Metabolomics (2014) 10:291–301
DOI 10.1007/s11306-013-0568-z

ORIGINAL ARTICLE

A new strategy based on real-time secondary electrospray ionization and high-resolution mass spectrometry to discriminate endogenous and exogenous compounds in exhaled breath

Christian Berchtold · Lukas Meier · Robert Steinhoff · Renato Zenobi

Received: 26 March 2013 / Accepted: 5 July 2013 / Published online: 14 July 2013

Abstract Breath is considered to be an easily accessible matrix, whose chemical composition relates to compounds present in blood. Therefore many metabolites are expected in exhaled breath, which may be used in the future for the development of diagnostic methods. In this article, a new strategy to discriminate between exhaled endogenous metabolites and exhaled exogenous contaminants by direct high-resolution mass spectrometry is introduced. The analysis of breath in real-time by secondary electrospray ionization mass spectrometry allows to interpret the origin of exhaled compounds. Exhaled metabolites that originate in the respiratory system show reproducible and significant patterns if plotted in real-time (>1 data point per second). An exhaled metabolite shows a signal that tends to rise at the end of a complete (forced) exhalation. In contrast, exogenous compounds, which may be present in room air, are gradually diluted by the air from the deeper lung and therefore show a trend of falling intensity. Signals found in breath by using this pattern recognition are linked to potential metabolites by comparison with online databases. In addition to this real-time approach, it is also shown how to combine this method with classical analytical methods in order to potentially identify unknown metabolites. Finally exhaled compounds following smoking a cigarette, chewing gum, or drinking coffee were investigated to underline the usefulness of this new approach.

Electronic supplementary material The online version of this article (doi:10.1007/s11306-013-0568-z) contains supplementary material, which is available to authorized users.

C. Berchtold · L. Meier · R. Steinhoff · R. Zenobi (✉)
Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zurich, Switzerland
e-mail: zenobi@org.chem.ethz.ch

Keywords Breath analysis · Mass spectrometry · Secondary electrospray ionization · Identification strategy

Abbreviations

APCI	Atmospheric pressure chemical ionization
Da	Dalton, molecular mass unit g/mol
ECB	Exhaled breath condensate
EI-MS	Electron impact mass spectrometry
HMDB	Human metabolite data base
HPLC	High performance liquid chromatography
MS	Mass spectrometry
PTR-MS	Proton transfer reaction mass spectrometry
SESI	Secondary electrospray ionization
SIFT-MS	Selected ion flow tube mass spectrometry
SPME	Solid phase micro extraction
UPLC	Ultra performance liquid chromatography

1 Introduction

An important aspect of modern metabolomics investigations is the development of diagnostic methodologies and techniques. One of the key challenges is always the identification of unknown compounds and potential biological markers. Metabolomic studies are usually performed using blood, urine or tissue samples (Fiehn 2002). A quite novel approach in metabolic research is the use of breath as a sample. Breath is a matrix that can sampled non-invasively, has a relatively low level of complexity, and provides information related to the actual health status. Therefore, breath is very interesting for metabolomics studies, especially for the development of diagnostic tests (Buszewski et al. 2007; Kim et al. 2012).

Breath is composed of atmospheric gases, such as nitrogen, oxygen and is enriched with carbon dioxide,

water vapor and nitric oxide in the vol% to ppmV range. Volatile organic compounds (VOCs), with a boiling point <100 °C and a molecular weight <200 Da, are usually found in the ppmV–ppbV range (Risby 2008). In addition to VOCs, compounds of low volatility are also present among the exhaled metabolites (Cathcart et al. 2012). These are exhaled in the ppbV–pptV concentration and are in some cases probably exhaled as aerosols (Johnson and Morawska 2009). Exhaled VOCs and compounds of low volatility represent metabolites that can serve as markers for various diseases (Buszewski et al. 2007; Kim et al. 2012). While the human metabolome consist of more than 200,000 compounds (Fiehn 2002), only about 2,000 VOCs have been detected in breath In addition to that, a small number of metabolites of low volatility has also been found (Effros et al. 2004). In order to gain an overall understanding of human metabolism and to access the broadest range of diagnostic molecules it is necessary to measure exhaled volatiles as well as compounds of low volatility. Diagnostic approaches based on exhaled metabolites are still quite rare although it has been known for centuries that smell of breath can be used as a diagnostic tool. Acetone in exhaled breath, as an example of volatiles, is known as a marker for diabetes. The detection of lactase deficiency (hypolactasia, lactose intolerance) via exhaled breath is fairly routine. It is detected by measuring exhaled hydrogen after lactose consumption (Oberacher et al. 2011). Exhaled urea is used as a marker for the infection by helicobacter prioli, and the detection of nitric oxide is used in asthma therapy (Risby and Solga 2006). Diagnostics based on breath have been investigated for cancer (Mazzone et al. 2012), COPD (Cazzola and Novelli 2010), or lung inflammation (Carpagnano et al. 2004), but are currently not used routinely.

Breath analysis can be divided into off-line and real-time techniques. The off-line collection of breath can provide a high level of up-concentration and therefore the highest sensitivity. Solid-phase micro extraction (SPME) in combination with gas chromatography (GC) and electron impact (EI) ionization has to be considered the gold standard for the analysis of VOCs in exhaled breath. Studies of exhaled breath by Phillips et al. identified almost 2000 volatile compounds in exhaled breath (Phillips et al. 1999). For compounds of lower volatility, filters or condensers have been used. Exhaled breath condensate (EBC), which is the most important technique to sample compounds of low volatility in breath, allows to pre-concentrate large volumes of exhaled breath and is therefore considered to be a good sampling method for very sensitive measurements (Beck et al. 2011; Cathcart et al. 2012; Dwyer 2004; Effros et al. 2004; Hunt 2002). Even viral DNA has been detected in EBC (Costa et al. 2011). EBC was also used to identify markers of low volatility such as isoprostane (Shahid et al. 2005) or leukotriene (Wardlaw et al. 1989), which are related to lung inflammation. Off-line methods provide high sensitivity and allow investigations of exhaled compounds by chromatographic methods. Therefore, compound identification strategies are similar to "classical" metabolomics using blood or urine.

A number of technical aspects have to be considered for real-time detection of compounds in exhaled breath. "Real-time" in this context means that exhaled air is analyzed breath-by-breath using a time resolution of more than one scan per second. The classical method for real-time breath analysis, such as APCI (Benoit et al. 1983) direct EI-MS (Van den Velde et al. 2008), PTR-MS (Smith and Španel 2011), or SIFT-MS (Španel and Smith 2011) are well suited to monitor compounds with molecular weights below 200 m/z; for example, metabolites such as isoprene (Schwoebel et al. 2011), acetone (Thekedar et al. 2009) or ammonia (Smith et al. 2008) have been monitored. However, the total number of metabolites detected in real time is only in the range of about 10–20 (Martinez-Lozano et al. 2011), and metabolites of higher molecular weight ($m/z > 200$) are not accessible with these methods.

Exhaled metabolites of higher molecular weight are potentially of enormous value. Firstly it is expected that high molecular weight metabolites are more specific for certain metabolic pathways. Secondly, if a higher number of metabolites are detected in real time, the chance of finding a truly specific biomarker or metabolite pattern is significantly enhanced. Therefore a new method is needed, which allows to detected volatiles as well as molecules of low volatility in real time.

