



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2016

Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry - a pilot study on tobacco smokers

Gaugg, Martin Thomas; Gomez, Diego Garcia; Barrios-Collado, Cesar; Vidal-de-Miguel, Guillermo; Kohler, Malcolm; Zenobi, Renato; Martinez-Lozano Sinues, Pablo

Abstract: Online breath analysis is an attractive approach to track exhaled compounds without sample preparation. Current commercially available real-time breath analysis platforms require the purchase of a full mass spectrometer. Here we present an ion source compatible with virtually any preexisting atmospheric pressure ionization mass spectrometer that allows real-time analysis of breath. We illustrate the capabilities of such technological development by upgrading an orbitrap mass spectrometer. As a result, we detected compounds in exhaled breath between 70 and 900 Da, with a mass accuracy of typically <1 ppm; resolutions between $m/\Delta m$ 22,000 and 70,000 and fragmentation capabilities. The setup was tested in a pilot study, comparing the breath of smokers ($n = 9$) and non-smokers ($n = 10$). Exogenous compounds associated to smoking, as well as endogenous metabolites suggesting increased oxidative stress in smokers, were detected and in some cases identified unambiguously. Most of these compounds correlated significantly with smoking frequency and allowed accurate discrimination of smokers and non-smokers.

DOI: <https://doi.org/10.1088/1752-7155/10/1/016010>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-131026>

Published Version



Originally published at:

Gaugg, Martin Thomas; Gomez, Diego Garcia; Barrios-Collado, Cesar; Vidal-de-Miguel, Guillermo; Kohler, Malcolm; Zenobi, Renato; Martinez-Lozano Sinues, Pablo (2016). Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry - a pilot study on tobacco smokers. *Journal of breath research*, 10(1):016010. DOI: <https://doi.org/10.1088/1752-7155/10/1/016010>

Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry—a pilot study on tobacco smokers

This content has been downloaded from IOPscience. Please scroll down to see the full text.

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 130.60.47.58

This content was downloaded on 19/01/2017 at 15:34

Please note that [terms and conditions apply](#).

You may also be interested in:

[Secondary electrospray ionization-mass spectrometry](#)

P Martínez-Lozano, L Zingaro, A Finiguerra et al.

[Real-time breath analysis with active capillary plasma ionization-ambient mass spectrometry](#)

Lukas Bregy, Pablo Martínez-Lozano Sinues, Maryia M Nudnova et al.

[Secondary electrospray ionization-mass spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study](#)

Pablo Martínez-Lozano Sinues, Elena Landoni, Rosalba Miceli et al.

[Mass spectrometry for real-time quantitative breath analysis](#)

David Smith, Patrik Španl, Jens Herbig et al.

[Compounds enhanced in a mass spectrometric profile of smokers' exhaled breath versus non-smokers](#)

Ievgeniia Kushch, Konrad Schwarz, Lukas Schwentner et al.

[First observation of a potential non-invasive breath gas biomarker for kidney function](#)

Ingrid Kohl, Jonathan Beauchamp, Ferguel Cakar-Beck et al.

[Exhaled breath analysis in suspected cases of non-small-cell lung cancer: a cross-sectional study](#)

M Corradi, D Poli, I Banda et al.

[Analysis of human breath samples using a modified thermal desorption: gas chromatography electrospray ionization interface](#)

James C Reynolds, Modupe A Jimoh, Cristina Guallar-Hoyas et al.



PAPER

OPEN ACCESS

RECEIVED
10 June 2015REVISED
12 December 2015ACCEPTED FOR PUBLICATION
12 January 2016PUBLISHED
11 February 2016

Original content from
this work may be used
under the terms of the
Creative Commons
Attribution 3.0 licence.

Any further distribution
of this work must
maintain attribution
to the author(s) and the
title of the work, journal
citation and DOI.



Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry—a pilot study on tobacco smokers

Martin Thomas Gaugg¹, Diego Garcia Gomez¹, Cesar Barrios-Collado^{1,2}, Guillermo Vidal-de-Miguel^{1,2}, Malcolm Kohler³, Renato Zenobi¹ and Pablo Martinez-Lozano Sinues¹

¹ Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland

² SEADM S.L., Parque Tecnológico de Boecillo, parcela 205, 47151 Boecillo (Valladolid), Spain

³ Pulmonary Division, University Hospital Zurich, Zurich, Switzerland

E-mail: pablo.mlsinues@org.chem.ethz.ch

Keywords: breath analysis, ionizer, mass spectrometry, real-time, heavy volatiles, high mass resolution

Supplementary material for this article is available [online](#)

Abstract

Online breath analysis is an attractive approach to track exhaled compounds without sample preparation. Current commercially available real-time breath analysis platforms require the purchase of a full mass spectrometer. Here we present an ion source compatible with virtually any preexisting atmospheric pressure ionization mass spectrometer that allows real-time analysis of breath. We illustrate the capabilities of such technological development by upgrading an orbitrap mass spectrometer. As a result, we detected compounds in exhaled breath between 70 and 900 Da, with a mass accuracy of typically < 1 ppm; resolutions between $m/\Delta m$ 22 000 and 70 000 and fragmentation capabilities. The setup was tested in a pilot study, comparing the breath of smokers ($n = 9$) and non-smokers ($n = 10$). Exogenous compounds associated to smoking, as well as endogenous metabolites suggesting increased oxidative stress in smokers, were detected and in some cases identified unambiguously. Most of these compounds correlated significantly with smoking frequency and allowed accurate discrimination of smokers and non-smokers.

Introduction

Exhaled breath contains valuable information about metabolic processes taking place within the human body. The development of suitable analytical tools to capture this information non-invasively could support the clinical diagnosis of various diseases in the near future [1–5]. Every exhalation contains hundreds of volatile compounds, including metabolites, inhaled exogenous substances and compounds produced in the oral cavity [6, 7]. Since Pauling discovered more than 200 different compounds in exhaled breath using gas chromatography-mass spectrometry (GC-MS) in the 1970s [8], interest in breath analysis has steadily grown. Over 800 compounds have been identified in exhaled breath of humans [9]. Apart from GC-MS, which is considered the workhorse in the field, several other techniques have evolved over the last decades to analyze breath, for example, electronic sensors, spectroscopic methods and ion mobility spectrometry [5].

Mass spectrometry-based methods provide the highest chemical selectivity, enabling the identification of breath metabolites. Real-time breath analysis is an advantageous approach to track physiological changes over short periods of time. One natural disadvantage accompanying real-time mass spectrometric methods is the lack of prior chromatographic separation. This compromises the identification of isobaric species, hence leading to a considerable loss of information when measuring with insufficient resolution. This is especially important in untargeted studies where one should ideally cover as many metabolites as possible.

