5

BREATHOMICS IN LUNG DISEASE

Tamara Paff, MD; Marc P.C. van der Schee; Paul Brinkman; Wim M.C. van Aalderen; Eric G. Haarman; Peter J. Sterk

Chest (2015); 147(1): 224-231
ABSTRACT

Volatile organic compounds (VOCs) are produced by virtually all metabolic processes of the body. As such, they have potential to serve as noninvasive metabolic biomarkers. Since exhaled VOCs are either derived from the respiratory tract itself or have passed the lungs from the circulation, they are candidate biomarkers in the diagnosis and monitoring of pulmonary diseases in particular. Good examples of the possibilities of exhaled volatiles in pulmonary medicine are provided by the potential use of VOCs to discriminate between patients with lung cancer and healthy control subjects and to noninvasively diagnose infectious diseases and the association between VOCs and markers of disease activity that has been established in obstructive lung diseases. Several steps are, however, required prior to implementation of breath-based diagnostics in daily clinical practice. First, VOCs should be studied in the intention-to-diagnose population, because biomarkers are likely to be affected by multiple (comorbid) conditions. Second, breath collection and analysis procedures need to be standardized to allow pooling of data. Finally, apart from probabilistic analysis for diagnostic purposes, detailed examination of the nature of volatile biomarkers not only will improve our understanding of the pathophysiologic origins of these markers and the nature of potential confounders but also can enable the development of sensors that exhibit maximum sensitivity and specificity toward specific applications. By adhering to such an approach, exhaled biomarkers can be validated in the diagnosis, monitoring, and treatment of patients in pulmonary medicine and contribute to the development of personalized medicine.
DIAGNOSIS BY OLFACTION

Aristotle taught doctors to use their sense of smell. Liver disease was identified by a fecal breath and stale beer smell suggested the presence of TB [1]. A more recent anecdote comes from Oscar the cat who was able to smell impending death at a Rhode Island nursing home [2]. Such observations stimulated research on the use of Volatile Organic Compounds (VOCs), the main molecular substrate underlying mammalian olfaction, as biomarkers. This review focuses on the validation of VOCs as novel biomarkers for pulmonary medicine. The rationale behind VOCs, validation of current results and their potential future use and the obstacles that need to be overcome are addressed.

VOCs reflect metabolism in humans

VOCs are gaseous organic molecules that are emitted from the fluid phase because they are highly volatile. Human VOCs are released from skin, with feces, urine and breath and are derived from many metabolic pathways. Since cellular metabolism is altered by disease, the resulting change in VOCs may serve as biomarkers for particular pathophysiological conditions. In pulmonary medicine breath is of special interest because of its intensive contact with the respiratory tract. The rate at which VOCs are exhaled is the net effect of several interacting (bio)chemical processes; intracellular and extracellular degradation, solubility of the compound in extracellular fluid, fat, and blood, the affinity with extracellular matrix and carrier proteins, the concentration gradient with the alveolar and bronchial air, the vapor pressure and alveolar ventilation [3]. This results in a chemical equilibrium of a given compound between breath, blood, and fat that can be described by that substance’s physiochemical partition constant [4].

Nobel laureate Pauling and colleagues were the first to isolate > 200 organic volatiles from a single breath sample [5]. To date, several thousands of individual VOCs have been identified, generally occurring in the parts per million/parts per billion range. This multitude of markers likely represents the complexity of human biology more accurately than do isolated biomarkers. “Breathomics” may, therefore, have potential for noninvasive diagnosis and monitoring of disease, which closely fits a personalized medicine approach [6]. As the interest in this field is growing rapidly, this may expedite the development of VOCs in lung disease.

Exhaled VOCs may be of local, systemic or exogenous origins (figure 1). Locally produced compounds diffuse directly into alveoli or the airway lumen along the respiratory tract. Volatiles of systemic origins are derived from the circulation after originating from metabolic processes elsewhere and dissolving into the blood. Therefore, even nonpulmonary diseases contribute to exhaled VOCs, which has successfully been used in the assessment of
nonpulmonary malignancies [7]. Exogenous VOCs can be inhaled or absorbed through the skin. They primarily originate from nonhuman sources and exist in three categories. First, VOCs introduce undesired noise by being inspired and expired without interaction with the body. A second group of exogenous VOCs does interact with human tissue and can be stored inside the body for extensive periods of time [3]. The latter volatiles can, therefore, serve as potential biomarkers for environmental exposures and buildup of toxins, such as cigarette smoke carcinogen N-Nitrosomine [8]. The third group of exogenous VOCs is of (resident) microbial origin (predominantly bacteria, but also fungi and viruses), making them of specific interest when identifying infectious diseases or diseases linked to changes in microbiome [9]. Since VOCs reflect this broad range of pathophysiological processes they have the potential to meet many of the criteria for an ideal biomarker (table 1).