To measure compounds of molecular weights of more than 200 Da in real time and with good sensitivity (<ppb), direct ionization methods are needed. Secondary electrospray ionization (SESI) and related techniques have been introduced as matrix-tolerant and efficient ionization method for the detection of compounds of low volatility in the gas phase by mass spectrometry (Alberici et al. 2010; Biasioli et al. 2011; Chen et al. 2009a, b; Huang et al. 2011). The detection limit of SESI–MS was tested for a variety of mass spectrometers and samples and estimated to be below pptV (Martinez-Lozano and Fernandez de la Mora 2009). Accordingly, fatty acids were quantified in breath and estimated to be in the range of 100 pptV (Martinez-Lozano et al. 2011). However, it was shown that the sensitivity of SESI for compounds of low volatility was about 100 times lower than when measuring them by direct ESI (Meier et al. 2011). The sensitivity is reduced due to the reduced ionization probability in the secondary ionization step (Sinues et al. 2012) and due to Coulomb repulsion in front of the MS (Meier et al. 2012c). To reduce this loss, Meier et al. introduced an atmospheric-pressure ion funnel for SESI, and were able to enhance the

sensitivity by a factor of ≈1,000 compared to conventional SESI (Meier et al. 2012a, c). This system was also hyphenated to a high-resolution mass spectrometer to reach a new level of resolution (>100,000) and sensitivity (pptV) (Meier et al. 2012a, c).

The main goal of this study was to investigate the potential of this novel ionization stage, interfaced with a high-resolution mass spectrometer, to discriminate between endogenous metabolites and exogenous contaminants in exhaled human breath. A new strategy to achieve this based on real-time breath patterns was developed. A combination of manual and automatic identification by the use of open access databases is proposed to allow tentative assignments of compounds. The ion source could also be equipped with an off-line sampling system in order to perform complementary analyses on the same instrument. This newly proposed strategy was applied to several exhaled compounds in breath, e.g. after smoking a cigarette, drinking coffee, or chewing gum.

2 Materials and methods

2.1 Materials and reagents

The calibration mix consisted of caffeine, MRFA (Met-Arg-Phe-Ala), a component of the ProteoMass^{TM} LTQ/FT-Hybrid ESI Pos. mode CalMIX, as well as nicotine, all of which were obtained from Sigma-Aldrich (Buchs, SG, Switzerland). Norfentanyl, morphine, cocaine, naloxone, fentanyl, and sufentanil were all obtained from Lipomed AG (Arlesheim, Switzerland). Methanol and acetic acid were obtained from Acros Organics (Geel, Belgium). Nanopure water with a resistivity of >18.1 MΩ cm was obtained from a NANOpure water purification system (Barnstead, IA, USA).

2.2 Samples

Samples were collected from a healthy volunteer (a member of our research team) For real-time experiments the person was asked to exhale constantly for about 30 s after taking a deep breath in order to ventilate the whole lung. The mouthpiece geometry as well as the size and shape of the ion funnel inlet limited the flow rate to about 7 L/min. For offline evaluation, about 30 breath strokes were collected on cotton in the same way. The cigarette (Parisienne Orange, British American Tobacco Switzerland) 4 mg tar, 0.4 mg nicotine) and chewing gum (strawberry flavor Stimorol, Kraft Foods Schweiz GmbH) were purchased in a local grocery store. Coffee was a large espresso brewed with Arabica beans (Lavazza).

2.3 LTQ-FT instrument and SESI-interface

All experiments in this study were conducted on a commercial linear ion trap Fourier-transform ion cyclotron mass spectrometer (LTQ-FT Ultra, Thermo Finnigan, San Jose, USA). Neutral molecules were transferred into an atmospheric pressure ion funnel and ionized by a SESI setup as published before (Meier et al. 2012a, c). Detailed information on the SESI interface as well as on the setup and construction has been reported elsewhere (Meier et al. 2012b). A small schematic is shown in the supplementary information (Fig. S1). Breath is introduced into the system via a disposable non-return valve mouth piece (ACE series, ACE GmbH, Freilassing, Germany) to prevent saliva from entering. A Teflon tube is used to transfer breath into the interface (ID 10 mm, OD 12 mm). For offline analyses, breath was collected on 100 mg cotton supported by a Teflon tube. The Teflon tube was placed in the spray unit after breath was collected. The charging spray unit was heated to 75 °C (less than the boiling point of the solvent mixture used) with three heating cartridges (Probag Wärmetechnik AG, Niederbuchsiten, Switzerland) to prevent sample molecules from condensing. Heat transfer through an aluminum adapter resulted in warming of the breath interface to roughly 40 °C.

The charging spray solvent consisted of a MeOH:H_2O mixture in a 1:1 ratio acidified with 1 % acetic acid, and three identical sprays were infused at 0.05 mL/min each. A voltage of 3 kV was applied. The transfer capillary was held at 250 °C. All other parameters were optimized for maximum ion yield. Mass spectra were collected in scanning mode in the *m/z* range of 100–500 Da. The resolution was set to nominally 150,000 and the maximum accumulation time to 100 ms. The linear trap was set to maximum transmission and the ion transfer was tuned and calibrated by the use of a self-made calibration mix. The best ion transmission was obtained in the range of 200–400 Da.

2.4 Validation of mass resolution and accuracy

The system was optimized for best ion transmission, resolution and mass accuracy. Mass calibration and tuning of all relevant parameters were performed by the standard ion source and the calibration solution. The ion funnel interface was then installed and optimized on the inlet of the LTQ-FT instrument. This optimization was previously described in more detail by Meier et al. (2012b). Mass resolution and accuracy were frequently checked by analyzing known background signals. The current setup for these experiments achieved a mass resolution of 102,460 and a mass accuracy of <0.9 ppm. The latter value was evaluated at the signal of dioctylphthalate, exact mass 391.284835 ($[M+H]^+$), measured mass 391.2845. On the

lower end of the mass spectrum *m/z* 130.1590 ($[M+H]^+$) was used, most likely dioctylamine, which showed an experimental mass accuracy of <4.4 ppm. These background signals were used to check the calibration during the study, to recognize, for example, any form of calibration shift. The calibration itself was always based on the calibrants injected by ESI prior to each measurement series. For calculations and further data interpretation, an accuracy of 5 ppm and a resolution of 100,000 were assumed, based on the data found for the background signals used (dioctylphthalate and dioctylamine).

2.5 Data acquisition and library searches

Data acquisition was controlled by the Xcalibur 2.0 software (Thermo Fisher Scientific, Waltham, MA, USA). The conditions were in general the same as published by Meier et al. recently, if not indicated otherwise (Meier et al. 2012b).

For library searches, ChemSpider (http://www.chemspider.com), the Human Metabolite Database (HMDB http://www.hmdb.ca) and METLIN (http://metlin.scripps.edu/) were used. ChemSpider is a very general database, which covers many structures from many sources (about 28 Million structures, Dec 2012). METLIN covers metabolites from any organism (about 65,000 structures, Dec 2012), whereas HMDB is focused on human metabolites only (40,000 structures, Dec 2012).

