White Paper 038
December 2020

Uncovering hidden compositional changes in breath profiles using untargeted chemometric workflows

This study describes the use of thermal desorption (TD) and GC×GC–TOF MS for exploratory profiling of biomarkers in breath, coupled with innovative data mining tools to uncover hidden compositional changes using automated untargeted workflows.

Introduction

Volatile organic compounds (VOCs) emitted in breath have great potential for use in non-invasive disease diagnosis. This is largely due to the discovery of so-called 'biomarkers', which provide indicators of normal or abnormal states.

In large-scale clinical trials, hundreds of samples may be collected across multiple sites (e.g. clinics or hospitals) over the course of many weeks. During this biomarker discovery phase, an incorrect identification can compromise the validity of an entire trial, meaning that both robust analytical techniques and confident data mining are required.

Thermal desorption (TD) coupled with GC–MS is known as the 'gold standard' for breath analysis, due to its ability to capture a complete breath profile with high sensitivity. Here, we combine TD with advanced separation and detection by GC×GC–TOF MS to gain greater insight into sample composition.

However, data acquisition is just the beginning – the information-rich chromatograms must then be transformed into meaningful results. Here, we demonstrate the use of ChromCompare+, a powerful data mining and chemometrics platform, to automatically find significant differences between sample classes in complex datasets.

ChromCompare+ allows the raw data to be imported directly, minimising manual steps and reducing the risk of operator error prior to applying chemometrics. An integral step of the workflow is automated alignment of the chromatograms to ensure that retention time drift (a common problem in long-term clinical studies) does not adversely impact the results. Feature Discovery is performed on the raw data to find significant differences between sample classes. In metabolomics

SepSolve Analytical Ltd
T: +44 (0)1733 669222 (UK) +1 888-379-3835 (USA)
E: hello@sepsolve.com
www.sepsolve.com

matrices, the diagnostic compounds are rarely of high abundance, but by utilising all the raw data, trace differences are less likely to be overlooked.

We will demonstrate how these innovative tools can allow automated untargeted workflows to be adopted, minimising laborious pre-processing steps and accelerating analytical workflows.

Experimental

Sampling: 1 L breath was collected using sampling bags from two groups of participants – five individuals in Group A and four individuals in Group B – and transferred to 'Biomonitoring' sorbent tubes (Markes International).

TD: Instrument: Centri® (Markes International) using the tube-based TD module with a 50-tube autosampler.

GC×GC: INSIGHT® flow modulator (SepSolve Analytical); PM 2.5 s.

TOF MS: BenchTOF-Select™ (SepSolve Analytical); Mass range: m/z 45–450; Acquisition rate: 100 Hz in Tandem Ionisation® mode (with 70 eV and 12 eV data acquired simultaneously).

Software: Full instrument control by ChromSpace®, with data processing and chemometrics in ChromCompare+ (SepSolve Analytical).

Please contact SepSolve for full analytical parameters.

Results and discussion

In this study, TD–GC×GC–TOF MS was used to capture nine comprehensive breath profiles from two groups of participants, with the goal of identifying compositional differences between the two groups (Figure 1).