Pioneering work from scientists at Sciencelife showed for the first time that their atmospheric pressure chemical ionization (APCI) mass spectrometer—dubbed TAGA (Trace Atmospheric Gas Analyzer)—had the potential to monitor breath metabolites in real-time [10–13]. TAGA no longer exists, but other mass spectrometric techniques have emerged which enable real-time analysis of trace gases. The most prominent

techniques are proton transfer reaction-mass spectrometry (PTR-MS) [14] and selected ion flow tube-mass spectrometry (SIFT-MS) [15]. In SIFT-MS and PTR-MS, in contrast to APCI-MS, ionization of neutral analytes takes place at reduced pressure. As an alternative to PTR-MS and SIFT-MS, secondary electrospray ionization-mass spectrometry (SESI-MS) has shown to be suitable for the sensitive analysis of trace gases in real-time. As in the case of TAGA, the ionization of neutral vapors in SESI takes place at atmospheric pressure, but instead of a corona discharge, an electrospray of pure solvent produces the reactant ions [16, 17]. As a result, one can implement this technique on any commercially available atmospheric pressure ionization (API) mass spectrometer, without having to purchase an entire mass spectrometer. API-MS systems with very high performance are available in many laboratories, and for vapor analysis via SESI-MS one could thus take full advantage of the high performance in terms of sensitivity, mass resolution/mass accuracy and MS/MS capabilities of state-of-the-art mass spectrometry. However, most commercial API-MS systems are designed for the analysis of liquid samples. As a result, the implementation of SESI-MS often requires some modifications of the front-end hardware, compromising its widespread use. Despite this difficulty, some groups have used this method for vapor analysis in different applications [18–33]. To overcome this shortcoming, we have recently developed an optimized SESI add-on, which can be interfaced with virtually any commercial API-MS. This add-on is based on a particularly efficient design, termed low-flow SESI [34], was developed to be incorporated in a trace explosive detector for cargo containers [35]. Similarly, a number of devices are available on the market to implement different ionization strategies such as direct analysis in real time [36] or desorption electrospray ionization [37].

Here, we present the analytical capabilities of this SESI source when coupled to a high-resolution mass spectrometer. We interrogated the breath of smokers on non-smokers with the aim of illustrating the possibilities of this technology for breath research. One motivation is the high interest of clinicians in tests which allow reliable determination of an individual's smoking status. At present, the gold-standard test involves the measurement of blood or urine levels of cotinine [38], which is the primary metabolite of nicotine [39]. Hence, quicker and non-invasive methods to determine the current smoking status are highly desirable.

Experimental section

Online mass spectrometric measurements

The measurements were performed using a commercial, highly efficient low-flow SESI ion source module (SEADM, Spain; figure 1) [34], coupled to an unmodified LTQ Orbitrap (Thermo Fisher). The covered mass range was m/z 50–1000 and the resolving power ranged between 70 000 (at m/z 91) and 22 000

(at m/z 963). Measurements were done in positive ion mode. For chemical identification of some selected compounds, real-time breath MS/MS experiments were carried out using He as the collision gas.

The ion source featured a heated breath sampling Teflon tube ($T = 190$ °C, length 80 cm, i.d. 1.48 mm) with a manometer to monitor the exhalation pressure. After each exhalation, the SESI chamber and sampling line were flushed with nitrogen stemming from the MS curtain-gas flow. The SESI source comprises a heated core ($T = 80$ °C) and focusing and impactor plates ($U_1 = 2.6$ kV, $U_2 = 1.3$ kV) to guide the ions formed towards the MS inlet. Formic acid (Merck, 98–100%, p.a.) 0.1% in H₂O (Merck, LiChrosolv[®], for chromatography), was used as primary electrospray solvent. The emitter was a PicoTip[™] TaperTip[™] nanospray capillary (50 ± 3 μm).

Subjects and sampling

The stability of the setup was initially tested by monitoring a healthy subject over 3 h. A total of 10 non-smoking subjects (two females/eight males) and nine regularly smoking subjects (two females/seven males) participated in this study (anthropometric data included as supplementary data in table S1 (stacks.iop.org/JBR/10/016010/mmedia)). All measurements were collected within 10 d. To ensure repeatability, all subjects exhaled six times for 20 s each, with a pressure of 20 mbar (monitored by an electronic manometer visible to the subjects). The subjects were asked to not eat, smoke, brush their teeth, use chewing gum or drink anything except water within one hour prior to the measurement, in order to minimize confounding factors. Other potential confounding variables such as for example exposure to second-hand smoke were not considered. The study was approved by the local ethical committee (EK 2012-N-49) and all subjects gave written informed consent to participate.

Data analysis

Mass spectra preprocessing

Data processing was done using home-written MATLAB (R2014a, Mathworks Inc.) scripts. In order to be readable by MATLAB, the raw files were converted into the mzXML file format using MSConvert (Proteowizard) [40]. Then, a peak list was generated by shape-preserving piecewise cubic interpolation (10^7 data points) and summation of all spectra. Afterwards, the continuum mass spectra were centroided by summing the intensities around each peak within the full width at half maximum (FWHM). After a baseline adjustment in the time dimension for all peaks, a filter was applied to extract all features that increased during the exhalation phases. To accomplish this task, 4-hydroxy-2,6-nonadienal (m/z 155.1067) and 4-hydroxy-2-decenal (m/z 171.1381) were used as references because they were found to be present in each exhalation of all participants. For the smoker

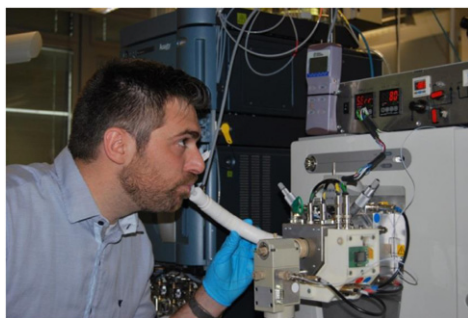


Figure 1. The SESI ion source interfaced with an orbitrap mass spectrometer used for this breath analysis study.

study, additional filtering was applied that eliminates all features, which are only present in fewer than 10 exhalations. Subsequently, the intensities within the last four exhalation phases of each subject were averaged, yielding the raw matrix of intensities (762×19 , $m/z \times \#$ of subjects). To correct for small instrumental fluctuations between the measurements, the matrix was normalized to the sum of intensities within the quantile range of 0.1 and 0.9, making the normalization more robust towards outliers [41]. This procedure (shown schematically in figure S1 of the supplementary data) yielded the final matrix used for further analysis.

