Breath Biopsy® OMNI® – Example Dataset

This document is provided as supporting documentation with the download of the example dataset for Breath Biopsy OMNI. The dataset provides an example of the data generated from samples of human breath VOCs by Breath Biopsy OMNI and illustrates the kind of feature table that is provided as a basic output from Breath Biopsy studies. Additional reporting options for your own Breath Biopsy studies include further statistical and biological analysis and interpretation where desired.

Contact us to discuss your research with us and to start designing your own Breath Biopsy study with support from our expert team.

Study Overview

The attached data is the result of an internal Owlstone Medical study performed in 2021 during the testing of Breath Biopsy OMNI. We continually make incremental improvements to our methods which mean our current methods may differ from those used here. The study is a breath or blank (BoB) study that we use to examine the detection of volatile organic compounds (VOCs) from breath samples. Our goal with this study was to assess the stability and yield of OMNI for studies consisting of over 100 samples, and to measure the capability of OMNI to detect on-breath VOCs through global (untargeted) analysis.

57 breath samples from four healthy volunteers were collected and analyzed by the Owlstone Medical team at a single site with each volunteer sampled on a separate day. An equivalent number of system blank samples were also collected on the same days and in the same location. All samples were collected using the Breath Biopsy Collection Station, which includes the ReCIVA® Breath Sampler and CASPER® Portable Air Supply. For this study, a pre-production prototype of the Breath Biopsy Mouthpiece was used with ReCIVA. Device settings matched our recommended defaults for Breath Biopsy OMNI studies.

In accordance with our standard practice, each sample consisted of four sorbent tubes, each containing VOCs captured from 1250ml of the lower fraction (alveolar air) from exhaled breath. Sample collection typically takes around 12-15 mins (max 30 mins) and subjects perform normal tidal breathing. Samples were dry purged to remove water before being stored until all samples for the study were collected.

Once all samples had been collected, they were analyzed in the Breath Biopsy Laboratory. VOCs were released using thermal desorption and then analyzed using high resolution accurate mass (HRAM) gas chromatography mass spectrometry (GC-MS). Each analysis pooled VOCs collected in two sorbent tubes, the equivalent of VOCs from 2.5 litres of alveolar breath.

On-breath VOCs

A key distinction from many other methods of breath analysis is that our approach focused specifically on detecting 'on-breath' compounds. We define on-breath compounds as those where the abundance on breath is more than three standard deviations above the average abundance detected in comparable blank samples. Signals of 100,000 or less were also removed from the analysis.

$$on\ breath\ VOC \geq mean_{blanks} + 3 \times std.dev_{blanks}$$

VOCs can arise from many sources, such as food, fragrances, pollution and cleaning products. As such, ambient air contains many VOCs that have no relevance to human biology. CASPER Portable Air Supply is a key part of our efforts to exclude ambient VOCs from breath samples but, in addition, collecting blank samples gives us a more complete awareness of the VOCs that are present in breath samples but that are not relevant to patient health. Therefore, we can be more confident that any VOCs that are significantly more abundant in exhaled breath than in inhaled air (blank samples) have interacted with human biology and so may be relevant as biomarkers.

Volunteer Restrictions

Volunteers were all nominally healthy and did not eat or drink (except water) for at least 90 mins prior to the start of collection until the end of the experiment. They also did not brush their teeth in the 60 mins before or during collection.

Data Overview

The file you have downloaded is a feature table that was generated from the GC-MS chromatograms produced during this study. The initial raw data has been processed using Compound Discoverer software (Thermo Fisher Scientific) and has been normalized to remove batch effects that can result from variation in analytical performance over time. Normalization was done by dividing each signal by the median signal calculated from eight internal standard compounds included with each sample.

The file contains data collected via GC-MS and reported as molecular features, where each feature is a measurable peak in the data. In principle, each feature should be the result of molecules of a specific VOC passing through the mass spectrometer. As such, most features can be tentatively assigned a molecular identity, which could be verified through further experimental work.

The feature table (Figure 1) is provided as both a spreadsheet (xlsx) and csv file. Each row represents a single sample and is labelled as breath or blank and with the associated sample collection session (v1-v4). Each column represents a molecular feature and includes retention time (RT) and, where possible, includes a tentative VOC identity assigned automatically based on the closest match from the national institute of standards and technology (NIST) mass spectral library.

The data includes 114 samples and 2658 detected features.

In a typical Breath Biopsy study, we perform manual curation of feature tables to ensure that, where possible, features are associated to a single unique VOC identity. Note that the most suitable identity is not always the one with the closest library match and that, without manual curation, several features may be assigned the same tentative identity. Since biological identities were not essential to the goals of this study, we have provided this dataset without manual curation. As such, there may be several features that have been assigned the same compound identity.