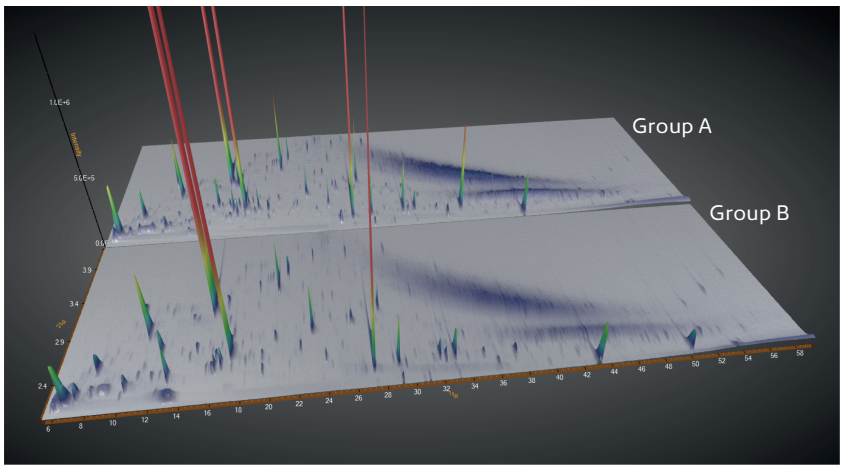

Figure 1
TD–GC×GC–TOF MS surface charts of an example breath profile from each participant group.

SepSolve Analytical Ltd
T: +44 (0)1733 669222 (UK) +1 888-379-3835 (USA)
E: hello@sepsolve.com

A key advantage of GC×GC–TOF MS is the generation of structured chromatograms, where chemical classes elute in bands, aiding in the identification of unknown compounds. The expanded region of the chromatogram in Figure 2 shows that an increase in second dimension retention time (^2t_R) is associated with a decrease in analyte polarity.

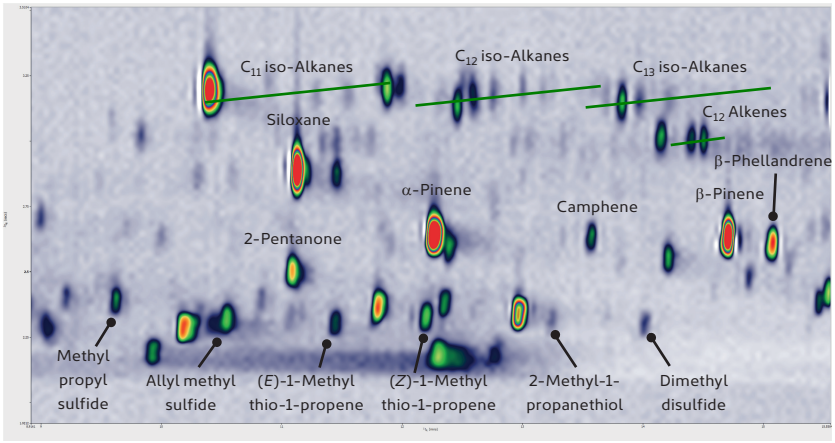

Figure 2

Expanded region of a TD–GC×GC–TOF MS colour plot for a breath profile from Group B, highlighting the structured ordering of GC×GC chromatograms.

TD–GC×GC–TOF MS clearly provides a wealth of information on sample composition but has often resulted in data analysis being the most challenging and time-consuming step of the analytical workflow.

A novel chemometrics platform, ChromCompare+, was used to tackle this challenge and automatically uncover the significant differences between the two groups of samples. The raw data was aligned and imported directly into ChromCompare+ using the innovative untargeted workflow described in Figure 3. This approach divides the chromatogram into small sections, allowing every m/z channel to be compared for every section of each chromatogram in the dataset. These sections can even be overlapped (in this case by 30%) to further reduce the risk of missing important details. This approach is performed in automated sequences, reducing the need for laborious pre-processing steps (such as integration and identification), thereby accelerating discovery workflows.

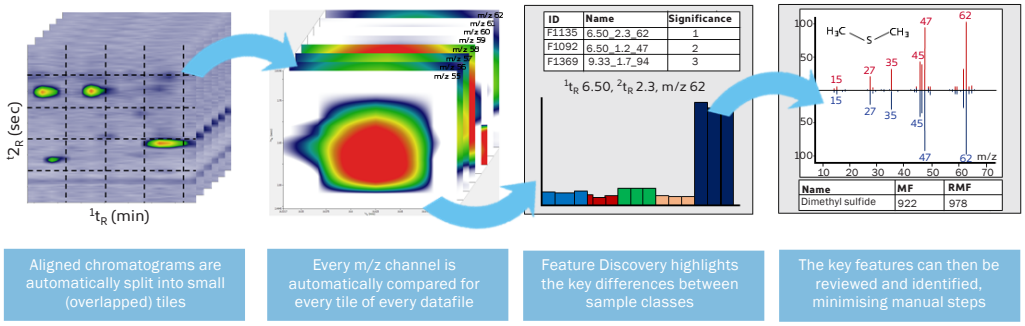

Figure 3

Overview of the automated untargeted workflow in ChromCompare+.