Univariate analysis

Once the working data matrix had been assembled, we sought to identify breath compounds that were exhaled at different concentrations between the two groups investigated. Initially, a two-sample *t*-test was performed. Due to the limited sample size, 100 000 bootstrap samples were used to compute *p*-values for the 762 features. In addition, an estimate for the false discovery rate (FDR) was calculated using a linear step-up procedure originally introduced by Storey [42]. It followed a correlation analysis between the peak intensities and the smoking habits (i.e. cigarettes per day) of the subjects.

Smoking status prediction

Subsequently, we sought to determine whether the mass spectral breath prints could be used to predict smoking/non-smoking status. The prediction ability was assessed by performing a leave-one-out cross validation (LOOCV). Instead of using the entire data set for prediction, we implemented a feature selection procedure. It is important to note that the feature selection was also performed without the left-out sample to be truly unbiased. To identify the best predictors for the classification model a genetic selection algorithm was used. In short, a two-sample *t*-test was used as a filtering method [43]. One to three peaks were randomly selected as a training subset from all features with a *p*-value below 0.001 and tested in an inner leave-one-out cross validation. If the misclassification rate was below 10%, the peaks were selected as good predictors. This procedure was

performed 500 times. The most frequently selected features were then used to classify the left-out sample. As classifier, we used a binary support vector machine algorithm [44].

Results and discussion

Real-time breath analysis by high resolution mass spectrometry

One of the main advantages of real-time breath analysis is that it provides a rapid response and therefore circumvents problems associated with sample collection, storage and manipulation. The price one pays is that some selectivity is sacrificed as compared to traditional GC-MS methods. One approach to minimize this drawback is to couple the real-time ionization source with a high resolution mass analyzer. To illustrate this point, we investigated the number of features detected in breath at increasing MS resolution. An orbitrap has the option of increasing resolving power at the cost of scan frequency and some sensitivity. Thus, we detected 660 breath features at a preset resolution of 7500, while at $m/\Delta m = 30\,000$ it increased to 1020 features (i.e. +55%). Figure S2 of the supplementary data provides an overview of the number of features detected and scan frequencies for the different resolution settings.

To illustrate the importance of mass resolution to capture as many compounds as possible in real time, figure 2(a) shows an example of how four isobaric species are resolved at increasing resolving power. At a resolution of 7350 one peak is observable at m/z 300.0753. Inspection of the corresponding time-trace for this feature indicates that it clearly rises during three consecutive exhalation maneuvers (figure 2(b)). By increasing the resolution (e.g. Res $\sim 147\,000$ at 300.0614), four distinct features are resolved. It turned out that all four resolved features rose upon exhalations (figure 2(b)). Hence, even at a typical time-of-flight (TOF) resolution ($m/\Delta m = 5000$), these fine details would have gone undetected in real-time measurements.

An initial stability test of one subject breathing frequently into the mass spectrometer during 3 h revealed that this novel SESI source can deliver robust analysis of breath vapors over extended periods of time and capture transient fluctuations. To illustrate this, figure 3 shows the signal intensity as a function of time (~ 11.00 – 14.00) for indole (tentative assignment), non-2-enal (previously characterized [45]) and an unidentified compound at m/z 479.4829 ($C_{32}H_{62}O_2$). The overall increase of indole during the course of the morning was of a factor of 3, which is consistent with previous observations suggesting that indole breath concentration fluctuates closely in a circadian fashion [46]. In contrast, non-2-enal remained essentially unchanged, while the heavy species at m/z 479.4829 ($C_{32}H_{62}O_2$) showed a decreasing tendency with time. Figure S4 of the supplementary data further illustrates the online detection of heavy species.

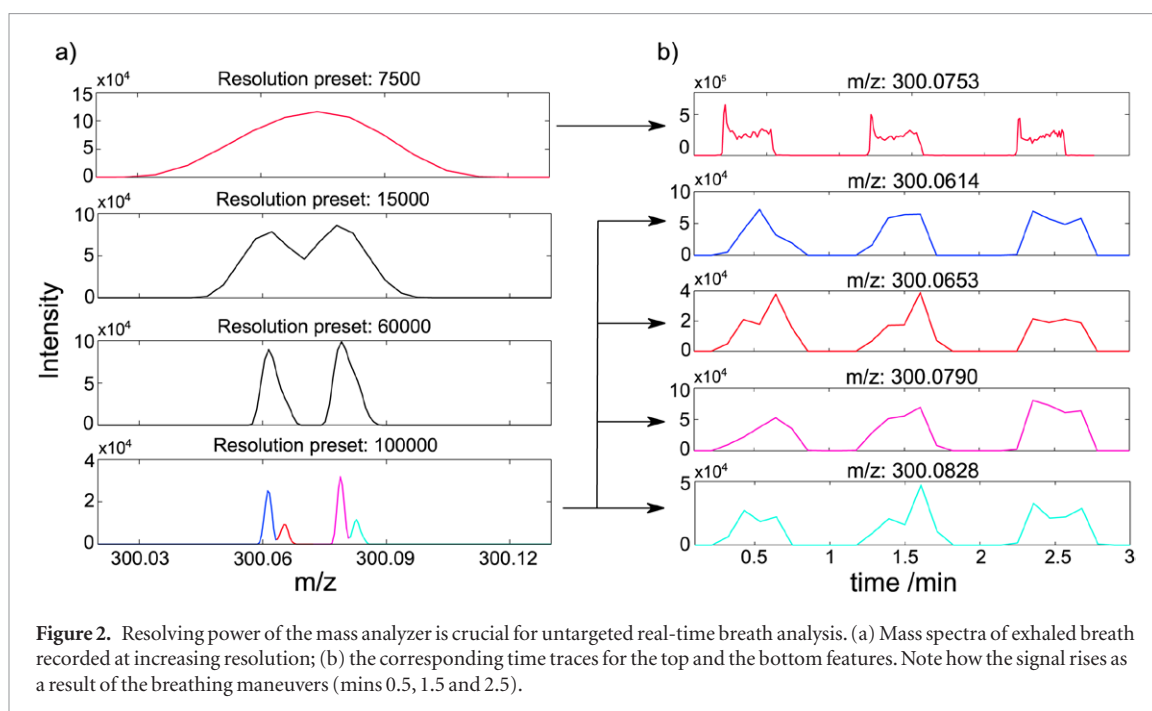


Figure 2. Resolving power of the mass analyzer is crucial for untargeted real-time breath analysis. (a) Mass spectra of exhaled breath recorded at increasing resolution; (b) the corresponding time traces for the top and the bottom features. Note how the signal rises as a result of the breathing maneuvers (mins 0.5, 1.5 and 2.5).