Figure 1. Systemic diagram of airways and alveolus with the various sources of exhaled VOCs. Exhaled VOCs can originate from (1) endogenous VOCs from conducting airways and alveoli, (2) exogenous VOCs inhaled and subsequently exhaled or originating from the resident microbiome, or (3) systemic VOCs generated elsewhere in the body and transported to the lungs via the blood circulation. VOC, volatile organic compound.

**VOC sampling and handling**

The potential to noninvasively sample breath VOCs is core to the attractiveness of these biomarkers. Depending on the specific application, a variety of techniques to collect exhaled breath are currently available. Progress is made in investigating the influence of these various techniques on exhaled VOCs to provide suggestions for standardization [11–13]. Fortunately, the lack of international guidelines for sampling of VOCs is currently being
addressed by task forces. An overview of the core factors that are relevant with respect to collecting and handling exhaled breath are provided in table 2.

Table 1. Criteria for an ideal biomarker

<table>
<thead>
<tr>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sensitive and specific for the diagnosis of the disease process</td>
</tr>
<tr>
<td>2. Reflect or be a very clear surrogate of the pathophysiological mechanism</td>
</tr>
<tr>
<td>3. Be stable and only vary with events known to relate to disease progression</td>
</tr>
<tr>
<td>4. Predict early-stage disease development</td>
</tr>
<tr>
<td>5. Predict disease progression</td>
</tr>
<tr>
<td>6. Be responsive to interventions known to be effective</td>
</tr>
</tbody>
</table>

Adapted from Stockley et al. [10]

Table 2. Key factors in breath collection and handling. [13]

<table>
<thead>
<tr>
<th>Key factors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air sampling</td>
<td>Direct sampling</td>
</tr>
<tr>
<td></td>
<td>Reusable collection bag with thorough cleaning</td>
</tr>
<tr>
<td></td>
<td>Disposable collection bag</td>
</tr>
<tr>
<td>Air collection device</td>
<td>VOCs derived from collection device</td>
</tr>
<tr>
<td></td>
<td>Disposable collection device</td>
</tr>
<tr>
<td></td>
<td>Reusable collection device (cleaning agent VOCs)</td>
</tr>
<tr>
<td>Air collection method</td>
<td>Nasal/Oral sampling</td>
</tr>
<tr>
<td></td>
<td>Tidal breathing vs forced exhalation</td>
</tr>
<tr>
<td></td>
<td>Exhalation after breath hold</td>
</tr>
<tr>
<td></td>
<td>Forced exhalation</td>
</tr>
<tr>
<td></td>
<td>Flow</td>
</tr>
<tr>
<td>Environmental influences</td>
<td>Baseline samples of environmental air</td>
</tr>
<tr>
<td></td>
<td>Wash-out period by inspiratory VOC filter</td>
</tr>
<tr>
<td>Storage</td>
<td>Direct analysis</td>
</tr>
<tr>
<td></td>
<td>Storage on sorbent tubes for prolonged stabilization of VOCs</td>
</tr>
<tr>
<td></td>
<td>Storage conditions</td>
</tr>
</tbody>
</table>

VOC, volatile organic compound.

VOC analysis techniques

The concepts of the analysis of volatile biomarkers can be understood best by discussing the two ends of the spectrum of available techniques (figure 2). On one hand, these encompass chemical analytical techniques allowing identification of specific compounds. At the other end of the spectrum are pattern-recognition-based techniques allowing probabilistic discrimination of biomarker profiles. It is, however, important to realize that many techniques share features with both of these basic concepts.
Figure 2. Analytical techniques for exhaled VOCs.