	Sample	Tags	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5		
1	Sample	Tags	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	F	Feature number
2		Name	Peak@1.6	Peak@1.6	Peak@1.6	Peak@1.6	Cyclotetra P		Assigned name
3		RT [min]	1.631	1.636	1.64	1.64	1.642		Retention time
4	Breath v1	Area: 9684_14_742525_742533_14.raw (F2)		166465	147540		842232		
5	Breath v4	Area: 9684_24_741931_742937_24.raw (F5)			139324		571682		
6	Breath v2	Area: 9684_30_762790_350749_30.raw (F7)		183706	162169		965952		
7	Breath v3	Area: 9699_11_746276_631743_11.raw (F9)			214769		62980?		
8	Breath v1	Area: 9699_20_737020_660910_20.raw (F12)			180701		570607		
9	Breath v4	Area: 9699_27_600369_625507_27.raw (F14)			212777		745329		
10	Breath v3	Area: 9709_20_746910_746903_20.raw (F18)		102854	171096		408908		
11	Breath v4	Area: 9729_11_762625_762623_11.raw (F23)		126276	187263		677887		
12	Breath v2	Area: 9729_20_718948_624791_20.raw (F26)			204199		613332		
13	Breath v3	Area: 9729_27_746246_661952_27.raw (F28)			218079		626385		
14	Breath v4	Area: 9739_20_762784_741866_20.raw (F34)			144354		634782		
15	Breath v3	Area: 9739_27_746906_746304_27.raw (F36)		103691	150589		761904		
16	Breath v1	Area: 9759_27_351130_632256_27...w (F44)					522311		

Sample type (breath or blank) and collection session (v1 to v4)

Tags — The name of the source file for this sample

Features — Each feature likely corresponds to a specific VOC species. Blank spaces mean there was no detectable on-breath signal for that feature

Figure 1: Feature table extract.

Columns A to CXH contain unnormalized data. Column CXL contains the median value calculated from eight internal standard compounds included in each sample and Columns CXP to GVW contain the data normalized relative to the internal standard median.

Library matching and feature naming

The chemical identities of compounds detected through mass spectrometry are often tentatively assigned by matching the detected features to libraries of mass spectra generated using standard compounds. The precise signal produced by each compound is dependent upon a range of factors including the sample, method and device used for analysis. This means that there will always be some disparity between a compound detected in a sample and the corresponding library standard. This is why we only assign tentative IDs based on library matching.

In the associated feature tables, each feature is assigned a tentative identity based on the closest match from the NIST library. It is important to note that these assignments are only tentative and would need further verification before being studied in more detail. To further improve the accuracy of our tentative assignments, we are developing our own Breath Biopsy Library using devices and methods that more closely match our standard procedures for breath analysis and, where possible, we use this to assign higher confidence tentative identities.

Where a suitable NIST library match was not available, features are named based on their retention time, the time at which they eluted from the gas chromatography column into the mass spectrometer e.g. "Peak@1.631".

Results/Performance Metrics

This study detected a total of 2658 features with a median of 1445 per sample, or a median 517 on-breath per sample. We have found the below representations to be highly effective in showing the results of BoB studies and allowing easy comparison across studies. Figure 2 provides an ordered representation of the percent of samples where each VOC was detected on breath. An improved method will detect more VOCs (x-axis) in a higher proportion of samples (y-axis). This representation allows a complete visualization of the data without needing to apply thresholds.
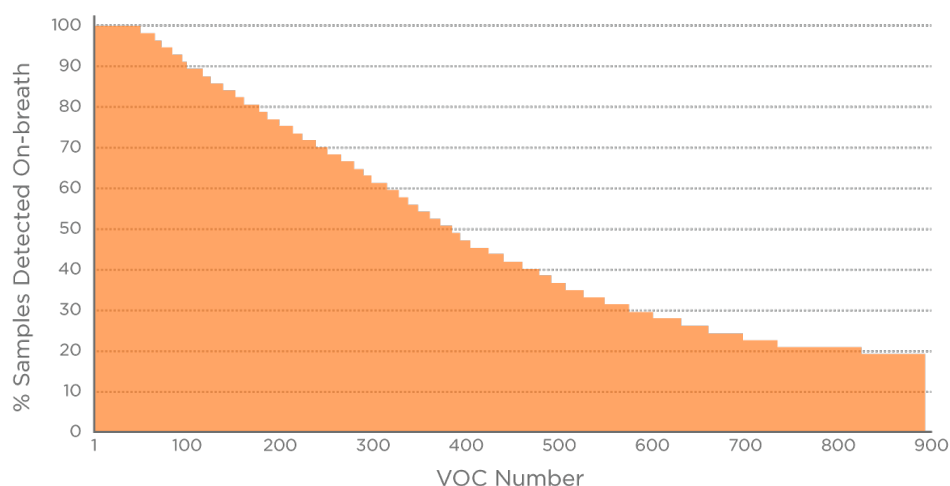
Figure 2: Visualizing on-breath VOCs by sample. A more effective method should be able to detect a larger number of on-breath VOCs in a higher percent of samples.

Figure 3 represents the level of variation in VOC abundances between samples as measured by relative standard deviation (RSD). The graph shows RSDs for the 384 VOCs that were detected on-breath in 50% or more of breath samples (as shown in Figure 2). The median inter-subject RSD across these VOCs was 49% and the median intra-subject RSDs for the four test subjects were 30%, 26%, 30% and 29% respectively.
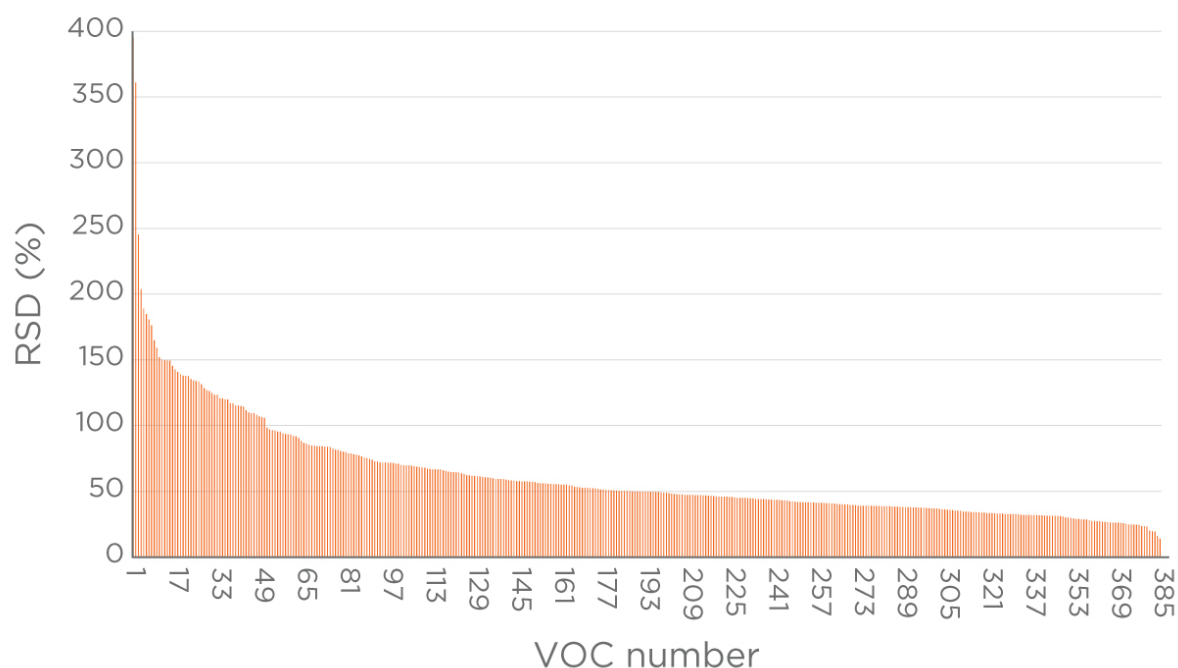

Figure 3: Calculating relative standard deviation for VOCs detected on breath in >50% of samples. Median RSD 49%.

The standard Breath Biopsy protocol includes eight deuterated internal standard compounds with all samples that are analyzed, the low RSD values that we observe for these standards give us confidence that much of the variation we observe within breath samples is due to experimental differences and not a result of analytical effects. The RSDs calculated for the internal standards analyzed in this study are provided in Table 1.

Internal Standard	RSD from breath samples
IS1	9.7%
IS2	13.0%
IS3	9.3%
IS4	10.3%
IS5	7.5%
IS6	5.2%
IS7	3.6%
IS8	3.4%

Table 1: RSD values calculated for eight internal standards analyzed alongside the example Breath Biopsy OMNI dataset

Outcomes

For breath testing to become a valuable diagnostic tool, we need more standardized methods for analysis and reporting that enable comparison across studies and allow us to better understand the relative performance of different technologies. In our view, BoB studies provide a simple and powerful approach that can be applied to any analysis method and produces results that can be easily compared.

We believe that BoB studies have the potential to unlock advancements across the breath field and encourage wider inclusion of BoB in new and upcoming breath research studies. We are continuing to use BoB studies to identify improvements to Breath Biopsy methodologies, contact us for information on our latest improvements.