It shows a mass spectrum including background chemical noise and breath signals. The inset shows one of such breath signals as a function of time for the ion m/z 670.1740, which corresponds to a molecular formula $[\text{C}_{18}\text{H}_{18}\text{O}_{13}\text{N}_7]^+$. This assignment was supported by its nearly perfect match with the theoretical isotopic pattern. Expanding the coverage of exhaled molecules to species well above 200 Da is one of the main strengths of SESI-MS. However, given that the response to vapor detection is instrument-dependent, real-time quantification is not directly available [7], unless a calibration procedure using standard vapors is incorporated [21].

Real-time breath analysis in smokers

To complete the evaluation of SEADM's ionizer, we further explored differences in smokers versus non-smokers in a pilot study. The t -test revealed 140 features with masses between 77 and 908 Da, which were significantly different ($p < 0.05$ and $\text{FDR} < 0.05$) between the two groups. Moreover, 100 out of these 140 features also showed a significant correlation ($p_{\text{corr}} < 0.05$) with smoking frequency (i.e. cigarettes per day). The top 62 features, which were found to be highly significant ($p < 0.01$ & upper 95%-CI(p) < 0.05 & $\text{FDR} < 0.05$) are listed in table S2. In total, 68 features were found to be significantly increased and 72 significantly decreased in the breath of smokers. Figure 4 displays the most significantly enhanced feature. It shows the raw mass spectra for all the participants in the region at m/z 114 (figure 4(a)). The corresponding box plot, showing the signal intensities after normalization, is shown in figure 4(b). A plot of signal intensity versus cigarettes/day suggests a strong correlation ($r = 0.88$; $p < 8.6 \times 10^{-7}$) between breath concentration of this particular compound and smoking frequency (figure 4(c)).

A natural advantage accompanying the separation of isobaric species is that isotopic distributions can assist determining molecular formulae with higher confidence. The high mass resolution and accuracy of the orbitrap mass analyzer (< 1 ppm) enabled the possibility of proposing molecular formulae for most of the highly significant features listed. Moreover, for the most abundant molecules, we conducted real-time MS/MS measurements, enabling unambiguous structural elucidation. For example, the highly discriminant compound shown in figure 4 was identified as trimethylsilylacetonitrile. Figure 5(a) shows the overlaid mass spectra of the breath mass spectra and a standard in the region m/z 114–116. The insets show a closer view of the regions of interest, where a nearly perfect match of the isotopic distributions is observed. Note how the high resolution ($R \sim 126\,000$ at 114.0733) enables resolving the isotopic peaks corresponding to ^{13}C (m/z 115.0766) and ^{29}Si (m/z 115.0729). The identification was further confirmed by MS/MS. Figure 5(b) shows the fragmentation spectra for the standard (top) and the breath signal (bottom). Both show a major fragment at m/z 73.047 corresponding to neutral loss of acetonitrile. To our knowledge, this is the first time trimethylsilylacetonitrile has been reported in breath. This compound has not been reported in tobacco, either [47]. Nevertheless, its closely related compound acetonitrile is a well-known compound present in tobacco smoke as well as smokers' breath [48, 49].

Further in-depth analysis revealed that most of the significant features were grouped in chemical families (i.e. $-\text{CH}_2$ -homologous series). In total, we identified seven significant homologous series: A ($\text{C}_5\text{H}_8\text{O}_2$ – $\text{C}_{11}\text{H}_{20}\text{O}_2$), B ($\text{C}_5\text{H}_6\text{O}_2$ – $\text{C}_9\text{H}_{14}\text{O}_2$), C ($\text{C}_6\text{H}_8\text{O}$ – $\text{C}_{11}\text{H}_{18}\text{O}$), D ($\text{C}_6\text{H}_{10}\text{O}_2$ – $\text{C}_{14}\text{H}_{26}\text{O}_2$), E ($\text{C}_5\text{H}_8\text{O}_3$ – $\text{C}_9\text{H}_{16}\text{O}_3$), F ($\text{C}_7\text{H}_{10}\text{O}_3$ – $\text{C}_{10}\text{H}_{16}\text{O}_3$) and G ($\text{C}_6\text{H}_8\text{O}_4$ – $\text{C}_{10}\text{H}_{16}\text{O}_4$).

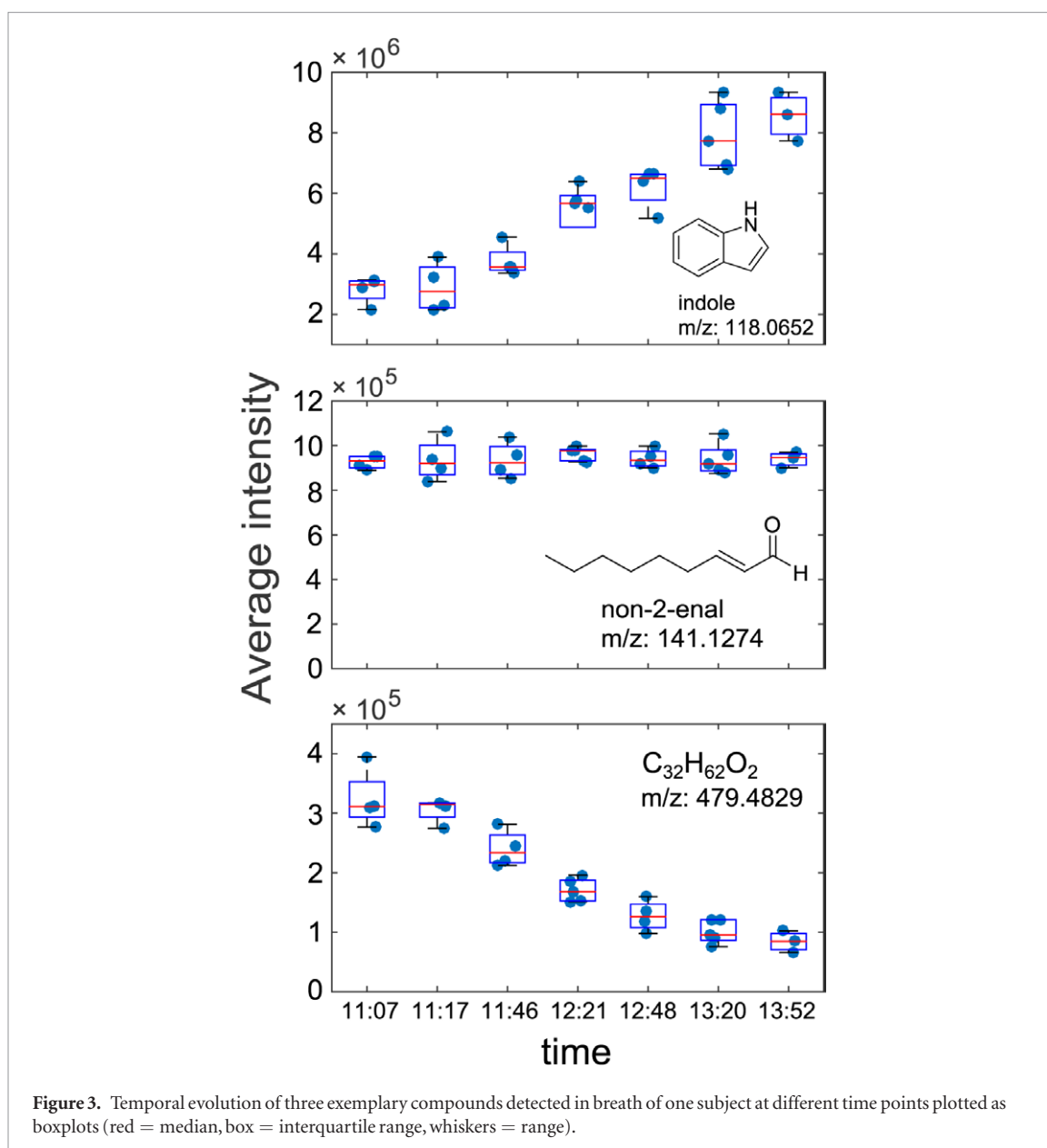
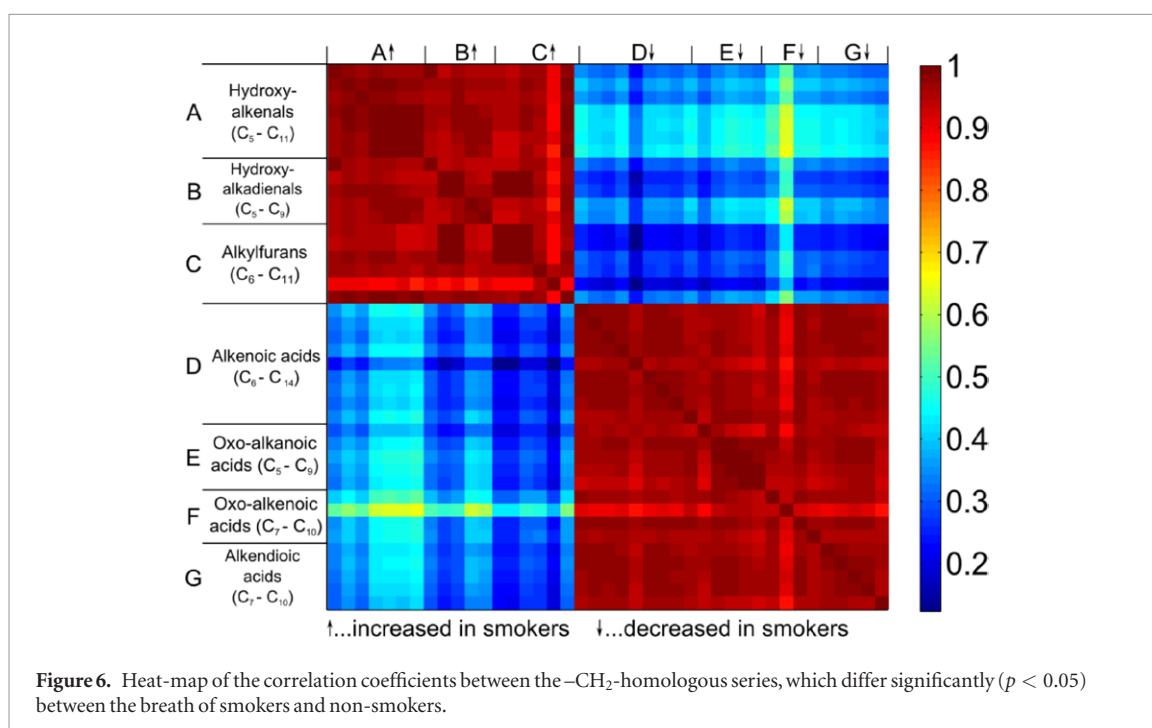
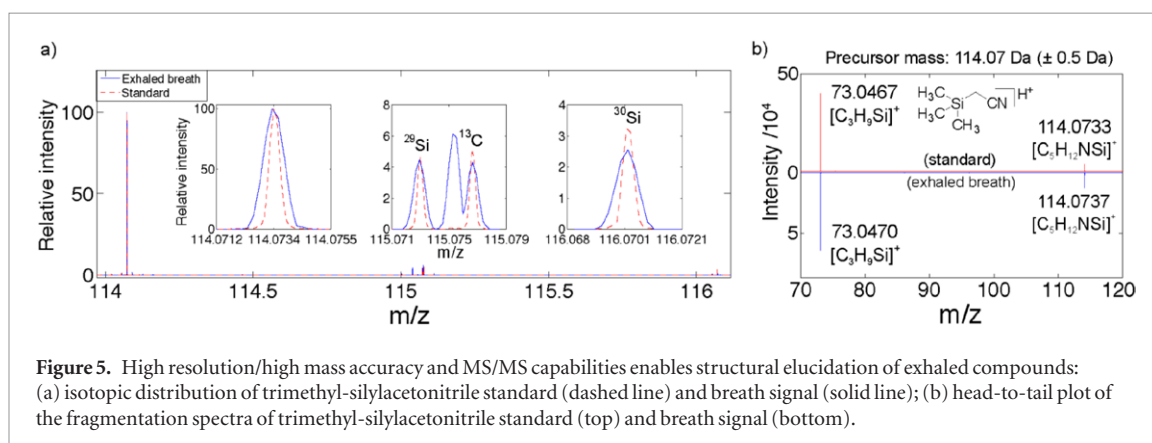
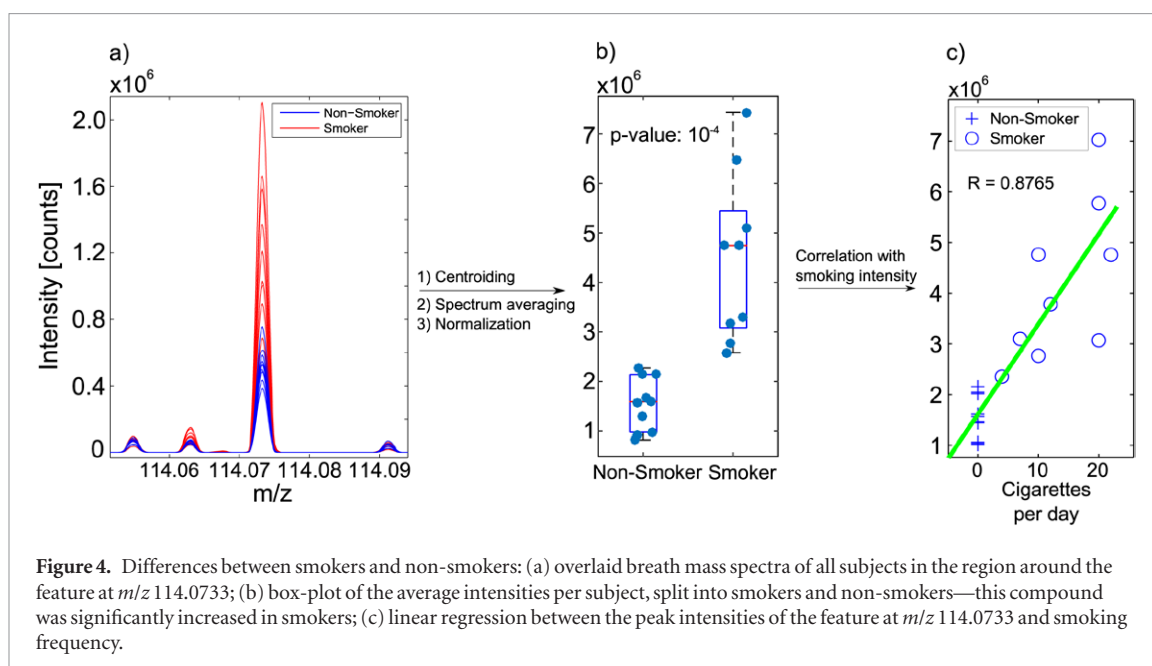


Figure 6 provides an overview of the relationships between the seven chemical families. It shows a heatmap of the correlation matrix between the identified homologous series. The first reassuring observation is the fact that within homologous series, the compounds are highly correlated. At the same time, we found that series A, B and C correlated with each other, whereas D, E, F and G formed another block of correlating compounds. Interestingly, series A, B and C were significantly increased in the breath of smoking subjects (table S3 of the supplementary data). Series A were 4-Hydroxy-2-alkenals and B 4-Hydroxy-dialkenals. For example, figure S3 of the supplementary data shows the MS/MS spectrum and theoretical and experimental isotopic pattern of hydroxy-2,4-hexadienal. This particular compound showed the highest average increase (1.84) in smokers within series B. Hydroxy-2,4-hexadienal is thought to be directly related to tobacco smoke because it is a metabolite of benzene, which is a prominent compound in tobacco smoke [47, 50]. Besides this exception, the rest of com-

pounds of series A and B could be proxy indicators of oxidative stress [51]. This is indeed expected because oxidative stress is one of the main consequences of tobacco smoking [52, 53]. Among these compounds, we found 4-hydroxy-2-nonenal ($p < 0.001$), which is perhaps the most widely studied lipid peroxidation product [54]. While a number of methods exist to monitor 4-hydroxy-2-nonenal in tissues and breath condensate [55, 56], only recently it has been unambiguously detected in real time in breath [45]. The fact that all the compounds, except for $C_7H_{12}O_2$, of the 4-hydroxy-2-alkenals (A) and 4-hydroxy-alkadienals (B) series correlate ($p < 0.05$) with smoking frequency suggests that the degree of oxidative stress was actually monitored by breath analysis. This is consistent with previous studies associating cigarette smoking and measures of lipid peroxidation such as breath ethane [57]. All these observations indicate that not only exogenous compounds (e.g. trimethylsilylacetonitrile) attributable to smoke itself can be



monitored, but also some of the physiological consequences of smoking.

In contrast to the rest of the series identified, series C (i.e. alkyl-furan derivatives), are all exogenous compounds associated with tobacco smoke [58]. Herein we observed alkylfurans with alkyl-residues expanding from C2 (i.e. ethylfuran or dimethylfuran) to C7, all being significantly increased in exhaled breath of smokers. Consistently with previous GC-MS breath studies [59], we found the smallest compounds of the series to be highly discriminating and, except for C₁₀H₁₆O, the whole family highly correlating with smoking frequency.

In contrast to the hydroxyl-alkenals, hydroxy-alkadienals and alkylfurans, series D, E, F and G were all significantly less abundant in the breath of smokers. Hence, it is reasonable to attribute them to a systemic origin rather than smoke/tobacco constituents. These series of compounds were all fatty acids (detected as [M + NH₄]⁺ adduct). These series were assigned to alkenoic acids (D), oxo-alkanoic acids (E), oxo-alkenoic acids (F) and alkendioic acids (G). The reasons why these compounds are systematically decreased in the breath of smokers are unclear. The fact that fatty acids are common precursors of aldehydes in lipid peroxidation routes suggests that the series of aldehydes A and B are enhanced in smokers at the cost of decreased levels of lipids. The connection of these compounds via metabolic routes, however, remains to be established. Along the same lines, prior metabolomics studies have suggested significantly altered profiles of plasma fatty acids profiles for smokers [60].

Finally, when we attempted to predict smoking status based of the breathprints, the feature selection and classification algorithm yielded an out-of-sample classification rate of 100% (sensitivity = 1, specificity = 1). The chosen predictors with their selection frequency in parenthesis were *m/z* 121.0317 (17), 114.0733 (14), 187.1147 (4), 113.0597 (2), 114.0631 (2), 115.0729 (1). The most frequently chosen predictor corresponds to a molecular formula of C₄H₈O₂S (0.4 ppm). The fact that this compound was found almost exclusively in the breath of smokers suggests an exogenous origin. While it could not be unambiguously identified, it might correspond to 1,1-dioxide-tetrahydrothiophene (sulfolan) because it is a known compound in tobacco smoke [47]. The second most frequently selected predictor was trimethyl-silylacetonitrile. Not surprisingly, its isotope at 115.0729 Da was also chosen once as a predictor. For the third predictor (*m/z* 187.1147), we found two possible formulae, C₉H₁₈O₂Si (1.2 ppm) and C₁₀H₁₈OS (2.4 ppm). However, the signal intensity was too low to confirm unambiguously the molecular formula based on its isotopic pattern. Despite the relatively high mass shift, a database and literature search indicated the latter as the more plausible formula, as 8-mercapto-pmenthan-3-one is a compound that has been reported in tobacco. The features *m/z* 113.0597 and 114.0631 correspond to the benzene metabolite 4-hydroxy-2,

4-hexadienal (figure S3 of the supplementary data). Follow-up measurements should be conducted to assess the prediction power of these four compounds.

Conclusions

We present a breath analysis evaluation of a commercial add-on to upgrade a pre-existing atmospheric pressure ionization mass spectrometer with a SESI source meant to analyze vapors at trace levels in real-time. As a result, (i) we were able to measure around 1000 breath features per subject, including species heavier than 900 Da, thus greatly expanding the available mass range of current state-of-the-art on-line breath analysis; (ii) the high resolution/high mass accuracy and MS/MS capabilities of the mass analyzer enabled us to provide molecular formulae and in some cases unambiguous identification of breath compounds; (iii) in a pilot study including nine smokers and 10 non-smokers, we detected multiple compounds in exhaled breath that were highly correlated with smoking frequency. Exogenous compounds as well as systemic metabolites related to oxidative stress were identified, suggesting that both tobacco chemicals as well as physiological implications of smoking could be simultaneously detected; (iv) such detailed instantaneous breathprints enabled predicting smoking/non-smoking status with 100% accuracy.

Acknowledgments

We gratefully acknowledge Dr J Zhang (Novartis AG) for the donation of the LTQ orbitrap instrument used in this study and the European Community's Seventh Framework Programme (FP7-2013-IAPP) for funding the project 'Analytical Chemistry Instrumentation Development' (609691). We are indebted to Christoph Baertschi (ETH workshop) for his assistance machining the ion source.

References

- [1] Martinez-Lozano Sinues P, Zenobi R and Kohler M 2013 Analysis of the exhalome: a diagnostic tool of the future *Chest* **144** 746–9
- [2] Haick H, Broza YY, Mochalski P, Ruzsanyi V and Amann A 2014 Assessment, origin, and implementation of breath volatile cancer markers *Chem. Soc. Rev.* **43** 1423–49
- [3] Amann A, de Lacy Costello B, Miekisch W, Schubert J, Buszewski B, Pleil J, Ratcliffe N and Risby T 2014 The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva *J. Breath Res.* **8** 034001
- [4] Amann A, Miekisch W, Schubert J, Buszewski B, Ligor T, Jezierski T, Pleil J and Risby T 2014 Analysis of exhaled breath for disease detection *Annu. Rev. Anal. Chem.* **7** 455–82
- [5] Amann A and Smith D 2013 *Volatile Biomarkers* (Boston, MA: Elsevier)
- [6] Risby T H and Solga S F 2006 Current status of clinical breath analysis *Appl. Phys. B* **85** 421–6
- [7] Smith D and Spanel P 2015 Pitfalls in the analysis of volatile breath biomarkers: suggested solutions and SIFT-MS quantification of single metabolites *J. Breath Res.* **9** 022001

- [8] Pauling L, Robinson A B, Teranishi R and Cary P 1971 Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography *Proc. Natl Acad. Sci. USA* **68** 2374–6
- [9] de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, Osborne D and Ratcliffe N M 2014 A review of the volatiles from the healthy human body *J. Breath Res.* **8** 014001
- [10] Lovett A M, Reid N M, Buckley J A, French J B and Cameron D M 1979 Real-time analysis of breath using an atmospheric pressure ionization mass spectrometer *Biomed. Mass Spectrom.* **6** 91–7
- [11] Thomson B A, Davidson W R and Lovett A M 1980 Applications of a versatile technique for trace analysis: atmospheric pressure negative chemical ionization *Environ. Health Perspect.* **36** 77–84
- [12] Benoit F M, Davidson W R, Lovett A M, Nacson S and Ngo A 1983 Breath analysis by atmospheric pressure ionization mass spectrometry *Anal. Chem.* **55** 805–7
- [13] Benoit F M, Davidson W R, Lovett A M, Nacson S and Ngo A 1985 Breath analysis by API/MS—human exposure to volatile organic solvents *Int. Arch. Occup. Environ. Health* **55** 113–20
- [14] Lindinger W, Hansel A and Jordan A 1998 On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research *Int. J. Mass Spectrom. Ion Process.* **173** 191–241
- [15] Smith D and Španěl P 1996 Application of ion chemistry and the SIFT technique to the quantitative analysis of trace gases in air and on breath *Int. Rev. Phys. Chem.* **15** 231–71
- [16] Martínez-Lozano Sinues P, Criado E and Vidal G 2012 Mechanistic study on the ionization of trace gases by an electrospray plume *Int. J. Mass Spectrom.* **313** 21–9
- [17] Fernandez de la Mora J 2011 Ionization of vapor molecules by an electrospray cloud *Int. J. Mass Spectrom.* **300** 182–93
- [18] Tam M and Hill H H 2004 Secondary electrospray ionization-mobility spectrometry for explosive vapor detection *Anal. Chem.* **76** 2741–7
- [19] Wu C, Siems W F and Hill H H 2000 Secondary electrospray ionization ion mobility spectrometry/mass spectrometry of illicit drugs *Anal. Chem.* **72** 396–403
- [20] Martínez-Lozano P 2009 Mass spectrometric study of cutaneous volatiles by secondary electrospray ionization *Int. J. Mass Spectrom.* **282** 128–32
- [21] Martínez-Lozano P, Rus J, Fernandez de la Mora G, Hernandez M and Fernandez de la Mora J 2009 Secondary electrospray ionization (SESI) of ambient vapors for explosive detection at concentrations below parts per trillion *J. Am. Soc. Mass Spectrom.* **20** 287–94
- [22] Dillon L A, Stone V N, Croasdel L A, Fielden P R, Goddard N J and Paul Thomas C L 2010 Optimisation of secondary electrospray ionisation (SESI) for the trace determination of gas-phase volatile organic compounds *Analyst* **135** 306–14
- [23] Zhu J, Bean H D, Kuo Y-M and Hill J E 2010 Fast detection of volatile organic compounds from bacterial cultures by secondary electrospray ionization-mass spectrometry *J. Clin. Microbiol.* **48** 4426–31
- [24] Bean H D, Zhu J and Hill J E 2011 Characterizing bacterial volatiles using secondary electrospray ionization mass spectrometry (SESI-MS) *J. Vis. Exp.* **52** e2664
- [25] Martínez-Lozano P, Zingaro L, Finiguerra A and Cristoni S 2011 Secondary electrospray ionization-mass spectrometry: breath study on a control group *J. Breath Res.* **5** 016002
- [26] Meier L, Berchtold C, Schmid S and Zenobi R 2012 High mass resolution breath analysis using secondary electrospray ionization mass spectrometry assisted by an ion funnel *J. Mass Spectrom.* **47** 1571–5
- [27] Berchtold C, Meier L, Steinhoff R and Zenobi R 2014 A new strategy based on real-time secondary electrospray ionization and high-resolution mass spectrometry to discriminate endogenous and exogenous compounds in exhaled breath *Metabolomics* **10** 291–301
- [28] Zhu J, Bean H D, Jiménez-Díaz J and Hill J E 2013 Secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse model study *J. Appl. Physiol.* **114** 1544–9
- [29] Zhu J, Jiménez-Díaz J, Bean H D, Daphtary N A, Aliyeva M I, Lundblad L K A and Hill J E 2013 Robust detection of *P. aeruginosa* and *S. aureus* acute lung infections by secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting: from initial infection to clearance *J. Breath Res.* **7** 037106
- [30] Martin H J, Reynolds J C, Riazanskaia S and Thomas C L P 2014 High throughput volatile fatty acid skin metabolite profiling by thermal desorption secondary electrospray ionisation mass spectrometry *Analyst* **139** 4279–86
- [31] Wolf J C, Schaer M, Siegenthaler P and Zenobi R 2015 Direct quantification of chemical warfare agents and related compounds at low PPT levels: comparing active capillary dielectric barrier discharge plasma ionization and secondary electrospray ionization mass spectrometry *Anal. Chem.* **87** 723–9
- [32] Reynolds J C *et al* 2010 Detection of volatile organic compounds in breath using thermal desorption electrospray ionization-ion mobility-mass spectrometry *Anal. Chem.* **82** 2139–44
- [33] Aernecke M J, Mendum T, Geurtsen G, Ostrinskaya A and Kunz R R 2015 Vapor pressure of hexamethylene triperoxide diamine (HMTD) estimated using secondary electrospray ionization mass spectrometry *J. Phys. Chem. A* **1197** 11514–22
- [34] Vidal-de-Miguel G, Macia M, Pinacho P and Blanco J 2012 Low-sample flow secondary electrospray ionization: improving vapor ionization efficiency *Anal. Chem.* **84** 8475–9
- [35] Vidal-de-Miguel G, Zamora D, Amo M, Casado A, de la Mora G F and de la Mora J F 2012 Method for detecting atmospheric vapors at parts per quadrillion (PPQ) concentrations *US Patent* US20120325024 A1
- [36] Cody R B, Laramée J A and Durst H D 2005 Versatile new ion source for the analysis of materials in open air under ambient conditions *Anal. Chem.* **77** 2297–302
- [37] Takáts Z, Wiseman J M, Gologan B and Cooks R G 2004 Mass spectrometry sampling under ambient conditions with desorption electrospray ionization *Science* **306** 471–3
- [38] Florescu A, Ferrence R, Einarson T, Selby P, Soldin O and Koren G 2009 Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology *Ther. Drug Monit.* **31** 14–30
- [39] Benowitz N L, Hukkanen J and Jacob P III 2009 Nicotine chemistry, metabolism, kinetics and biomarkers *Nicotine Psychopharmacology* (Berlin: Springer) pp 29–60
- [40] Kessner D, Chambers M, Burke R, Agus D and Mallick P 2008 ProteoWizard: open source software for rapid proteomics tools development *Bioinformatics* **24** 2534–6
- [41] Callister S J, Barry R C, Adkins J N, Johnson E T, Qian W-J, Webb-Robertson B-J M, Smith R D and Lipton M S 2006 Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics *J. Proteome Res.* **5** 277–86
- [42] Storey J D 2002 A direct approach to false discovery rates *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **64** 479–98
- [43] Das S 2001 Filters, wrappers and a boosting-based hybrid for feature selection *ICML* **1** 74–81
- [44] Cristianini N and Shawe-Taylor J 2000 *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge: Cambridge University Press)
- [45] García-Gómez D, Martínez-Lozano Sinues P, Barrios-Collado C, Vidal-De-Miguel G, Gaugg M and Zenobi R 2015 Identification of 2-alkenals, 4-hydroxy-2-alkenals, and 4-hydroxy-2,6-alkadienals in exhaled breath condensate by UHPLC-HRMS and in breath by real-time HRMS *Anal. Chem.* **87** 3087–93
- [46] Martínez-Lozano Sinues P, Tarokh L, Li X, Kohler M, Brown S A, Zenobi R and Dallmann R 2014 Circadian variation of the human metabolome captured by real-time breath analysis *PLoS One* **9** e114422
- [47] Rodgman A and Perfetti T A 2013 *The Chemical Components of Tobacco and Tobacco Smoke* (Boca Raton, FL: CRC Press)
- [48] Kohl I, Herbig J, Dunkl J, Hansel A, Daniaux M and Hubalek M 2013 *Volatile Biomarkers* ed A Amann and D Smith (Boston, MA: Elsevier) pp 89–116