SepSolve Analytical Ltd
T: +44 (0)1733 669222 (UK) +1 888-379-3835 (USA)
E: hello@sepsolve.com

Once the raw data was imported in to the ChromCompare+ platform, various data reduction and visualisation tools were used to investigate the differences between the sample classes (in this case, Groups A and B).

Firstly, Feature Discovery was used to reduce the total number of features from over 800,000 to the top 100 discriminants. The principal components analysis (PCA) score plot clearly shows distinct clusters for the two sample groups after this feature selection step (Figure 4), with PC1 differentiating the two sample groups and PC2 displaying intraclass differentiation, likely due to biological variance between participants in the study.

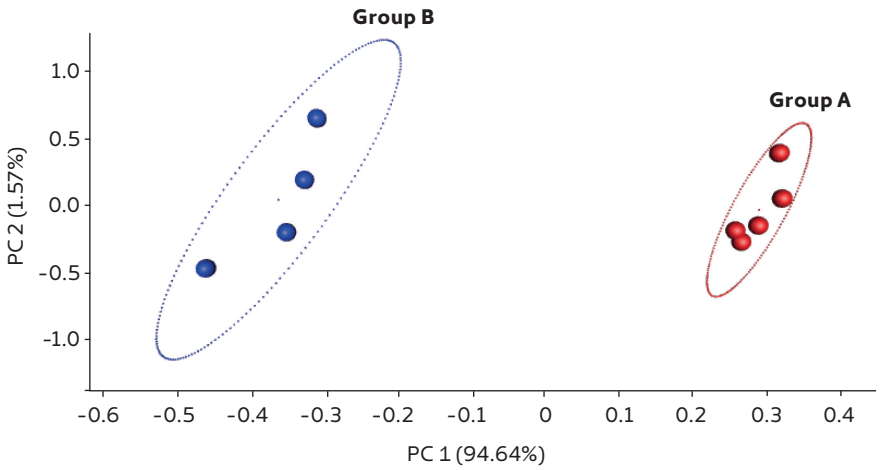

Figure 4

Principal components analysis (PCA) score plot in ChromCompare+ software, showing two distinct clusters for the participant groups after Feature Discovery.

This reduced feature list was then viewed as a volcano plot (Figure 5) displaying the statistical significance of differences between classes for each feature versus the magnitude of this difference, ensuring that small yet significant differences were not overlooked. This plot conveniently highlights analytes that were statistically increased (red) or decreased (blue) in the Group B samples relative to the Group A samples.

Figure 5

A volcano plot in ChromCompare+ for 100 features selected using Feature Discovery. Five statistically significant features that were found in increased abundance in the Group B samples are annotated (A–E) and identified in Figure 6.

SepSolve Analytical Ltd
T: +44 (0)1733 669222 (UK) +1 888-379-3835 (USA)
E: hello@sepsolve.com

In this case, five statistically significantly features stand out in the volcano plot (labelled A–E). Using the retention time and m/z information provided in the feature list, the analytes that these features represent could be identified (Figure 6). The workflow is therefore streamlined, as it eliminates the need to integrate and identify hundreds of peaks that may not be relevant.

In this case, the most significant differences were shown to be sulfur-containing species that were found in the Group B samples only. Despite their trace abundance (an intensity three orders of magnitude lower than the highest-loading components), the untargeted raw data approach used in ChromCompare+ ensured that these differences were not overlooked. This is imperative in discovery studies, where the biomarkers of interest are not yet known.

