Review

CellPress

Taking your breath away: metabolomics breathes life in to personalized medicine

Nicholas J.W. Rattray[1*], Zahra Hamrang[2*], Drupad K. Trivedi[1], Royston Goodacre[1], and Stephen J. Fowler[3,4]

[1] School of Chemistry and Manchester Institute of Biotechnology, University of Manchester, Manchester, UK
[2] Manchester Pharmacy School, University of Manchester, Manchester, UK
[3] University of Manchester, Manchester Academic Health Science Centre, NIHR Respiratory and Allergy Clinical Research Facility, University Hospital of South Manchester, Manchester, UK
[4] Respiratory Medicine, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK

Breath-based metabolomics (breathomics) is an exciting developing area of biotechnology that centers on the capture, identification, and quantification of volatile organic compound (VOC) patterns in human breath and their utilization as tools in the diagnosis of a broad spectrum of medical problems. With the age of personalized medicines demanding rapid bespoke diagnosis and treatment, this area of molecular diagnostics is beginning to see an upsurge in biotechnological advancement. Here, we discuss recent improvements and directions in the development of breath VOC analysis and diagnosis platforms that offer the potential for disease biomarker discovery and disease prognosis.

VOCs as an unmapped source of biomarker potential within the breath

Breathing is something we all do approximately 20 000 times a day and is a medium that contains a wealth of potential for non-invasive disease detection. Currently, components of breath studied include breath condensate [1], exhaled breath particles [2], and in some instances, combined with carbon isotope-labeled substrates to investigate drug metabolism [3]. Given the broad nature of this field, we highlight current trends and the future direction of the potential of gas-based VOCs (see Glossary) as a source of disease diagnosis.

The modern era of breath gas analysis started in 1971 when Linus Pauling's pioneering paper [4] to the National Academy of Sciences communicated that a typical human breath signature comprised over 250 spectral features containing information potentially originating from VOCs within the breath. These compounds are the products of numerous highly dynamic and regulated metabolic processes not only locally in the lung, but also throughout the body, whereby VOCs generated peripherally are transported to the breath via the pulmonary circulation and subsequent alveolar blood–gas exchange.

The relatively slow pace of breath biomarker research since is in part due to technological limitations associated with the complexity of reliably capturing breath, the analytical intricacy of extrapolating potential biomarkers from endogenous signals, and the inability of the chemical detectors used to achieve sufficient sensitivity and analyte specificity for the many low concentration compounds.

With current human biomedical monitoring focusing on tissue samples and bodily fluids as a source of diagnosis [5], breath-based biomarkers remain among the least developed despite having huge potential as a non-invasive diagnostic source.

Over the past few years, 3400 individual VOCs have been detected from deep alveolar breath, and breath diagnostics has garnered increasing attention from the media (www.bbc.co.uk/news/science-environment-22013700) and governing bodies [6]. The renaissance in breath diagnostics is fuelled by advancements in adaptive sampling methodology and an explosion in the diversity, versatility, and sensitivity of associated detection platforms. It is important that this upward trend in biotechnological advancement is continued in order to develop non-invasive real-time disease diagnostics that will be available in the clinic and at the bedside, at the point of care (PoC) [7,8].

The breath analysis pipeline

The breathomics analysis pipeline (Figure 1) is comparable to other metabolomics-based procedures for other substrates, such as biofluids and tissues [9,10]. A sample is collected in a standardized manner and chemically profiled by a suitable platform to gain the largest representative (in terms of reproducibility and robustness) metabolomic signature [11]. This can be performed on larger systems, such as gas chromatography-mass spectrometry (GC-MS). Subsequent data preprocessing and statistical or chemometric analysis of the platform data are carried out to identify significant metabolic panels that can be linked to

Corresponding authors: Rattray, N.J.W. (Nicholas.Rattray@manchester.ac.uk, nikrattray@gmail.com); Fowler, S.J. (Stephen.Fowler@manchester.ac.uk).
Keywords: breathomics; volatile organic compounds; personalized medicine; point-of-care; eNose; mass spectrometry.
[*]These authors contributed equally to this work.

