, S. Marco, S. Marco, AlpsNMR: An R package for signal processing of fully untargeted NMR-based metabolomics, Bioinformatics 36 (2020) 2943–2945, https://doi.org/10.1093/bioinformatics/btaa022.

[70] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (3) (2003) 166–173, https://doi.org/10.1002/cem.785.

[71] J. Rudnicka, M. Walczak, T. Kowalkowski, T. Jezierski, B. Buszewski, Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography-mass spectrometry versus trained dogs, Sensors Actuators, B Chem. 202 (2014) 615–621, https://doi.org/10.1016/j.snb.2014.06.006.

[72] A. Amann, B.d.L. Costello, W. Miekisch, J. Schubert, B. Buszewski, J. Pleil, N. Ratcliffe, T. Risby, The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva, J. Breath Res. 8 (3) (2014) 034001, https://doi.org/10.1088/1752-7155/8/3/034001.

[73] O. Lawal, W.M. Ahmed, T.M.E. Nijsen, R. Goodacre, S.J. Fowler, Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis, Metabolomics 13 (10) (2017), https://doi.org/10.1007/s11306-017-1241-8.

[74] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemom. 28 (2014) 213–225, https://doi.org/10.1002/cem.2609.

[75] H.H. Nir Peled, Meggie Hakim, Paul A. Bunn, York E. Miller, Timothy C. Kennedy, Jane Mattei, John D. Mitchell, Fred R. Hirsch, Non-invasive breath analysis of pulmonary nodules, J. Thorac. Oncol. 7 (2013) 1528–1533, https://doi.org/10.1097/JTO.0b013e3182637d5f.Non-Invasive.

[76] G. Peng, M. Hakim, Y.Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten, U. Tisch, H. Haick, Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors, Br. J. Cancer 103 (4) (2010) 542–551, https://doi.org/10.1038/sj.bjc.6605810.