3 Results and discussion

3.1 Discerning metabolites in exhaled breath

Although the investigation of breath is comparable to that of other matrices, the direct ionization combined with the real-time nature of our acquisition requires new data interpretation steps. Since no separation is performed prior to ionization (e.g. chromatography), the use of high-resolution mass spectrometry is a good approach to separate potential metabolites. In the mass range of 50–500 *m/z*, up to 5,700 signals are detected in each breath stroke. This includes isotopic patterns, solvent clusters, contaminants, and exhaled metabolites. Figure 1 shows the mass spectra of humidified air, clean breath, breath after smoking, chewing gum, and drinking coffee. Only a fraction of these signals are actually related to metabolites or exhaled exogenous compounds.

Figure 2 shows the general strategy we propose for distinguishing exhaled endogenous metabolites and exogenous compounds with ambient ionization mass spectrometry. The strategy is divided into three main steps, as shown in Fig. 2a–c; this is discussed in the following

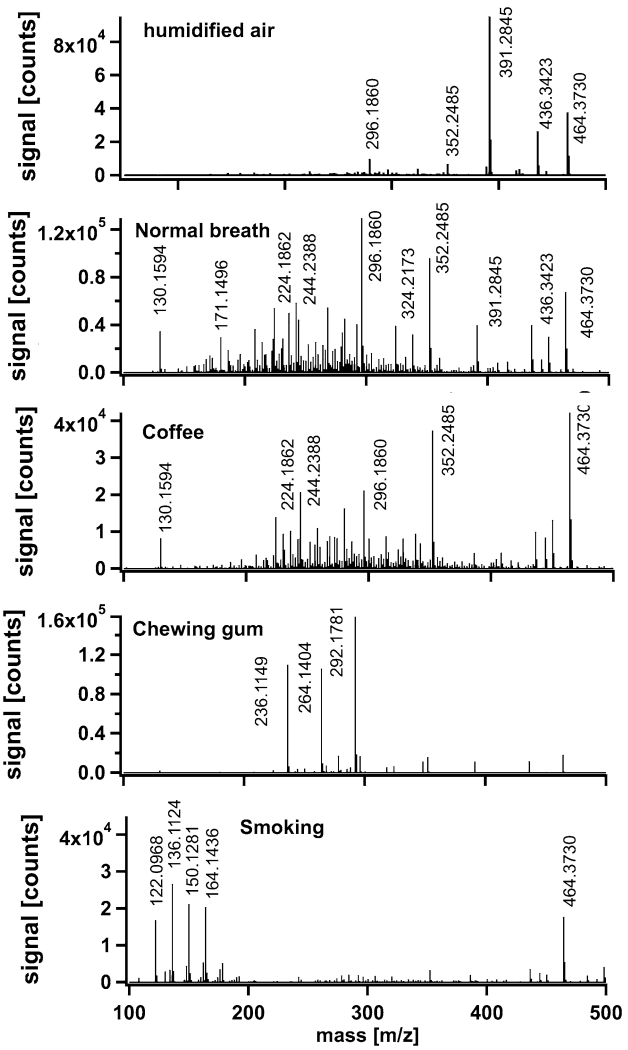

Fig. 1 SESI mass spectra of exhaled breath. Mass spectra of real-time detection of **a** humidified air, **b** normal breath, **c** breath after drinking coffee, **d** chewing gum and **e** smoking a cigarette are shown. Each spectrum represents the summarized spectra of one breath stroke. The intensity is different since the duration of a measurement, the humidification as well as the temperature are not always identical

paragraph. This strategy is based on the real-time data evaluation of exhaled breath and not on chromatographic separation, therefore it represents a new approach for discovering metabolites. In step A (Fig. 2a) the mass spectrum is collected breath-by-breath and all the detected masses are listed. For each breath stroke a mass spectrum is collected, as shown in Fig. 1. It is not possible to perform background subtraction to distinguish exhaled metabolites from background signals: the signal intensity in SESI is strongly influenced by humidity and sample gas flow rate (Dillon et al. 2010; Sinues et al. 2012). Temperature, flow rate and the level of humidification of breath are neither readily reproducible nor stable; as such, a subtraction of the signals that appear in the background from those observed in a breath stroke is not helpful as this might actually cause

artifacts. In addition to this, no reference blank exists for endogenous metabolites in breath. Moreover, there is no pure instrumental blank. In classical LC–MS metabolomics, the appearance of signal in a blank sample would be classified as a contaminant; this is not possible in real-time breath analysis. Nevertheless, recording a blank (such as humidified air) is necessary to assist the identification of unknown metabolites during the breath-by-breath evaluation. Step B of our interpretation strategy (Fig. 2b) takes this new aspect into account.

To keep the number of signals manageable for manual interpretation we reduced the total number of signals to 500 (the 500 strongest signals). Many of these signals had matches in open access databases such as METLIN or HMDB. This process was very fast and provided approximately 130 hits for 25 masses. These 25 masses represented good candidates for further investigation and their limited number entailed that it was feasible that further manual interpretation could be applied. Nevertheless, we recognized that many unknown metabolites could not be processed if only database hits were used. As such, the 30 most intense signals that did not produce any hits in the databases were also added for further investigations. For positive ion mode, a total of 55 signals underwent further processing in order to identify metabolites, exhaled exogenous compounds or internal background signals by analyzing their real-time pattern.

The real-time pattern of an exhaled compound is significantly influenced by its origin. This is due to the physiology of breathing and the method of sampling. Since the volunteer in our experiment was asked to exhale for as long as possible and rest in between breath strokes, each stoke appeared quite similar and represented the whole lung volume. The first 150 mL of exhaled air is the so-called dead space, which consists of freshly inhaled air that has very little or no interaction with the lung surface (alveoli, bronchiole). Approximately 500 mL of the exhaled air after this is the normally exhaled volume (for a calm person) that already contains some metabolites, but also a high fraction of room air. This volume was slightly enhanced as the result of the fact that the subject took a deeper inhalation and was holding breath for a few seconds, which will lead to the accumulation of metabolites. The residual air (about 5 L in total) originated from deep inside the lung and contained the main fraction of exhaled metabolites (tidal breath). This physiological fact directly affected the real-time breath spectrum. This aspect is problematic for quantitative studies in breath, but very helpful for untargeted metabolite studies. An exhaled metabolite should rise slowly during one breath stroke or remain at least stable, whereas an exogenous compound should start with a high value before being gradually diluted by the air from the deeper lung. This pattern should