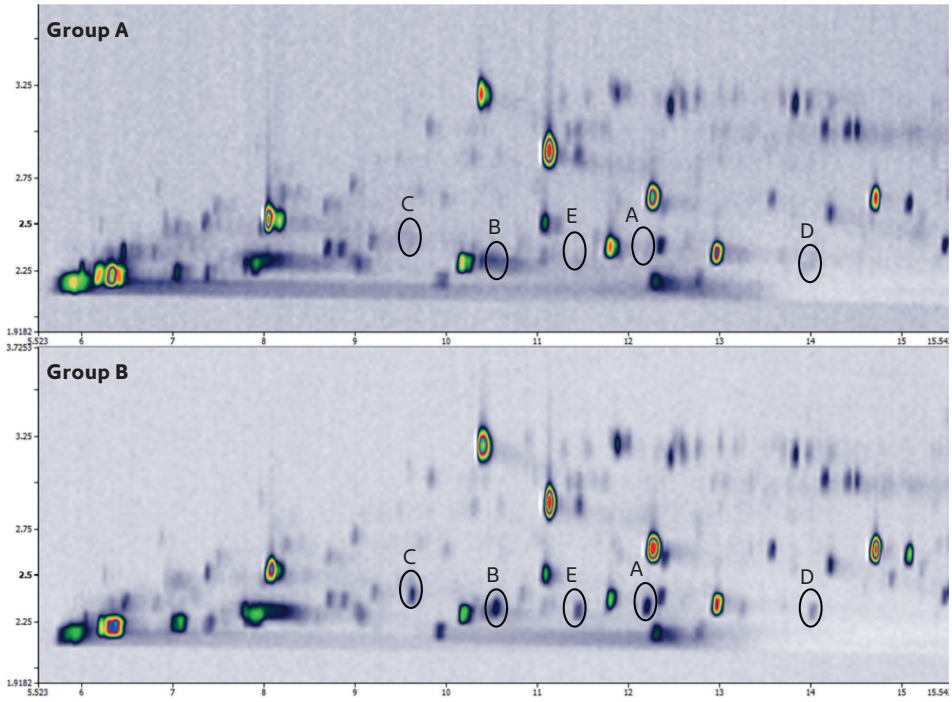

Figure 6

TD–GC×GC–TOF MS colour plots showing the identification of the five significant differences between Groups A and B, as labelled A–E in Figure 5.

Label	^1t_R (min)	^2t_R (s)	Identification	Significance	Fold-change
A	12.200	2.319	(*Z*)-1-Methylthio-1-propene	14.86	11.08
B	10.557	2.307	Allyl methyl sulfide	16.82	3.36
C	9.626	2.384	Methyl propyl sulfide	20.06	2.28
D	14.012	2.290	Dimethyl disulfide	25.63	2.62
E	11.455	2.299	(*E*)-1-Methylthio-1-propene	28.37	1.99

In fact, the samples analysed in this study were taken from participants before lunch (Group A) and after lunch (Group B), so it is not surprising that sulfur species were detected as they are likely to result from the food and drink consumed. However, the study does illustrate the power of TD–GC×GC–TOF MS combined with chemometrics to identify changes in breath profiles during clinical trials, which could indicate a change in metabolism caused by disease.

It is also worth noting that these trace sulfur components would have co-eluted with a number of high-loading peaks in a traditional 1D GC separation, but the enhanced separation of GC×GC–TOF MS, combined with the powerful chemometrics in ChromCompare+, ensures that they are identified with confidence (Figure 7).