Two ends of the spectrum of VOC analysis techniques are depicted. Left: Exhaled VOCs are a complex mixture of different compounds (depicted by triangle, circle, and square symbols). The top of the image displays GC and MS as prototype chemical analytical technique allowing identification of individual components by comparing the mass to charge ratios of individual compounds with a previously established reference library. The bottom of the image displays the e-Nose as a prototype pattern recognition-based technology. e-Nose sensor are cross-reactive to multiple compounds in the VOC mixture, resulting in a pattern of sensor responses driven by, and characteristic of, the composite mixture of VOCs. These patterns can be compared with those previously encountered by pattern recognition algorithms to allow classification of individual cases. e-Nose, electronic nose; GC, gas chromatography; MS, mass spectrometry.

Chemical analytical techniques

Gas chromatography (GC) coupled to mass spectrometry (MS) is the current gold standard for the identification of individual chemical compounds [14]. GC-MS allows identification of individual compounds in a breath sample on the basis of their elution time from a capillary column and their mass to charge ratio. Clinical implementation of GC-MS is, however, relatively complex because of the requirement for highly trained personnel, the laborious analysis, and, therefore, the costly application. Recent years have, however, seen major advances in these techniques, such as the development of selected ion flow tube mass spectrometry, which is well suited for breath analysis because it allows real-time quantitative analysis of VOCs down to the parts per trillion range and is technologically less demanding [15]. Furthermore, miniaturization of devices such as ion-mobility spectrometry may allow low-cost, on-site detection of specific compounds [16].
**Pattern recognition-based sensors**

At the other end of the spectrum are multiple-sensor devices resembling mammalian olfaction, therefore dubbed electronic noses (eNoses) [14]. VOCs competitively interact with cross-reactive sensors, allowing multiple VOCs to interact with the same sensor based on their affinity for both the sensor and its substrate. Likewise, multiple sensors interact with the same volatile. This is comparable to the powerful mammalian olfactory system and results in a pattern of firing sensors that is driven by, and characteristic of, the composite mixture of VOCs [17]. These patterns can be compared with those previously encountered by pattern recognition algorithms. Notably, this technique does not identify individual compounds but rather provides probabilistic recognition, which forms the basis of assessing diagnostic accuracy. eNose technology is relatively cheap, easy to use, and provides on-site results holding promise for its use as a point-of-care tool if properly validated and standardized [14].

**CURRENT STATE OF VOC RESEARCH AND ITS FUTURE POTENTIAL**

The key opportunities and critical challenges for the application of VOCs in clinic are illustrated by the collective experience of their use in three diseases: lung cancer, respiratory infections and obstructive lung disease.

**Lung cancer: breath volatiles in screening**

The noninvasive nature of VOC analysis makes it well suited for population-based screening for lung cancer. Studies to date using both GC-MS and eNose have shown a sensitivity and specificity ranging from 50-100% and 80-100%, respectively, for discrimination of healthy control subjects and patients with lung cancer in vivo [7, 18, 19]. Both research on in vitro tumor cell lines and in vivo breath analysis has aided in the identification of compounds driving these signals, which predominantly are hydrocarbons, alcohols, aldehydes, ketones, esters, nitriles and aromatic compounds [4, 20, 21]. Some of these markers have shown to decrease after resection of non-small cell lung carcinoma, providing further evidence they are linked to tumor presence [4].

It is important to realize that the exact origin of most of these VOCs is unknown. This means there is insufficient evidence that volatiles that discriminate between “gold-standard” healthy control subjects and carefully selected patients with lung cancer have sufficient accuracy to establish lung cancer in an intention-to-diagnose population. The latter can be characterized by comorbid conditions, such as smoking and COPD, potentially hampering the specificity of identified volatiles for lung cancer. This may be of even greater importance with respect to pattern-based techniques that do not identify individual compounds. It is, therefore, essential to apply VOC analysis with an a priori objective regarding the confirmation (high...
specificity) and/or the exclusion (high sensitivity) of disease in the population in which the clinical application is intended [22]. Two small studies applying such principles showed that patients with malignant pleural mesothelioma could be discriminated from patients with benign asbestos-related disease and subjects with asbestos exposure with a reasonable accuracy of 80.8% [23, 24].

An appealing alternative to identify potential volatile biomarkers is that employed by the team of Hoassam Haick [25]. They identified mutation-specific volatiles by GC-MS analysis of headspace VOCs of tumor cell lineages with a known oncogenic K-Ras or epidermal growth factor receptor (EGFR) mutation. These were used to create a nanosensor array discriminating the cell lines with an accuracy of 84% to 96%. Although the specificity of these tailor-made sensors has to be determined in vivo, the concept of building sensors for disease-specific volatiles is appealing because it can boost the accuracy of breath-based tests and simultaneously improve knowledge of the origins of these VOCs.