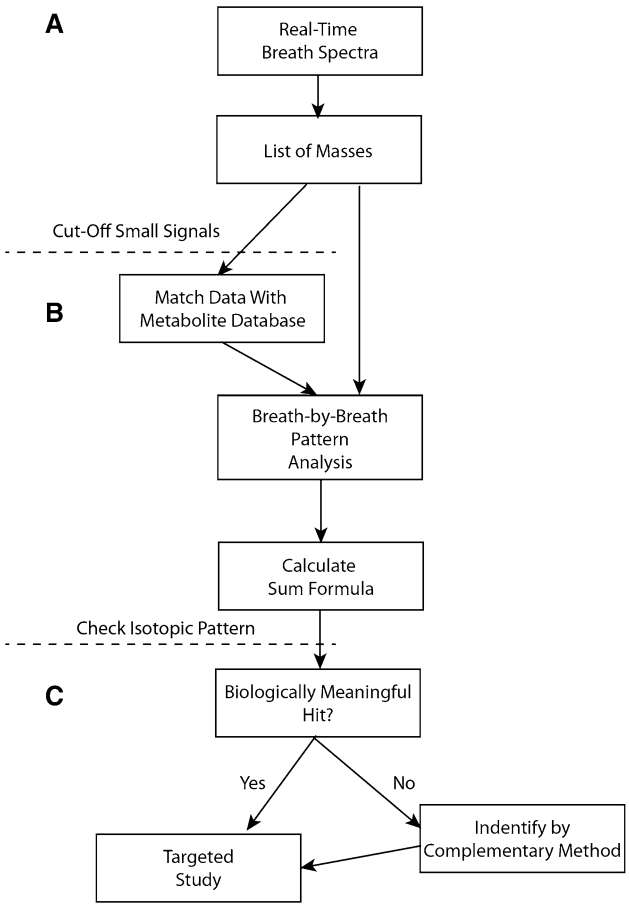

Fig. 2 Strategy for assignment to exhaled compounds. The spectra of one or several breath strokes are averaged, which will provide about signals 5,700 for clean breath. This mass list collected in section (*a*) is the basis for further investigations. By eliminating the smallest signals (<40 counts/scan) a list of about 500 signals is generated. This mass list is either matched directly with databases or analyzed according to the exhaled breath-by-breath pattern. About 25 compounds that provided significant database hits and the 30 most intense signal without hits were process further. These about 55 signals were processed according to the breath-by-breath behavior. About 30 compounds were identified as exogenous exhaled compounds or exhaled metabolites. The sum formulae of these compounds were calculated and crosschecked by databases and the isotopic pattern. About 26 exhaled compounds were identified by the use of this strategy. If there is a conclusive biological hit in a database, this marker may be investigated directly in a targeted study. If there is no hint for identity according the accurate mass, the identification has to be performed with complementary methods prior to a targeted study. In section (*c*) this additional identification loop, which may include offline sampling and classical HPLC–MS, is shown

be reproducible from breath stroke to breath stroke, since some background signals may behave in an unpredictable fashion.

Figure 3 depicts the example of two ions (*m*/*z* 391.2850 and 293.1795), which regularly appeared in exhaled breath in experiments over the course of weeks. The signal of *m*/*z* 391.2850 represented a trend that indicates a gradual dilution by the breath from the deeper lung, whereas

m/z 293.1795 remained stable, which indicates a metabolite. Humidified air (before the first stroke in Fig. 3) was used as a baseline; signals that are intensity enhanced by humidity and flow could be compared and identified with this baseline. Additional patterns of putative endogenous and exogenous compounds are shown in the supplementary Fig. S2–S4. Using the breath-by-breath pattern analysis for the 55 extracted masses (positive ion mode), about 21 potential metabolites and 5 potentially exogenous exhaled compounds could be distinguished. The remaining of signals were most likely contaminations originating from inside the mass spectrometer or from the electrospray solvent; they did not show a reproducible pattern. This interpretation step, which was performed manually in the present case, can potentially be fully automated and could be applied to all signals detected in the first step. This would also allow the detection of metabolites of very low intensity.

The qualified list of compounds derived from steps A and B in the strategy (Fig. 2) is confirmed and refined by identification algorithms to identify compounds detected by high-resolution mass spectrometers. Kind and Fiehn (2007) introduced a strategy to identify metabolites measured by HPLC–MS, called the "seven golden rules". These rules were adapted to the data obtained by our method to generate and confirm sum formulae. The sum formulae were determined and validated by help of the associated isotopic patterns. The 26 signals listed in Table 1 represent the "survivors" from this process in positive ion mode, including the suggested sum formula. The table also shows the numbers of hits in the database. The signals in negative ion mode, which were evaluated in the same way, are also listed. The main difference is that in negative ion mode, fewer signals were detected (<600 in total). This strategy (Fig. 2a, b) provides a first list of potential exhaled metabolites and exogenous compounds.

The final step of the process is to identify the exhaled compounds (Fig. 2c). A few of the compounds listed were already matched with hits in the database. As an example, lactate (*m/z* 89.0244) was identified with some confidence, since it has been reported in breath before (Marek et al. 2010; Martinez-Lozano and Fernandezde la Mora 2008). Some exogenous exhaled compounds, such as the dioctylphthalate, a very common plasticizer, were also easily identified (*m/z* 391.2845). Other compounds provide hits that were theoretically possible, but have never been found in exhaled breath. One such candidate was $m/z = 352.2850$, which could be capnine (according to METLIN). Capnine is a well-known metabolite of the lipid metabolism of bacteria (Abbanat et al. 1985; White 1984). Either there is an unknown metabolic pathway in humans which produces this compound, or bacteria living in the mouth cavity or the lung produce it. In the future, these hits need to be identified and substantiated, e.g., by comparison with blood levels, by the use of authentic standards, and by working with a statistically relevant number of volunteers. This may be work for the compounds already known, however many of the compounds listed did not result in

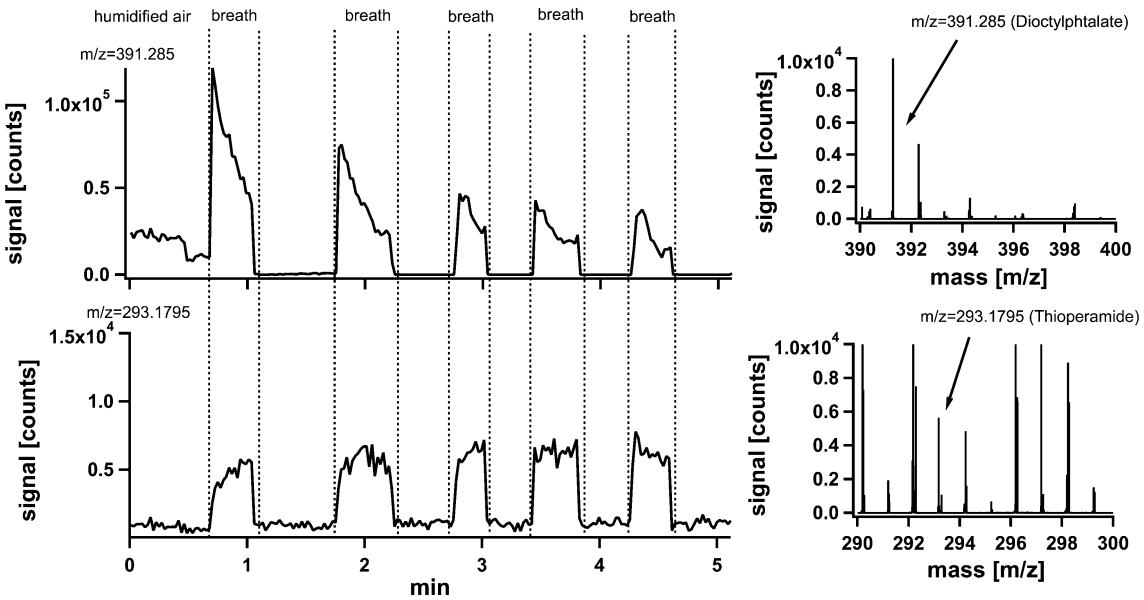

Fig. 3 Breath-by-breath patterns from the real-time extracted ion chromatogram. The *dotted lines* separate long exhalation phases from brakes between measuring breath samples. At the beginning of the experiment, humidified air was blown into the system. The *top* graph shows the extracted ion chromatogram for *m/z* 391.285. It most likely represents an exogenous contaminant, probably dioctylphthalate. The breath strokes enhance the signal and transport the phthalate into the ion source. $M/z = 293.1795$ could refer to the singly protonated ion of thioperamide, which may represent a metabolite

