Large-scale targeted biomarker analysis of volatile organic compounds in breath by TD-GC-MS

Darren Barrington-Light, Thermo Fisher Scientific, Altrincham, United Kingdom
Mircea Martiniuc, Owlstone Medical Ltd, Cambridge, United Kingdom; Laercio Pires Fernandes Jr, Thermo Fisher Scientific, Altrincham, United Kingdom

Abstract

Purpose: The aim of this work was to develop a targeted TD-GC-MS analytical workflow for breath biomarker research, with a data workflow fully contained in Thermo Scientific™ Chromeleon™ software.

Introduction

- Breath Biopsy® is emerging as a powerful technique in the development of novel non-invasive disease biomarkers and precision medicine applications. Gas chromatography (GC) coupled to mass spectrometric (MS) detection remains the gold-standard for the identification and quantitation of the diverse breath metabolome.
- Typical discovery workflows rely on untargeted analysis, which has major limitations in terms of robust compound identification (ID), translatability and cross-study comparisons. On the other hand, targeted analysis provides higher confidence data but is currently limited to small panels of compounds (e.g. carbonyl compounds (Xie et al 2024)).
- Large-scale targeted analysis like the one described herein, will bridge the gap between the two, allowing us to build the Breath Biopsy VOC Atlas®, Owlstone Medical's growing repository of breath-related VOCs and rich metadata, accelerating the growth of breathomics applications.

Materials and methods

Sample Preparation

Sample collection is performed using Owlstone Medical's ReCIVA® Breath Sampler coupled to the CASPER® Portable Air Supply to minimise ambient contamination, as previously described by Arulvasan et al (2024). Approximately 1.25 L of breath are collected onto each sorbent tube. Deuterated internal standards (n=36) are injected onto all calibration standards and sample tubes.

Test Method

Samples (single tube) are analysed on a thermal desorption (TD) unit (Markes TD100xr) coupled to a GC-MS (Thermo Scientific™ Orbitrap™ Q Exactive™ mass spectrometer) system on the method known as the Breath Biopsy OMNI®.

The target list contains 200 compounds originating from breath, based on our previously published work (Arulvasan et al 2024), with additional targets from Arulvasan et al (2025). A solution containing 36 deuterated internal standards (IS) is injected onto all calibrators, check standards and unknowns.

Data Analysis

Processing methods, View Settings and Report Templates were created and run in Chromeleon CDS due its compliance, audit trail functionality and ability to control and run HRAM instruments in a networked environment.

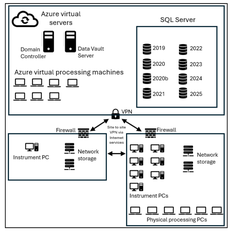
BREATH® BIOPSY

Network Architecture

For this method, approximately 96 MB of data is generated per sample, with each full sequence producing around 6 GB. Across the company, data acquisition has reached 5 to 6.6 TB annually, reflecting the substantial volume of high-resolution accurate mass (HRAM) data generated through routine operations.

The management of instruments and handling of the data are streamlined using Chromeleon CDS, version 7.3.2 MUb, which is deployed as a lab-wide enterprise solution within a cloud environment. This centralized, scalable infrastructure ensures efficient data processing, secure storage, and enhanced accessibility to support scientific workflows.

Figure 1. Cloud-based network allows users to control 9 instruments and perform data analysis remotely with seamlessly integrated data vaults, instrument PCs and data processing virtual machines.



Sample and Data Integrity

System Suitability Testing (SST) with Intelligent Run Control (IRC)

Thermal desorption analysis is destructive, therefore known instrument and sample injection failure modes (e.g. failed sorbent tube pressure test) are put in place to prevent analysis which is expected to be unsuccessful. For more extreme failures, acquisition stops and relevant users are notified by email which allows for rapid resolution.

Patient-derived samples generate critical data and are often irreplaceable, making reinjection impractical or impossible. Therefore, proactive management of out-of-specification (OOS) conditions is essential. To preserve sample integrity and avoid invalid analyses, the system is configured to halt runs when OOS conditions are detected either in the system itself or in preceding injections.

Figure 2. Interactive Results showing SSTs per injection with color-coded pass/fail criteria, allow the analyst to assess the viability of the sample and sequence.

	A	B	C	D	E	F	G
1	Number	Name	Eval. Result	Operator	Ref. Value 1	Ref. Value 2	Result
2	1	Ionisation check	188628591.25	>	1000000.000	n.a.	Passed
3	2	Heated valve temp	200.27	between	195.000	205.000	Passed
4	3	TD Transferline temp	200.21	between	195.000	205.000	Passed
5	4	GC oven temp	270.00	=	270.000	n.a.	Passed
6	5	Back SSL pressure	28.69	between	28.400	29.000	Passed
7	6	Tube Leak (ST)	269.65	between	267.500	272.500	Passed
8	7	Column Bleed - Increase/step change during final hold	1.38	<	1.300	n.a.	Failed
9	8	Column Bleed - Decrease during final hold	1.47	>	1.000	n.a.	
10	9	Column Bleed - Decrease mid-chromatogram - transferline nut	1.17	>	1.000	n.a.	Passed
11	10	FilamentCharging	1.18	<	1.200	n.a.	Passed
12	11	Column bleed	35956760.00	<	40000000.000	n.a.	Passed
13	15	Theoretical Plates (EP) - 4-Ethylphenol-d10	2325445.00	>=	2000000.000	n.a.	Passed
14	19	DiffLok cap test (TBD)	149837.47	>=	15000.000	n.a.	Passed
15	20	Peak Asymmetry	1.06	<=	1.200	n.a.	Passed
16	21	Peak Width (50%)	0.03	<=	0.020	n.a.	Failed