By combining appropriately designed studies with technical expertise, the potential for breath-based diagnostics in lung cancer becomes realistic. For screening purposes, this requires excluding disease by maximizing sensitivity and negative predictive value enabling selection of patients for more elaborate testing. This will require large-scale prospective cohort studies to identify the suitable screening population, frequency of screening, and subsequent diagnostic approach, which will require many years of research. Developing such a breath-based low-cost, noninvasive screening test for an at-risk population may enable early detection and reduction of health-care costs with a better cost-effectiveness and risk profile than current screening programs.

Respiratory infections; point of care diagnostics

With respect to respiratory infections, the primary value for VOC-based analysis lies in early detection of infection both in vitro and in vivo and in monitoring of the host response [9, 26]. For TB, the development of a volatile biomarker test has many potential advantages over current diagnostic techniques because such a test could be low-cost and easy to implement in low-income countries [27]. In vitro analysis of the headspace of cultures revealed several TB-associated VOCs [28, 29]. Probabilistic analysis by eNose reached promising overall specificity and sensitivity of 91% and 89% respectively [30], although the accuracy in smear negative samples was only 69% [31]. The diagnostic potential of these in vitro volatiles can subsequently be assessed in vivo under the (unlikely) assumption that in vivo and in vitro metabolism of Mycobacterium tuberculosis is identical and that pathogen volatiles are not altered by the host. In fact, some studies did reproduce volatiles similar to those identified in cultures in the breath of subjects [32], whereas others did not [33]. By combining both host
response and pathogen derived VOCs, such as in vivo studies reached accuracies of 79% till 85%, albeit not all in the intention-to-diagnose population [32–34].

Besides TB, volatile biomarkers have potential in diseases such as ventilator-associated pneumonia (VAP) [35] and cystic fibrosis (CF) [36]. Studying infections in CF, however, is exemplary of the extra level of complexity inherent to such studies because VOCs may relate to both the primary disease process and the infection. The presence of multiple simultaneous infections and nonspecificity of biomarkers to a single infection further complicates such analysis [37, 38]. The first proof-of-concept studies have shown patients with CF with and without Pseudomonas infection can be discriminated by a panel of VOCs [36]. Furthermore, children with and without exacerbations could be discriminated based on their VOC profile [39]. If detailed understanding of both host response and pathogen VOCs becomes available, this may ultimately change the way we diagnose infectious disease and monitor treatment response. Prior to effectuating such a promise, research will need to focus on combining in vitro and in vivo studies to identify key markers for specific pathogens, after which these can be validated in studies emulating clinical practice.

**Obstructive lung diseases: phenotyping and monitoring of individual patients**

The value of VOCs in obstructive lung diseases lies primarily in the monitoring and phenotyping of disease. The first proof-of-concept studies found 86% to 100% accuracy in discriminating healthy control subjects and gold standard patients with asthma [40, 41]. Fens and colleagues were able to discriminate patients with asthma and COPD by eNose and confirmed this discrimination by external validation (sensitivity 85%; specificity 90%) [42, 43]. Furthermore, a small study in children identified eight candidate asthma volatiles by GC-MS [44].

The largest unmet need in asthma and COPD is stratification of patients for antiinflammatory therapy with inhaled steroids or new biologicals, based on the inflammatory profile in the airways [45]. Several independent studies already linked VOC (profiles) in asthma and COPD to the inflammatory cell type in sputum and disease activity [46–48]. Indeed, steroid responsiveness in patients with asthma was shown to be predicted by eNose even with greater accuracy than by sputum eosinophils or exhaled nitric oxide [46]. Furthermore, patients with asthma exhaled increased levels of pentane during an exacerbation, decreasing to the level of control subjects when the exacerbation subsided [49]. In line with this, recent data suggest that VOCs allow discrimination of exacerbations in childhood asthma [50]. Especially in such heterogeneous diseases, individual daily monitoring and tailoring of therapy may empower patients to regain control of their disease and its therapy [51]. With respect to obstructive lung diseases, large multicenter trials such as U-BIOPRED (www.ubiopred.eu) validating these biomarkers are currently underway. These aim to provide evidence to determine whether and how clinical implementation of these techniques is beneficial to patient care.
Challenges and future directions

The potential benefits of VOC biomarkers are its noninvasiveness, speed, low-costs, and applicability in low-income countries. Much work, however, is needed before VOC-based diagnostic tools meet the criteria in table 1 and can be implemented, because current progress can only be classified as phase 2 to 4 on the 10-step technology readiness assessment scale [52].