Table 1 Masses with estimated metabolites or contaminants in real time breath are listed

m/z	Elemental composition	Chem spider	METLIN	HMDB	Pattern	Possible assignment	Off-line
Positive mode							
130.1590	$C_8H_{19}N$	110	1	7	E	Octylamine	Yes
168.1242	$C_9H_{15}N_2O$	17	0	0	M		No
173.1654	$C_9H_{20}ON_2$	421	0	0	M		No
203.2133	$C_{11}H_{26}N_2O$	36	0	0	M		No
212.1509	$C_9H_{17}N_5O$	56	1	0	M		No
214.1801	$C_{12}H_{23}NO_2$	488	0	92	M		No
216.1961	$C_{12}H_{25}NO_2$	229	1	37	M	12-Amino-dodecanoic acid	No
217.2287	$C_{12}H_{28}N_2O$	17	0	0	M		No
223.0980	$C_{13}H_{10}N_4$	172	0	0	E		No
230.2117	$C_{13}H_{27}NO_2$	185	4	13	M		No
231.2430	$C_{13}H_{30}N_2O$	158	0	0	M		No
256.1693	$C_{17}H_{21}NO$	1,947	4	4	M		No
264.1448	$C_{11}H_{21}NO_6$	56	0	8	M		No
268.1550	$C_{14}H_{21}NO_4$	1,043	0	7	M		No
279.1590	$C_{16}H_{22}O_4$	925	7	19	E		No
296.1854	$C_{16}H_{25}NO_4$	657	2	6	M		No
302.1934	$C_{11}H_{23}N_7O_3$	1	0	1	M		No
302.2333	$C_{16}H_{31}NO_4$	0	0	6	M		No
302.2696	$C_{17}H_{35}NO_3$	15	0	0	M		No
324.2176	$C_{18}H_{29}NO_4$	371	0	2	M		No
340.2641	$C_{23}H_{33}NO$	139	1	0	M	Evocarpine	Yes
352.2514	$C_{17}H_{37}NO_4S$	3	1	0	M	Capnine	Yes
372.3112	$C_{21}H_{41}NO_4$	59	2	11	M		Yes
391.2849	$C_{24}H_{38}O_4$	368	115	7	E	Dioctylaphtalate	Yes
398.3625	$C_{24}H_{47}NO_3$	31	1	0	M		No
437.3408	$C_{30}H_{44}O_2$	65	1	0	M		No
Negative mode							
89.0244	$C_3H_6O_3$	31	8	34	M	Lactate	N.A.
105.0192	$C_3H_6O_4$	11	3	7	M	Glyceric acid	N.A.
112.9854	$C_4H_2O_4$	11	0	1	M		N.A.
119.0348	$C_4H_8O_4$	80	13	19	E		N.A.
135.0297	$C_4H_8O_5$	17	5	13	M	Erythronic acid	N.A.
149.0453	$C_5H_{10}O_5$	129	15	30	M	Ribose	N.A.
158.9908	$C_5H_4O_6$	5	0	1	M		N.A.
177.0558	$C_{10}H_{10}O_3$	641	17	22	M		N.A.
179.0558	$C_6H_{12}O_6$	221	24	45	E		N.A.

The hits found in METLIN database are shown

N.A. not analyzed

any conclusive database hits. Even if there were a few hits, such as for *m/z* 302.1934, none of them were published before in a biological context, which makes it interesting, but also challenging, for further studies.

Even at high resolution, mass spectrometric methods are not sufficient to positively identify unknown compounds, since isomers are not separated and the mass spectrum is limited to one-dimensional information. Therefore it is necessary to add at least one additional dimension to identify such molecules. In a study by Gamez et al., a new metabolite of valproic acid was found, and it was possible to identify this metabolite through the use of collision induced fragmentation (CID) in a high-resolution instrument (Gamez et al. 2011). This new metabolite could be identified since the signal was very intense and therefore the MS/MS approach was feasible. However, the signals of virtually all metabolites detected with our configuration were too weak and the mass selection of the linear trap was

Table 2 Compounds found by external contamination are listed which are clearly identified as related to the contamination

m/z	Elemental composition	Possible assigment	Detected offline
Chewing gum			
236.1149	$C_{12}H_{15}O_3N_2$		N.A.
264.1404	$C_{15}H_{21}NOS$		N.A.
264.1469	$C_{14}H_{219}O_3N_2$		N.A.
288.1495	$C_{19}H_{17}N_3$		N.A.
292.1781	$C_{16}H_{23}O_3N_2$		N.A.
297.1952	$C_{19}H_{24}N_2O$		N.A.
353.2599	$C_{23}H_{32}N_2O$		N.A.
Coffee			
80.0497	C_5H_5N	Pyridine	N.A.
Cigarette			
108.0811	C_7H_9N	Benzylazanium/*o*-toluidine/*N*-methyl-aniline	No
122.0968	$C_8H_{11}N$	Phenylethylamine/*N*-ethylaniline/xylidine/*N*,*N*-dimethylaniline	No
136.1124	$C_9H_{13}N$	Dextroamphetamine/2,4,5-trimethylaniline	No
148.1125	$C_{10}H_{13}N$		No
150.1281	$C_{10}H_{15}N$	*N*-Butylaniline/phentermine/2,6-diethylaniline	No
160.1125	$C_{11}H_{13}N$		No
163.1234	$C_{10}H_{14}N_2$	Nicotine	Yes
164.1438	$C_{11}H_{17}N$	3 Hexylpyridine	No
176.1438	$C_{12}H_{17}N$		No
184.1700	$C_{11}H_{21}NO$		No
188.1438	$C_{13}H_{17}N$		Yes
204.1751	$C_{14}H_{21}N$		No
207.1497	$C_{12}H_{18}N_2O$		No
221.1653	$C_{13}H_{20}N_2O$		No

The compound names refer to hits in METLIN

N.A. not analyzed

not sufficiently narrow (mass window was too broad) to perform MS/MS on them. As seen in Table 1, some of the signals observed differed in mass by less than 0.1 Da, which would mean that both would fragment together in a MS/MS experiment. Therefore, a different approach is needed to collect additional structural information for these compounds. Since there is no other method that is able to access compounds of low volatility in breath in real time, only an off-line method can deliver comparable data. In order to extend our initial strategy, it was necessary to combine it with an off-line method to allow the combination of our new method with classical metabolomics tools. Therefore Teflon tubes, filled with cotton were used and tested. The absorbed breath was desorbed using humidified air and directly measured in the ion funnel SESI interface. Using cotton, only a few compounds were successfully trapped, nevertheless it shows that the combination of real-time and off-line detection using SESI is possible. The result of this experiment is shown in Table 1 last column. Since absorbance materials are always selective, it is very helpful to make sure that the compound of interest is trapped successfully by using the same ionization method. Therefore, this intermediate step between real-time and off-line evaluation is quite useful. A compound that is trapped successfully, is easily analyzed further by the use of classical metabolic methods such as HPLC or UPLC–MS and finally identified by the use of authentic standards, if available (an example of how this can be done is shown in the supplementary information).