Easy to identify for user to take remedial actions

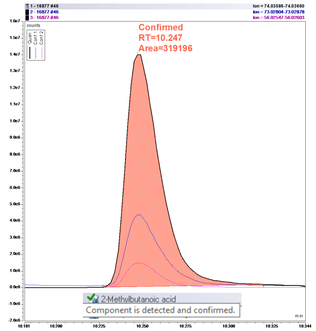
Suggestive of consumable degradation

Compound Identification

Composite Scoring

The identity of each deuterated and target peak is confirmed using Composite Scoring, thus achieving high-confidence (Tier 1) IDs (adapted from Schrimpe-Rutledge et al 2016), by a combination of mass accuracy (≤ 5 ppm), fragment ion coelution (≤ 0.01 min) and ion ratios (20% window), as compared to authentic standards analysed in each analytical sequence. Additionally, in-house HRAM libraries containing >1000 entries can be searched within the software to guide the identification of compounds which are not present in the target list.

Figure 3. Target compound 2-Methylbutanoic acid in breath sample with "Confirmed" Composite Scoring result.

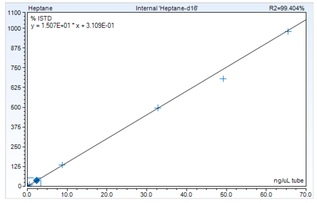


Quantitation

Automatic Gain Control (AGC)

IS correction and compound-specific linear regression with weighting (typically 1/Amount^2) results in direct quantitation of each target compound over linear dynamic ranges between 20 and 4000. Quantitation ranges were set to cover or exceed previously measured breath concentrations, minimizing the need for re-analysis of out-of-range samples.

Figure 4. Calibration curve for heptane showing breath sample with amount within calibration range (0.05-65 ng).



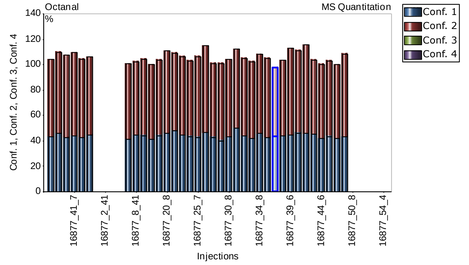
AGC function enables wide linear quantitation ranges
(up to 3 orders of magnitude between LLoQ and ULoQ)

Data Curation

View Settings

Integrated tools for dataset tabulation and visualization allow the creation of customizable, sequence-wide tables and plots with interactive elements to support data analysis

Figure 5. Interactive Chart showing ion ratios for a target compound throughout a sequence, with check standards (used as reference) on the left.



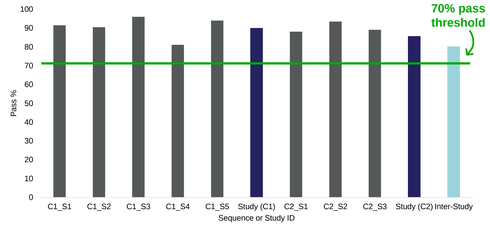
Interactive sequence-wide dataset visualization for rapid data quality assessment

Method Validation

Real-World Results

A method validation and verification study was performed using two breath sample cohorts. The full analytical workflow was applied, incorporating all elements of intermediate precision, with added major instrument maintenance between the two sample cohorts. Precision across the entire workflow was assessed.

Figure 6. Proportion of target compounds (n=200) with check standard (n=6-7 per sequence) precision below 20% RSD threshold within individual sequences, studies and between studies. C = Cohort, S = Sequence. Internal threshold of 70% (≥140 compounds) applied.



Sequence
Study
Inter-study

High precision of quantitation maintained when pooling data across sequences and studies, facilitating longitudinal data collation

Conclusions

An analytical workflow was developed fully within Chromeleon software, including but not limited to:

- **A fully integrated, compliant, and traceable method for data acquisition, analysis, and reporting.**
- **Enterprise-wide deployment in a cloud environment provides centralized instrument control, scalable storage, and remote data access.**
- **Automated system suitability and performance checks (SST/IRC) to prevent loss of critical patient samples by halting acquisition in the event of instrument issues, while ensuring key method parameters remain within validated acceptance criteria.**
- **Robust compound identification through composite scoring and ion ratio confirmation against reference standards, supporting high analytical confidence, improved data quality, and broad compound coverage.**
- **Interactive visualization and data review tools that support efficient curation and ensure the stability and reproducibility of quantitative results across sequences.**

Output: Fully integrated, compliance-ready quantitation workflow for targeted analysis of breath biomarkers.

References

Arulvasan, W. *et al.* (2024) 'High-quality identification of volatile organic compounds (VOCs) originating from breath', *Metabolomics: Official journal of the Metabolomic Society*, 20(5), p. 102.

Arulvasan, W. *et al.* (2025) 'Optimized breath analysis: customized analytical methods and enhanced workflow for broader detection of VOCs', *Metabolomics: Official journal of the Metabolomic Society*, 21(1), p. 17.

Schrimpe-Rutledge, A.C. *et al.* (2016) 'Untargeted metabolomics strategies-challenges and emerging directions', *Journal of the American Society for Mass Spectrometry*, 27(12), pp. 1897–1905.

Xie, Z. *et al.* (2024) 'Detection of COVID-19 by quantitative analysis of carbonyl compounds in exhaled breath', *Scientific reports*, 14(1), p. 14568.

Acknowledgements

The authors would like to thank Tilly Woodland, Kayleigh Arthur, Annamaria Szabo, Steven Levett and Lourenco Saco from Owlstone Medical Ltd for their contributions to the development and validation of the analytical workflow.

Trademarks/licensing

Watch our video: