Breathomics in pulmonary disease
Fens, N.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Perspective: Breathomics in medicine

Niki Fens, Marc P. van der Schee and Peter J. Sterk
ABSTRACT
The present procedures for assessment of obstructive airway diseases include spirometry, the response to inhaled bronchodilators and/or challenge with bronchoconstrictive stimuli. As an alternative empirical approach, omics techniques allow analysis of the overall molecular pattern of complex samples such as exhaled breath in a single measurement. This perspective describes the recent advances, current status and future developments of breathomics using gas chromatography mass spectrometry and electronic noses both from a medical and technical point of view. Examples of medical advantages of breathomics are the noninvasiveness of the test, the possibility to store breath samples, the availability of portable devices and the applicability of the technique in many diseases. Disadvantages are the inability to distinguish between endogenous and exogenous compounds and the influence of many factors, such as medication and diet, on the composition and quality of air samples. Technical advantages of eNose measurements are the limited costs, reversibility and short response time of sensors, and the use of pattern recognition algorithms. The latter enables a single device to be used in different diseases by simply changing the diagnostic algorithm. Technical drawbacks are the necessity of training the device for each application, incompatibility between devices and limited stability of sensors. This perspective focuses on advances made in the field of pulmonology, but recommendations for application can be easily translated to other fields of medicine. Steps towards introduction of metabolomic breath tests into daily clinical practice are currently undertaken and require joined forces between medical and technical task forces.
INTRODUCTION: OMICS IN PULMONARY MEDICINE

The present procedures for the clinical assessment of airway diseases are rather complex. These include spirometry, the response to inhaled bronchodilators and/or challenge with bronchoconstrictive stimuli [1,2]. These tests have internationally been standardized, and are generally considered to be reliable. However, they are cumbersome, time consuming and not widely applicable. This has seriously limited the implementation of these techniques at all necessary levels of medical care. The strength of current diagnostic procedures is that they are measuring a core component of both diseases, namely, (variable) airways obstruction. Hence, the tests are derived from pathophysiological reasoning, similar to many other diagnostic tests in medicine. This fits into a tradition in which diagnostic tests are developed based on understanding of disease mechanisms.

An increasingly successful alternative of pathophysiological reasoning in diagnosis, monitoring and pathogenetic research is the empirical approach, in which the tests are selected based on probabilistic evidence only. Such procedures are fully evidence-based, whilst being essentially hypothesis-free. The best examples for this are the recently developed high-throughput methods, the so-called ‘omics’ techniques, which allow analysis of the overall molecular pattern of complex biological samples by a single set of measurements [3]. This can for instance be done at the level of RNA (transcriptomics), proteins (proteomics) or metabolites (metabolomics). The information provided by these techniques is not based on detecting single and separate molecular signals, but is exclusively derived from pattern recognition among an array of signals by using powerful bioinformatics. This is based on cluster analysis and learning algorithms (artificial intelligence), which can be trained and applied for classification and identification of disease.

The omics techniques have the potential to be applied in almost every field of medicine. This can be done by an integrative systems medicine approach [4], providing ‘fingerprints’ of disease. In other disease areas, such as cardiology and neurology, this has already led to improved prediction of disease outcome [5,6]. The procedures to evaluate their diagnostic accuracy are essentially the same as for any other diagnostic test. After determining their test accuracy in diagnostic research (including patients with the disease, healthy subjects and subjects suspected for the disease) such tests can be implemented in the clinical or epidemiological setting [7-10].

METABOLOMICS IN EXHALED AIR

Can omics techniques be used in the diagnosis and monitoring of asthma and COPD? Application of proteomics in blood [11] and sputum [12] has provided promising results. As a non-invasive alternative, metabolomics in exhaled breath (breathomics) is potentially suitable for molecular diagnosis in asthma and COPD.

Volatile organic compounds

Already back in 1971 Pauling et al demonstrated that exhaled breath contains a complex mixture of several hundreds of volatile organic compounds (VOCs) [13]. The VOCs in human
Chapter 2

Exhaled breath represents various profiles of hydrocarbons, such as formaldehyde, methanol, ethanol, hydrogen sulphide, benzene, acetaldehyde, propanal, acetone, dimethyl sulphide, isoprene, toluene, phenol, xylene, and many others [14]. These VOCs can potentially be used as non-invasive biomarkers of various biochemical pathways that are operative in health and disease [15]. During all metabolic processes many volatile substances are produced. These VOCs are a result of homeostatic processes as well as disease processes taking place locally in the airways, as is the case in asthma and COPD, as well as elsewhere in the body. The latter substances are transported by the bloodstream to the alveoli, where diffusion into the exhaled breath takes place [16]. VOCs present in exhaled breath samples are most probably the result of metabolic fractioning of larger molecules. Oxidative stress is one such metabolic process, occurring in many diseases, including in COPD and to a lesser extent in asthma. Therefore, VOC patterns rather than individual compounds need to be studied [17].