3.2 Discerning exogenous compounds in exhaled breath

Exogenous compounds in this context are understood as exhaled compounds that are not metabolites, but enter into the breath from elsewhere, usually from the surrounding room air. It important that further study designs take into account the influence of external contaminations on the exhaled breath measured, which is also influenced by the subject's living habits, as well ass the behavior and surroundings prior to the experiment. The influence of exogenous compounds on breath was tested after the volunteer chewed gum, drank coffee, or smoked a cigarette. Figure 1c–e shows the mass spectra of breath after consuming

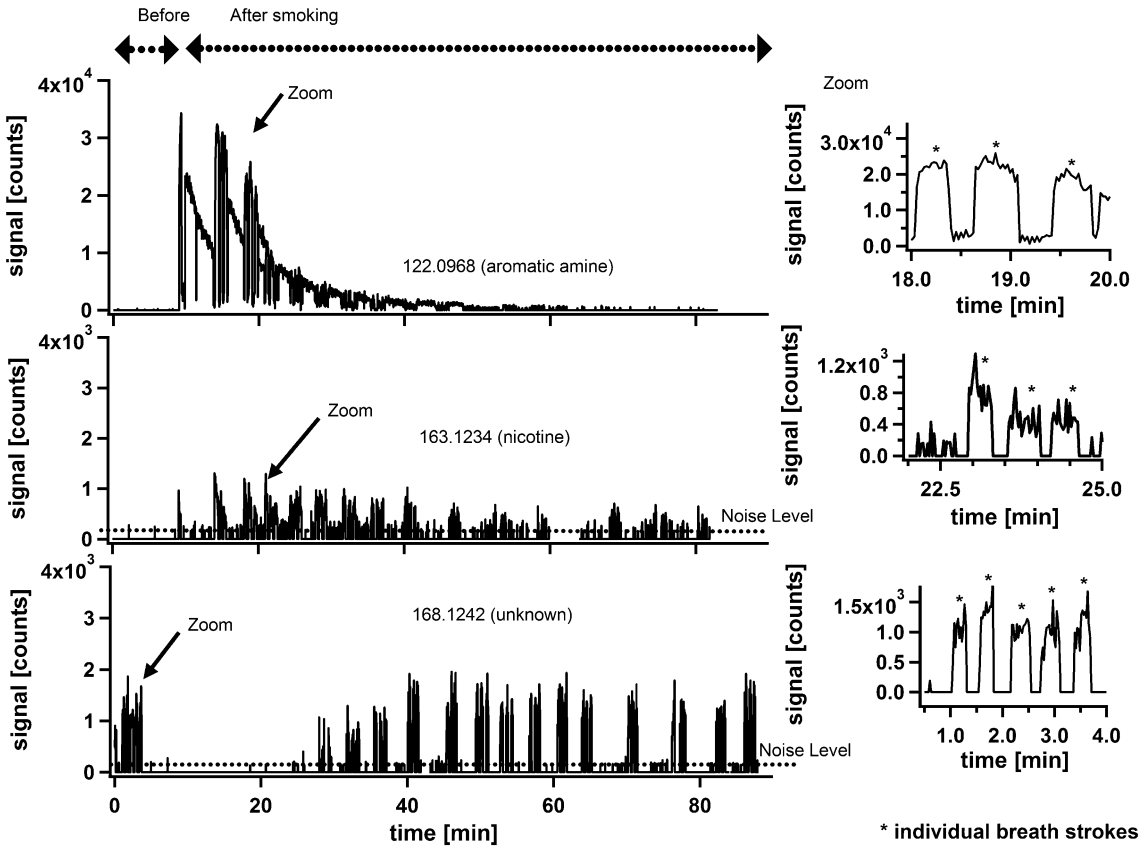

Fig. 4 The real-time extracted ion chromatograms from several *m/z* values after smoking is shown. The ion at *m/z* 122.0968 represents an aromatic amine (likely a mixture of isomers), which is directly exhaled after smoking and fades away fast. *M/z* 163.1234 is nicotine, which shows the expected kinetics. *M/z* 168.1242 is an unidentified signal, which has a pattern that suggests an exhaled metabolite. The disappearance of this signal right after smoking implies ion suppression, probably due to the high signal intensities of the tar compounds. This suppression could take place during ionization, although it is also possible that the high level of contamination in the respiratory system directly suppresses the release of this compound. The *insets* show zooms of the individual breath-strokes, which are not very well resolved in the overview graph. The *dashed line* in the extracted ion chromatogram indicates the noise level

these products. Table 2 shows the additional signals observed in exhaled breath after the consumption. Strong signals were observed after the subject smoked, which may originate from compounds in tar. These homologs have been previously reported for cigarette smoke (side stream as well as main stream) and are also found as environmental contaminants (Stabbert et al. 2003; Torikaiu et al. 2005). Nevertheless, to our knowledge, there is no study yet which proves that these compounds are directly exhaled after smoking. Moreover, we have already observed a similar mass spectral pattern in a study by APCI–MS in 2011 (Berchtold et al. 2011), but did not further identify the signals. Some of the exhaled compounds from cigarette smoke are plotted as time dependent signals in Fig. 4. The patterns observed indicate that they originate from within the respiratory system (at least the mouth cavity). The signal $m/z = 122.0968$ (Fig. 4, top) disappeared quite fast. Nicotine ($m/z = 163.1234$), as the main active compound and the only clearly identified hit in all the databases, was very clearly observed. Free nicotine in the blood flow is potentially exhaled, which fits the known pharmacokinetic behavior (Kardani et al. 2010). It appears that the other signals, which may refer to metabolites, are partially suppressed after smoking. As soon as the main exogenous compounds disappeared, these signals re-appeared again. For example, the signal at *m/z* 168.1242 demonstrates this behavior (Fig. 4 bottom). This signal is an unknown metabolite. For chewing gum these effects were even stronger. There were a few strong signals that suppressed all other potential metabolite peaks (Fig. 1c). In the experiment with chewing gum, none of the signals were identified: there was no hit in any of the databases that could be matched with an aroma or other well-known compound, thus no assignment is proposed. Only sum formulae and isotopic patterns were confirmed and are listed in Table 2. Coffee does not influence the exhaled pattern as much as chewing gum or smoke. In the experiment with coffee, pyridine was identified, a well-known compound that can be expected in the breath of people who drink coffee (Meier et al. 2012b). All these results are

summarized in Table 2. For cigarette smoke, the use of cotton tubes for offline detection was tested as well. Nicotine was detected in offline-mode SESI, whereas the tar compounds were not detected. Nicotine as well as some of the tar compounds were also detected by UPLC–MS after extracting the tube with methanol (see supplementary).

4 Concluding remarks

The analysis of the breath-by-breath pattern enables to distinguish between exhaled exogenous compounds, exhaled metabolites, and background signals. Signals that are identified as endogenous are further assigned to possible compounds using database matching (Fig. 3, supplementary Fig. S2–S4). To identify signals that do not provide a conclusive hit, the combination with off-line analysis will become important. The use of cotton-filled Teflon tubes shows that in principle it is possible to use off-line SESI ionization (also for studies where patient are not able to move in front of the instrument) as well as classical HPLC–MS or UPLC–MS to identify unknowns. For instance, the compounds listed in the table should be rigorously identified by employing orthogonal methods (such as UPLC–MS) and by the use of authentic reference standards if available.

Contamination by large concentrations of exogenous compounds was found to suppress the ionization of endogenous exhaled compounds. For example, we found that smoking even 90 min prior to an experiment still resulted in suppression effects. For the successful analysis of endogenous compounds in exhaled breath by SESI, it must thus be assured that a subject does not smoke, chew gum, use mouth wash, etc. for about 2 h before the measurements are carried out.

Finally, the strategy presented here might help to find and identify many new metabolites in exhaled breath. The whole process can be automated, which allows systematic screening for unknown metabolites. This proposed strategy is independent of the ionization method or instrument used as, long as the time resolution, mass resolution and sensitivity are sufficient. Real-time breath analysis is thus a promising new method to perform metabolomics studies.

Acknowledgments We thank the Swiss National Science Foundation for financial support (grant no. K-23K1-122264). Pablo Martinez-Lozano Sinues is acknowledged for critical review of the manuscript. Special thanks goes to the team from the Functional Genomics Center at the University Zürich for the access to their instrument.

References

Abbanat, D. R., Godchaux, W., Polychroniou, G., & Leadbetter, E. R. (1985). Biosynthesis of a sulfonolipid in gliding bacteria. *Biochemical and Biophysical Research Communications, 130*, 873–878. doi:10.1016/0006-291x(85)90497-8.

Alberici, R. M., et al. (2010). Ambient mass spectrometry: Bringing MS into the "real world". *Analytical and Bioanalytical Chemistry, 398*, 265–294. doi:10.1007/s00216-010-3808-3.

Beck, O., Sandqvist, S., Eriksen, P., Franck, J., & Palmskog, G. (2011). Determination of methadone in exhaled breath condensate by liquid chromatography-tandem mass spectrometry. *Journal of Analytical Toxicology, 35*, 129–133.

Benoit, F. M., Davidson, W. R., Lovett, A. M., Nacson, S., & Ngo, A. (1983). Breath analysis by atmospheric-pressure ionization mass-spectrometry. *Analytical Chemistry, 55*, 805–807.

Berchtold, C., Meier, L., & Zenobi, R. (2011). Evaluation of extractive electrospray ionization and atmospheric pressure chemical ionization for the detection of narcotics in breath. *International Journal of Mass Spectrometry, 299*, 145–150. doi:10.1016/j.ijms.2010.10.011.

Biasioli, F., Yeretzian, C., Maerk, T. D., Dewulf, J., & Van Langenhove, H. (2011). Direct-injection mass spectrometry adds the time dimension to (B)VOC analysis. *Trac-Trends in Analytical Chemistry, 30*, 1003–1017. doi:10.1016/j.trac.2011.04.005.

Buszewski, B., Kesy, M., Ligor, T., & Amann, A. (2007). Human exhaled air analytics: Biomarkers of diseases. *Biomedical Chromatography, 21*, 553–566. doi:10.1002/bmc.835.

Carpagnano, G. E., Barnes, P. J., Francis, J., Wilson, N., Bush, A., & Kharitonov, S. A. (2004). Breath condensate pH in children with cystic fibrosis and asthma: A new noninvasive marker of airway inflammation? *Chest, 125*, 2005–2010. doi:10.1378/chest.125.6.2005.

Cathcart, M. P., Love, S., & Hughes, K. J. (2012). The application of exhaled breath gas and exhaled breath condensate analysis in the investigation of the lower respiratory tract in veterinary medicine: A review. *Veterinary Journal, 191*, 282–291. doi:10.1016/j.tvjl.2011.08.016.

Cazzola, M., & Novelli, G. (2010). Biomarkers in COPD. *Pulmonary Pharmacology and Therapeutics, 23*, 493–500. doi:10.1016/j.pupt.2010.05.001.

Chen, H. W., Gamez, G., & Zenobi, R. (2009a). What can we learn from ambient ionization techniques? *Journal of the American Society for Mass Spectrometry, 20*, 1947–1963. doi:10.1016/j.jasms.2009.07.025.

Chen, H., Hu, B., Hu, Y., Huan, Y., Zhou, Z., & Qiaoc, X. (2009b). Neutral desorption using a sealed enclosure to sample explosives on human skin for rapid detection by EESI–MS. *Journal of the American Society for Mass Spectrometry, 20*, 719–722.

Costa, C., Bucca, C., Bergallo, M., Solidoro, P., Rolla, G., & Cavallo, R. (2011). Unsuitability of exhaled breath condensate for the detection of Herpesviruses DNA in the respiratory tract. *Journal of Virological Methods, 173*, 384–386. doi:10.1016/j.jviromet.2011.02.004.

Dillon, L. A., Stone, V. N., Croasdell, L. A., Fielden, P. R., Goddard, N. J., & Thomas, C. L. P. (2010). Optimisation of secondary electrospray ionisation (SESI) for the trace determination of gas-phase volatile organic compounds. *Analyst, 135*, 306–314. doi:10.1039/b918899a.

Dwyer, T. M. (2004). Sampling airway surface liquid: Non-volatiles in the exhaled breath condensate. *Lung, 182*, 241–250. doi:10.1007/s00408-004-2506-3.

Effros, R. M., Dunning, M. B., Biller, J., & Shaker, R. (2004). The promise and perils of exhaled breath condensates. *American Journal of Physiology-Lung Cellular and Molecular Physiology, 287*, L1073–L1080. doi:10.1152/ajplung.00069.2004.

Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Molecular Biology, 48*, 155–171. doi:10.1023/a:1013713905833.

Gamez, G., et al. (2011). Real-time, in vivo monitoring and pharmacokinetics of valproic acid via a novel biomarker in

exhaled breath. *Chemical Communications, 47*, 4884–4886. doi:10.1039/c1cc10343a.

Huang, M.-Z., Cheng, S.-C., Cho, Y.-T., & Shiea, J. (2011). Ambient ionization mass spectrometry: A tutorial. *Analytica Chimica Acta, 702*, 1–15. doi:10.1016/j.aca.2011.06.017.

Hunt, J. (2002). Exhaled breath condensate: An evolving tool for noninvasive evaluation of lung disease. *Journal of Allergy and Clinical Immunology, 110*, 28–34. doi:10.1067/mai.2002.124966.

Johnson, G. R., & Morawska, L. (2009). The mechanism of breath aerosol formation. *Journal of Aerosol Medicine and Pulmonary Drug Delivery, 22*, 229–237. doi:10.1089/jamp.2008.0720.