A key issue with volatile biomarkers identified to date is the relative absence of independently reproduced biomarkers undermining the reliability of identified biomarkers. Part of these differences in established markers is likely due to the current lack of exchangeable collection techniques, devices and, data-analysis protocols. A European Respiratory Society task force is currently underway to provide guidelines for standardized methodology to maximize the compatibility of different datasets and minimize false discovery rate through stringent statistical approaches [53]. This furthermore requires a stepwise development and validation of omics-based biomarkers in clinical diagnosis and management, for which concrete criteria are available [54].

A second source of variation between breathomics studies are potential confounders. Indeed, factors known to influence VOCs include age, sex, pregnancy, medication, diet, and smoking, among many others [55–58]. In fact, even very mild exertion can change exhaled volatiles for hours [59, 60]. This knowledge is important for the interpretation of identified VOCs, but strict avoidance of all potential confounders is likely not worth pursuing. First, there is no evidence that dictating diet and minimizing exertion results in a more reproducible detection of biomarkers. Second, development of more specific and reliable sensor technology will reduce the effect of confounding VOCs on detecting target molecules. Foremost, however, extensive constraints on measurement protocols would render the technique impracticable, whereas its ease of use is central to its attractiveness.

A further constraint on the interpretation of established volatile markers is the fact that many of the predominant volatiles are related to a multitude of clinical conditions [61]. It is, therefore, unlikely that a disease can be characterized by a single biomarker. Fortunately, the multitude of VOCs enables the use of volatile biomarker profiles that may more robustly identify a clinical condition in the presence of confounding volatiles than a single biomarker [6]. This was illustrated by a study showing that prediction of steroid responsiveness in asthma was better achieved by the composite signal of exhaled VOCs measured by an electronic nose as compared with single biomarkers, such as sputum eosinophil count or exhaled nitric oxide [41].
With respect to sensor technology, the major advances can be expected from increased sensitivity and the development of sensors tailored toward detection of particular (classes of) VOCs and thereby specific conditions. Increasing sensor sensitivity is important, because the concentrations of some relevant VOCs have shown to be below the detection limit for certain sensors. Progress in this field is illustrated by promising developments lowering sensor detection limit and increasing selectivity [62]. This can either be achieved by selection of preassembled sensors through trial and error, potentially even without detailed knowledge of the target volatiles, or by building customized sensors that allow key-and-lock identification of disease-specific volatiles. Unfortunately, the relative lack of detailed knowledge on the core volatile markers in pulmonary disease has hampered the progress in sensor development relative to, for instance, acetone breath analysis in diabetes. Progress in this area can be reached by pooling of datasets and combining metabolomic measurements in tissue, blood, and breath. Furthermore, blocking specific enzymatic pathways and studying animal models may help to delineate exact pathways of VOC metabolism [63].

The adoption of breathomics in medical research and clinical practice would be greatly accelerated by the development of a so-called breath cloud, wherein exhaled breath VOC profiles could be stored alongside with anonymized clinical disease characteristics. This would allow real-time comparison of VOC patterns to provide an accurate diagnosis and phenotyping and will allow every subsequent measurement to contribute to the diagnostic algorithm, improving its accuracy while it is implemented in practice. This would furthermore allow a detailed comparison of the accuracy of VOC biomarkers with established biomarkers for that disease, which is essential to assess the value of such tests in clinical practice.

In conclusion, based on the current proof-of-concept data, clinical application of exhaled breath-based diagnostics and phenotyping is warranted, provided that the required (external) validation steps have adequately been taken [21]. To that end, breathomics needs to overcome key challenges such as focusing on the intention-to-diagnose population, using correct and standardized methodology, optimizing negative and/or positive predictive values, improving knowledge of VOC origins and developing tailored sensors. These milestones are all within reach when the field collectively follows the international guidelines on testing diagnostic accuracy and validation of composite biomarkers [22, 54].
REFERENCES