Over 3000 different VOCs have been detected in human exhaled breath, of which approximately 10% is associated with either normal metabolism or pathophysiological mechanisms (see Figure 1) [18,19]. Only a small portion of VOCs is identified as disease-related [17]. However, as mentioned before and endorsed by the fact that different techniques of exhaled breath testing based on different chemical and physical interactions are able to differentiate between (lung) diseases and health, it is likely that (lung) diseases are characterized by distinctive signatures. These signatures, or breathprints, are not based on single markers or molecules from the same chemical group, but formed by several kinds of VOCs [17].

### Figure 1

**VOC-profiles in exhaled breath.**

<table>
<thead>
<tr>
<th>Greater than 3000 VOCs identified in breath</th>
</tr>
</thead>
<tbody>
<tr>
<td>90% Individual VOCs:</td>
</tr>
<tr>
<td>- Constitution</td>
</tr>
<tr>
<td>- Diet</td>
</tr>
<tr>
<td>- Habits, cosmetics</td>
</tr>
<tr>
<td>- Medication</td>
</tr>
<tr>
<td>- Environment</td>
</tr>
<tr>
<td>10% Common human VOCs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>+/− 300 common VOCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOC-profiles associated with normal metabolism</td>
</tr>
<tr>
<td>VOC-profiles associated with disease</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disease-related VOC-profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altered patterns of:</td>
</tr>
<tr>
<td>- Alkanes</td>
</tr>
<tr>
<td>- Methylalkanes</td>
</tr>
<tr>
<td>- Aldehydes</td>
</tr>
<tr>
<td>- Benzenes</td>
</tr>
<tr>
<td>- Alcohols</td>
</tr>
<tr>
<td>- Etc.</td>
</tr>
<tr>
<td>Associated with processes:</td>
</tr>
<tr>
<td>- Lipid peroxidation, oxidative stress</td>
</tr>
<tr>
<td>- Cholesterol metabolism</td>
</tr>
<tr>
<td>- Smoking</td>
</tr>
<tr>
<td>- Inflammation</td>
</tr>
<tr>
<td>- Etc.</td>
</tr>
</tbody>
</table>
GC-MS
The gas chromatography mass spectrometry (GC-MS) method is the gold standard method of measurement of exhaled breath, but also has limitations ranging from the lack of standardized collecting methods to the difficulties in detecting VOCs that are present in nanomolar (10^{-9} mol) or picomolar (10^{-12} mol) concentrations [19-21].

GC-MS has initially been used for assessment of lung cancer [22-25]. Subsequently, it has been employed in the study of inflammatory lung diseases, including asthma [26-28], cystic fibrosis [29], and COPD [30]. The strength of these measurements is that it allows broad spectrum individual VOC identification, which can aid in the discovery of yet unidentified pathophysiological pathways and new targets for therapy.

ELECTRONIC NOSES
The objective of using electronic noses (eNoses) is recognition of gas mixtures, by using a variety of sensors that capture multiple VOCs [31]. Interestingly, this resembles mammalian olfaction (see Figure 2) [32]. Hence, eNoses strictly follow an empirical approach, in which the suitability of the instrument is determined by the selection of sensors in relation to the gas mixture of interest, and the choice of the pattern recognition algorithms. Medical application

![Figure 2](image)

**Figure 2** The similarities between mammalian and artificial olfaction.
of eNoses is emerging [31,33,34] and goes hand in hand with rapid instrumental and statistical developments. The output of eNoses is a signature of the expired VOC-mixture, which can be regarded as a fingerprint of the exhaled air (breathprint).

There are several principles for eNose sensors [31], varying from conducting polymers [35-37], metal oxide [38], metal oxide field effect transistors, surface or bulk acoustic waves [20], optical sensors [39], colorimetric sensors [40,41], ion mobility spectrometry [42], infrared spectroscopy, gold nanoparticles [43], but eNoses can also be based on GC-MS [31]. Most of these eNoses have been developed for a broad range of applications [31,35,36], which may be a pro when exploring medical application.

A critical methodological step is the breath collection. It has been established that expiratory flow, expired volume, relative humidity, either or not filtering inspired air by VOC-filter, and total versus late expired sampling are all influencing the breathprint, and thereby need to be standardized [26,27,44,45]. Expired air should be sampled either directly or indirectly after short (< 1 hr) storage in an inert collapsible bag [46]. The latter method obviously makes bedside measurements feasible. If prolonged storage is required, thermal desorption tubes should be used, to minimize decomposition and loss of compounds. Another advantage of indirect sampling is the possibility to pre-concentrate breath samples by means of these sorbent traps or by solid-phase microextraction (SPME). Finally, the sample should be related to (filtered) environmental air as background to correct for inspired ambient VOCs [44,47].