0167-7799/
 http://dx.doi.org/10.1016/j.tibtech.2014.08.003

Glossary

Breathome: the profile of detectable biomolecules in the gaseous and/or liquid phase of exhaled breath.
Chemical arrays: chemiresistors based on a random network of carbon nanotubes with polyaromatic hydrocarbons containing a hydrophobic mesogen and terminated with ether groups of 2-ethyl hexane groups. Chemiresistors were based on gold nanoparticles coated with various hydrocarbons to ensure a high degree of cross-reactivity of ligands with a diverse range of VOCs.
Chronic obstructive pulmonary disease (COPD): a lung disease largely caused by smoking, associated with airways inflammation, recurrent exacerbations, and progressive difficulties in breathing associated with the narrowing and collapse of the airways.
Differential mobility spectrometry (DMS): ions are subjected to various electric field strengths for different periods, such that ions with certain mobilities will remain.
eNose: an artificial sensor system enabling quantitative or qualitative analysis of VOC mixtures.
Field asymmetric waveform ion mobility spectrometry (FAIMS): an ion separation technique in which ions are separated at atmospheric pressure via application of a high-voltage asymmetric waveform.
Flame ionization detector (FID): a detector type, frequently used after gas chromatography, that quantifies the concentration of organic species in a gas stream.
Gas chromatograph mass spectrometry (GC-MS): analytical method used to resolve and identify complex mixtures in the gaseous form.
Hierarchical cluster analysis (HCA): a statistical method used for finding homogeneous clusters of cases based on measured characteristics by reducing number of clusters gradually until there is only one significant cluster remaining.
Kernel-based orthogonal projections to latent structures (K-OPLS): algorithms used to predict variance between responses and variables for certain data sets.
Principal component analysis (PCA): a multivariate analysis test that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, called principal components.
ProteoWizard: a set of modular and extensible open-source, cross-platform tools, and software libraries that facilitate proteomics data analysis.
Proton transfer reaction mass spectrometry (PTR-MS): a sensitive online tool for monitoring VOCs that comprises an ion source directly connected to a drift tube and an analyzer.
R: a software package for statistical data analysis. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS.
Receiver operating characteristic (ROC): ROC curves are graphical tools to select possibly optimal models and to discard suboptimal ones independently from the class distribution.
Selected ion flow tube mass spectrometry (SIFT-MS): a quantitative approach that involves chemical ionization of low concentration volatile compounds by selected positive precursor ions during a defined period along a flow tube.
Solid-phase micro-extraction (SPME): a sample preparation technique that involves the use of a fibre coated with an extracting phase (liquid or solid) that can extract various types (volatile or non-volatile) of analyte from gaseous or liquid media.
Surface acoustic wave (SAW): a detector in gas chromatography that quantifies concentrations directly in proportion to the frequency of a surface acoustic wave.
Thermal desorption (TD): a technology that utilizes heat to volatilize a gaseous substances from a solid matrix.
Time warping: a basic time-warping algorithm computes the time axis stretch that optimally maps one time series (query) onto another (reference); it outputs the remaining cumulative distance between the two.
Volatile organic compounds (VOCs): organic molecules displaying a high vapor pressure at ambient pressure (VOCs can be detected in the part per billion range).

the biology of the disease in question. The final step is the advancement of targeted, miniaturized systems that can be used in a clinical setting by healthcare professionals who are not necessarily experts in bioanalysis.

Breath analysis is encountering several distinct challenges in the attempts to miniaturize inherently complex technology. The main areas of sample collection and analytical processing via MS analysis have been the foci of improvement to date. The cycle of development starts with the collection of a pilot sample set (preferably in duplicate). Duplicate sampling allows for tandem analysis. A large MS platform can positively identify VOCs, concurrently permitting external validation and signal correlation of the 'breathprint' created by a smaller sensor-based device. The smaller devices have more prospect of acceptance at PoC within a clinical setting. Subsequent chemometric analysis, which must be appropriately validated [12,13] and reported so that others can reproduce this process [14], is used to classify potential patterns that can be linked to specific disease pathology and any potential biomarker relations.

Current VOC-capturing technologies

Measurement of exhaled-breath VOCs requires analytical methodologies that reproducibly capture analytes of interest while minimizing interference from the sample matrix. For metabolomic profiling, the sampling process should also be as chemically unselective as possible. Breath-capture methods range from directly breathing into an analysis platform [15] and the relatively simplistic collection within plastic Tedlar® and aluminized Mylar bags or BioVOC™ bottles, to more complex systems that try to maximize sample quality and minimize the influence of the main challenges in breath analysis. Effective capture of a sample requires minimizing interference from exogenous environmental VOCs [16] or VOCs that do not originate from the area of interest, such as the lower respiratory tract or systemic compartment. VOCs from the upper respiratory tract and mouth are the targets of interest. Using a buffered-end tidal system [17] is one way to achieve efficient breath capture, along with using the breath collection unit developed by Ionicon. By contrast, uncontrolled end exhaled-breath sampling does not reliably and reproducibly measure concentrations of alveolar VOCs, and a steady breathing pattern should be established before analysis [18]. The breath-sampling apparatus must be acceptable to target patient groups (which may include those in respiratory distress or under ventilation) and to the operator. The apparatus must be safe and should comply with appropriate infection control requirements for use in the clinical environment. Current VOC-trapping technologies have their relative merits and drawbacks (Table 1).

Once the breath has been sampled, it needs to be introduced into the analytical platform. In the case of GC-MS, this is commonly via thermal desorption (TD) [19], solid-phase microextraction (SPME) [20], or porous monotraps (http://www.hichrom.com/product_range/existing_products/GLS/Monotrap.htm). These methods fall under the umbrella of active absorbent-based trapping and involve drawing the breath sample over a fiber or through a tube containing a combination of adsorbent polymers or activated charcoal. The adsorbent materials are heated to release trapped VOCs, which are then applied to a GC column. These methods benefit from being inherently designed to preconcentrate VOCs, can be tailored to target specific groups of VOCs, and initial sampling can be done by portable equipment before analysis via MS.

Current VOC identification platforms

Currently, the main analytical platforms for biomarker discovery and identification use time-of-flight (ToF) MS,