Kardani, F., Daneshfar, A., & Sahrai, R. (2010). Determination of nicotine, anabasine, and cotinine in urine and saliva samples using single-drop microextraction. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 878*, 2857–2862. doi:10.1016/j.jchromb.2010.08.041.

Kim, K. H., Jahan, S. A., & Kabir, E. (2012). A review of breath analysis for diagnosis of human health. *Trac-Trends in Analytical Chemistry, 33*, 1–8. doi:10.1016/j.trac.2011.09.013.

Kind, T., & Fiehn, O. (2007). Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics, 8*, 105. doi:10.1186/1471-2105-8-105.

Marek, E. M., Volke, J., Hawener, I., Platen, P., Muckenhoff, K., & Marek, W. (2010). Measurements of lactate in exhaled breath condensate at rest and after maximal exercise in young and healthy subjects. *Journal of Breath Research, 4*, 017105. doi:10.1088/1752-7155/4/1/017105.

Martinez-Lozano, P., & Fernandez de la Mora, J. (2009). On-line detection of human skin vapors. *Journal of the American Society for Mass Spectrometry, 20*, 1060–1063. doi:10.1016/j.jasms.2009.01.012.

Martinez-Lozano, P., & Fernandezde la Mora, J. (2008). Direct analysis of fatty acid vapors in breath by electrospray ionization and atmospheric pressure ionization-mass spectrometry. *Analytical Chemistry, 80*, 8210–8215. doi:10.1021/ac801185e.

Martinez-Lozano, P., Zingaro, L., Finiguerra, A., & Cristoni, S. (2011). Secondary electrospray ionization-mass spectrometry: Breath study on a control group. *Journal of Breath Research, 5*, 016002. doi:10.1088/1752-7155/5/1/016002.

Mazzone, P. J., et al. (2012). Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. *Journal of Thoracic Oncology, 7*, 137–142. doi:10.1097/JTO.0b013e318233d80f.

Meier, L., Berchtold, C., Schmid, S., & Zenobi, R. (2012a). Extractive electrospray ionization mass spectrometry enhanced sensitivity using an ion funnel. *Analytical Chemistry, 84*, 2076–2080. doi:10.1021/ac203022x.

Meier, L., Berchtold, C., Schmid, S., & Zenobi, R. (2012b). High mass resolution breath analysis using secondary electrospray ionization mass spectrometry assisted by an ion funnel. *Journal of Mass Spectrometry, 47*, 1571–1575. doi:10.1002/jms.3118.

Meier, L., Berchtold, C., Schmid, S., & Zenobi, R. (2012c). Sensitive detection of drug vapors using an ion funnel interface for secondary electrospray ionization mass spectrometry. *Journal of Mass Spectrometry, 47*, 555–559. doi:10.1002/jms.2982.

Meier, L., Schmid, S., Berchtold, C., & Zenobi, R. (2011). Contribution of liquid-phase and gas-phase ionization in extractive electrospray ionization mass spectrometry of primary amines. *European Journal of Mass Spectrometry, 17*, 345–351. doi:10.1255/ejms.1146.

Oberacher, M., Pohl, D., Vavricka, S. R., Fried, M., & Tutuian, R. (2011). Diagnosing lactase deficiency in three breaths. *European Journal of Clinical Nutrition, 65*, 614–618. doi:10.1038/ejcn.2010.287.

Phillips, M., et al. (1999). Volatile organic compounds in breath as markers of lung cancer: A cross-sectional study. *Lancet, 353*, 1930–1933. doi:10.1016/s0140-6736(98)07552-7.

Risby, T. H. (2008). Critical issues for breath analysis. *Journal of Breath Research,*. doi:10.1088/1752-7155/2/3/030302.

Risby, T. H., & Solga, S. F. (2006). Current status of clinical breath analysis. *Applied Physics B-Lasers and Optics, 85*, 421–426. doi:10.1007/s00340-006-2280-4.

Schwoebel, H., et al. (2011). Phase-resolved real-time breath analysis during exercise by means of smart processing of PTR-MS data. *Analytical and Bioanalytical Chemistry, 401*, 2079–2091. doi:10.1007/s00216-011-5173-2.

Shahid, S. K., Kharitonov, S. A., Wilson, N. M., Bush, A., & Barnes, P. J. (2005). Exhaled 8-isoprostane in childhood asthma. *Respiratory Research, 6*, 79. doi:10.1186/1465-9921-6-79.

Sinues, P. M.-L., Criado, E., & Vidal, G. (2012). Mechanistic study on the ionization of trace gases by an electrospray plume. *International Journal of Mass Spectrometry, 313*, 21–29. doi:10.1016/j.ijms.2011.12.010.

Smith, D., & Španel, P. (2011). Direct, rapid quantitative analyses of BVOCs using SIFT-MS and PTR-MS obviating sample collection. *Trac-Trends in Analytical Chemistry, 30*, 945–959. doi:10.1016/j.trac.2011.05.001.

Smith, D., Wang, T., Pysanenko, A., & Španel, P. (2008). A selected ion flow tube mass spectrometry study of ammonia in mouth- and nose-exhaled breath and in the oral cavity. *Rapid Communications in Mass Spectrometry, 22*, 783–789. doi:10.1002/rcm.3434.

Španel, P., & Smith, D. (2011). Progress in Sift-MS: Breath analysis and other applications. *Mass Spectrometry Reviews, 30*, 236–267. doi:10.1002/mas.20303.

Stabbert, R., Schafer, K. H., Biefel, C., & Rustemeier, K. (2003). Analysis of aromatic amines in cigarette smoke. *Rapid Communications in Mass Spectrometry, 17*, 2125–2132. doi:10.1002/rcm.1161.

Thekedar, B., Szymczak, W., Hollriegl, V., Hoeschen, C., & Oeh, U. (2009). Investigations on the variability of breath gas sampling using PTR-MS. *Journal of Breath Research, 3*(2), 027007. doi:10.1088/1752-7155/3/2/027007.

Torikaiu, K., Uwano, Y., Nakamori, T., Tarora, W., & Takahashi, H. (2005). Study on tobacco components involved in the pyrolytic generation of selected smoke constituents. *Food and Chemical Toxicology, 43*, 559–568. doi:10.1016/j.fct.2004.12.011.

Van den Velde, S., Nevens, F., Van Hee, P., van Steenberghe, D., & Quirynen, M. (2008). GC–MS analysis of breath odor compounds in liver patients. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 875*, 344–348. doi:10.1016/j.jchromb.2008.08.031.

Wardlaw, A. J., Hay, H., Cromwell, O., Collins, J. V., & Kay, A. B. (1989). Leutokrines, LTC4 and LTB4, in bronchoalveolar lavage in bronchial-asthma and other respiratory-disease. *Journal of Allergy and Clinical Immunology, 84*, 19–26. doi:10.1016/0091-6749(89)90173-5.

White, R. H. (1984). Biosynthesis of the sulfonolipid 2-amino-3-hydroxy-15-methylhexadecane-1-sulfonic acid in the gliding bacterium Cytophaga-Johnsonae. *Journal of Bacteriology, 159*, 42–46.