The next vital issue is the data analysis. This includes data normalization methods, pattern recognition algorithms and classification techniques [48]. There are rapid developments in these analysis techniques, and there is no doubt that novel, so-called ‘systems medicine’ approaches will more and more benefit from analytical improvements [3,4]. The risk of these high-dimensional analyses is false discoveries, which must be dealt with very carefully by using adequate recommendations [49]. Finally, at present eNoses are not identical, not even from the same brand and sensor type. Therefore, a final critical step for future implementation of this technology will be to link multiple eNose systems so that information from one eNose can be used by another [50].

**ENOSE IN LUNG DISEASE**

eNose analysis has potential in the diagnosis, phenotyping and monitoring of diseases like asthma and COPD. It appears from studies with different eNose sensor systems that asthmatics can be discriminated from healthy controls with cross-validated accuracies between 80-100% (chapters 3 and 4) [26,27,51,52]. Interestingly, COPD patients can also be discriminated from asthmatics [51,52] as well as from patients with lung cancer [53]. Lung cancer is an obvious candidate for eNose assessment [40,43,54-56] as well as infectious diseases [41,57] or diabetes [58-62].

All these data are based on cross-validation procedures in so-called training sets. The critical step in these diagnostic assessments is external validation [10,49]. Interestingly, the first data based on including validation sets of newly recruited patients with asthma and COPD
are showing highly successful identification and differential diagnosis (chapter 4) [52]. This is all probabilistic modeling, in which the breathprints contribute to molecular phenotyping of patients. eNoses do not contribute to the discovery of specific biomarkers or targets for therapy. The latter requires identification of individual molecular compounds e.g. by GC-MS.

Interestingly, eNose signals seem to be associated with inflammatory phenotypes in COPD (chapter 6) [63], which resembles similar associations between exhaled breath condensate metabolomics and inflammatory phenotypes in asthma [64]. These developments are very promising in view of simplifying future office-based monitoring procedures in selecting patients eligible for anti-inflammatory therapy [65,66].

**MEDICAL PRO’S AND CON’S**

Exhaled breath metabolomics has many advantages, the greatest of which is certainly the non-invasive nature of the measurement. Depending on the technique, patients are asked to exhale in a bag or device to provide an air sample. Little coordination or understanding is needed, making this technique suitable for small children as well. Another advantage is the short duration of the test from the patient’s perspective. When compared to pulmonary function tests, metabolomic breath assessment takes a fraction of the time to perform an adequate measurement, making it a convenient, repeatable and safe alternative. The possibility to store breath samples on sorbent traps, and the availability of portable electronic nose devices even brings application of this technique in primary care or in-house monitoring a step closer. At the other end of the spectrum, the world-wide spread of diseases such as tuberculosis may be monitored and controlled more closely [67].

Exhaled breath testing seems feasible in many diseases. Not only in diseases related to the respiratory tract, such as asthma and COPD, but also in systemic diseases. Many of the VOCs in exhaled breath are blood borne and may therefore reflect pathophysiological processes taking place anywhere in the body [31].

Last but not least, with ever increasing health care costs, eNose assessment may serve as a cheap alternative to repetitive lung function testing in asthma and COPD and as a simple screening instrument for diseases such as lung cancer in the near future.

However, several medical disadvantages of metabolomic breath profiling can be identified. An important disadvantage is the inability to distinguish between endogenous and exogenous VOCs in the exhaled breath. Of course this also holds for the GC-MS technique. Many factors influence the composition and the quality of exhaled breath samples, not only diseases, but also gender, food intake, environmental influences, smoking and use of cosmetics [16,31,68]. Analysis of VOCs can not readily replace measurements of biomarkers in blood. Although many VOCs are blood borne, their concentrations vary enormously due to physiochemical properties influencing the diffusion across the alveolar-capillary membrane [16].

A drawback specific for electronic noses is the inability to identify disease-related VOCs and thereby unraveling disease-associated pathways that may eventually lead to new targets.
for therapy. To overcome this limitation, research programs that include GC-MS analysis of breath samples should be conducted. In turn, the identification of critical VOCs may lead to the development of more sensitive and specific electronic nose sensors.

**TECHNICAL PRO’S AND CON’S**

Metabolomic profiling of breath has several technical advantages. First, the matrix of breath is much less complex than that of urine or blood, even though over 3000 different VOCs have been identified [18,69,70]. Exhaled breath is therefore a suitable specimen for diagnostic tests, that requires little sample handling.