- [49] Jordan A, Hansel A, Holzinger R and Lindinger W 1995 Acetonitrile and benzene in the breath of smokers and non-smokers investigated by proton transfer reaction mass spectrometry (PTR-MS) *Int. J. Mass Spectrom. Ion Process.* **148** L1–3
- [50] Monks T J, Butterworth M and Lau S S 2010 The fate of benzene-oxide *Chem. Biol. Interact.* **184** 201–6
- [51] Masashi T, Makoto N, Yoji K, Wakako M and Masayo S-N 2005 *Oxidative Stress, Inflammation, and Health* ed Y-J Surh and L Packer (Boca Raton, FL: CRC Press) pp 423–44
- [52] Halliwell B B and Poulsen H E 2007 *Cigarette Smoke and Oxidative Stress*. (Berlin: Springer)
- [53] Rahman I 2005 *Oxidative Stress, Inflammation, and Health* ed Y-J Surh and L Packer (Boca Raton, FL: CRC) pp 291–368
- [54] Spickett C M 2013 The lipid peroxidation product 4-hydroxy-2-nonenal: advances in chemistry and analysis *Redox Biol.* **1** 145–52
- [55] Rahman I and Biswas S K 2004 Non-invasive biomarkers of oxidative stress: reproducibility and methodological issues *Redox Rep.* **9** 125–43
- [56] Corradi M, Gergelova P, Di Pilato E, Folesani G, Goldoni M, Andreoli R, Selis L and Mutti A 2012 Effect of exposure to detergents and other chemicals on biomarkers of pulmonary response in exhaled breath from hospital cleaners: a pilot study *Int. Arch. Occup. Environ. Health* **85** 389–96
- [57] Miller E R, Appel L J, Jiang L and Risby T H 1997 Association between cigarette smoking and lipid peroxidation in a controlled feeding study *Circulation* **96** 1097–101
- [58] Alonso M, Castellanos M and Sanchez J M 2010 Evaluation of potential breath biomarkers for active smoking: assessment of smoking habits *Anal. Bioanal. Chem.* **396** 2987–95
- [59] Filipiak W *et al* 2012 Dependence of exhaled breath composition on exogenous factors, smoking habits and exposure to air pollutants *J. Breath Res.* **6** 036008
- [60] Müller D C, Degen C, Scherer G, Jahreis G, Niessner R and Scherer M 2014 Metabolomics using GC-TOF-MS followed by subsequent GC-FID and HILIC-MS/MS analysis revealed significantly altered fatty acid and phospholipid species profiles in plasma of smokers *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **966** 117–26