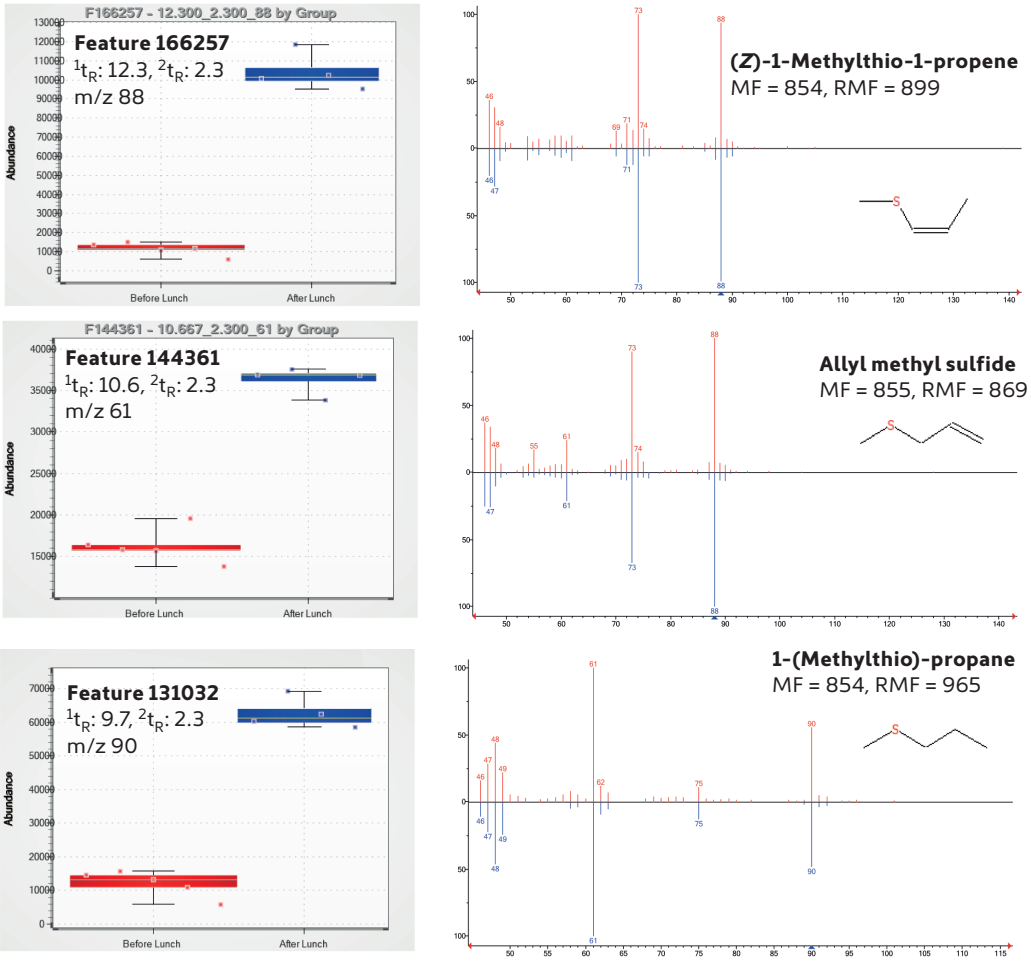

Figure 7

Left: ChromCompare+ feature box and whisker plots highlighting the abundance and intra-class variance for a selection of significant features. Right: Identification of the features by comparison of BenchTOF spectra (red) against the NIST17 library (blue).

Conclusions

This study has shown that the TD–GC×GC–TOF MS workflow described offers the following advantages for analysis of breath biomarkers:

- TD–GC×GC–TOF MS captures comprehensive breath profiles with high sensitivity to gain maximum insight into sample composition.
- Fully automated data analysis in ChromCompare+ minimises laborious pre-processing steps and accelerates workflows.
- Importing the entire raw dataset reduces the risk of missing important trace differences, thereby increasing confidence in results.
- Interactive charts, such as PCA plots, volcano plots and box plots, to show trends and differences between samples.

For more information on this application, or any of the techniques or products used, please contact SepSolve.

D0077_1_011220

SepSolve Analytical Ltd
T: +44 (0)1733 669222 (UK) +1 888-379-3835 (USA)
E: hello@sepsolve.com