As opposed to the standard analytical method for exhaled breath samples, GC-MS, breathomics using electronic noses poses no need for highly skilled personnel and fully equipped laboratories. The limited costs of the materials are a second major advantage. Besides this, the response time of the gas sensors is generally very short, enabling on the spot diagnosis in large groups of subjects and shortening doctor’s delay before start of treatment. Another technical advantage is the reversibility of the binding of VOCs to the gas sensors [35]. The device can be used repeatedly without previous measurements influencing future samples.

The omics technique for the detection of disease (states) benefits from the fact that concentrations of VOCs in breath generally are not higher or lower than in healthy controls. In fact, pattern recognition seems more suitable to detect the whole spectrum of VOCs, which changes in disease [16]. Because this technique is based on pattern recognition, it has the potential to be applicable in many different diseases, ranging from pulmonary diseases [26,27,51,52] to metabolomic changes in homeostasis as seen in cancer [23,54,71], diabetes [58-62], and even schizophrenia [72,73]. Hence, a single device may be used for different diseases requiring merely a different diagnostic algorithm.

eNose technology also has some technical drawbacks. In order to recognize patterns of VOCs associated with disease, devices have to be trained first. New incoming data are compared to an established set of VOC profiles, before a classification can be made [67]. In practice, this means that devices have to be trained for every clinically relevant disease state.

Second, raw sensor data from different electronic nose devices, even from the same brand, are incomparable. This prevents direct signal exchange between devices.

A third limitation is the stability of sensors. Gas sensors that are continuously stable are rare and most sensors tend to shift in their response over a long period of time [67]. This can have several causes, including sensor poisoning or irreversible saturation of the sensors with VOCs, ageing of sensors or environmental changes such as relative humidity, temperature and composition of the reference gas or surrounding air [67,74,75]. To overcome this limitation, several methods have been proposed, the most obvious being the creation of stable sensors that have no or little tendency to drift. Other options are the use of calibration gases with a known composition to correct for any drift and sensor modeling through a time-adapting dynamic system which predicts the sensor responses of each sensor based on the responses of the other sensors [76].
STATISTICAL APPROACHES
Besides technical considerations, the statistical approach of eNose data is the key for reliable results. When exposed to a gas mixture, a pattern of sensor responses is created. The first stage of data analysis is data acquisition, in which raw sensor data is converted to an electrical signal pattern that can be processed by pattern recognition techniques [48]. Subsequently, the number of dimensions is reduced by feature extraction. Feature extraction by principal component reduction and subsequent pattern recognition by linear discriminant analysis are the most commonly used types of data analysis for eNose data. However, there are several other analysis techniques suitable for eNose data, including principal component regression, discriminant factor analysis, analysis of variance between groups (ANOVA), partial least squares, and cluster analysis methods. In addition, more intuitive approaches of pattern recognition suitable for highly non-linear data, the artificial intelligence techniques including neural networks and fuzzy logic, are also available. Each application of eNose breath analysis may have its optimal analysis method depending on the pattern of VOCs of interest. Besides this, test accuracy may be optimized for the goal of the test (in- or exclusion of disease) by choosing the appropriate cutoff value. The common principle in these analyses is to limit false-positive discrimination, for which explicit guidelines have been given [49]. These were adopted in the present thesis.

FUTURE DEVELOPMENTS
Tailor-made eNoses or diagnostic algorithms for each disease may be a feasible next step, when the determining VOCs of the disease of interest have been established [54]. As a mobile medical application, future eNoses need to consist of miniaturized sensors that are highly responsive and chemically tunable, wireless and operable at low power (i.e. on batteries) [36,37]. As pointed out before, efforts should be directed towards optimization of the sensitivity of the test with high negative predictive values first, before specificity and positive predictive values are targeted. An important technical target for future development is the translation or comparability between devices, eliminating the necessity of training each individual device for each disease. Ideally, an umbrella database of disease breathprints is created to which newly uploaded profiles could be compared, and as soon as the diagnosis has been established, can be added to expand the database.

In conclusion, achievements during the past 20 years paved the way for application of advanced electronic noses in medical practice. Before exhaled breath metabolomic tests are ready for introduction into general clinical practice, joined medical and technical task forces should be established to provide guidelines on the standardization of breath collection, sampling, analysis and interpretation of exhaled breath metabolomic test results.
REFERENCES

34. Friedrich MJ. Scientists seek to sniff out diseases: electronic “noses” may someday be diagnostic tools. JAMA 2009;301:585-6.


73. Puri BK, Ross BM, Treasaden IH. Increased levels of ethane, a non-invasive, quantitative, direct marker of n-3 lipid peroxidation, in the breath of patients with schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 2008; 32:858-62.

