WHITE PAPER

Breath Biopsy® OMNI®: Data Normalization Project

BREATH® BIOPSY

OMNI®

Owlstone Medical Novel Insights

Authors: Owen Birch[1], Elen Shaw[1], Oliver Dean[1]
[1]Owlstone Medical, Cambridge, UK

Keywords: Breath Biopsy, normalization, internal standards, volatile organic compounds, GC-MS

Discover more about Breath Biopsy OMNI:

Key Points

- Normalization is key to reducing analytical variability in data.
- There are issues associated with using the median of internal standards to capture analytical variability.
- Owlstone Medical has developed an improved method of normalizaton to capture and remove analytical variability from OMNI data more accurately.

Introduction

Breath Biopsy® OMNI® is the most advanced service for global analysis of volatile organic compounds (VOCs) on breath and is a platform that we are constantly innovating. OMNI is a complete end-to-end solution incorporating expert support for study design, data analysis and interpretation as well as specialist technology for breath collection and analysis with high resolution accurate mass (HRAM) mass spectrometry.

Analytical variability is the tendency of an analytical platform to produce slightly different outcomes for repeated meausrements, often as the platform changes over time. Many factors contribute to these changes, such as mechanical wear and tear, replacing consumable components such as ion sources, and recalibration of the platform over many analytical runs. The variability introduced can mask the ground truth of the data, as identical inputs can yield different outputs, and make it difficult to assess the variability of interest.

There are many methods of tracking analytical variability, ranging from quality control (QC) samples to pooled samples. OMNI uses internal standards (ISs); 8 deuterated compounds that are injected into every sample at the same concentration. The increase and decrease over time of peak areas associated with compounds over time can be interpreted as the analytical variability associated with the platform for a specific time frame, such as during a study where samples are analyzed over multiple batches. Analytical variability is clearly undesirable as it can mask any potential biological variability, and so normalization (also known as batch effect correction) is a process that attempts to remove this variability, while preserving variability of interest.

Our previous normalization method consisted of dividing the peak area of every VOC in a sample by the median peak area of its eight ISs. This method leads to only two ISs dictating the scaling factor of every VOC in a sample, and in a past OMNI experiment, we found that these two standards were the same for every VOC in every sample as the relative sensitivity to the different ISs was orders of magnitude larger than the analytical variability.

It has been frequently observed that IS compounds can behave differently from each other over time, so taking only two ISs into account is unlikely to be optimal. Methods which use additional information about ISs (e.g., chemical properties) to select the internal standards used in the normalization of a specific VOC are less likely to over or under correct data, and are more likely to restore truth in our data.

The Normalization Project was run with the aim of developing metrics that quantify the effect of normalization on OMNI data and improve upon our previous normalization method. This new method was to be a robust, fit-for-purpose method that removed analytical variability in OMNI data, while restoring/preserving 'truth' (biological variability) for both targeted and untargeted OMNI data.

Background

The previous method of normalization derived a scaling factor from the median peak area of the eight ISs in a sample, then divided every VOC peak area in the sample by that scaling factor. This is a very common method used in metabolomics – other common methods are the selection of a single IS, or the use of the closest eluting IS. The quality of the normalized data was assessed using principal component analysis (PCA), a multivariate approach that can reduce the dimensionality of data by extracting maximal, orthogonal sources of variation into single features, known as Principal Components (PCs).

A batch is a sequence of up to 30 samples of interest and a number of quality control samples, which are analyzed using the OMNI method; a single study may comprise of many batches. If batch effects are a significant source of variation in the data, clustering by batch is expected in the first few PCs, upon visual inspection. For example, in Figure 1, a PCA of an OMNI dataset containing 4 different types of samples is shown. The two primary sources of variation in the data, PC001 and PC002, account for ~40% of the variance in the data and show a clear clustering by analytical batch with earlier batches standing out from the rest (Figure 1a). Clustering by sample type is shown in later sources of variation, PC003 and PC004 (Figure 1b), which account for 8.5% and 6.6% of the total variance respectively. Differing sample types are inherently distinct, and seeing clustering by them in PCA is desirable, however, in this case undesirable sources of variation (batch) are more prominent thant the desirable variaton between sample types.

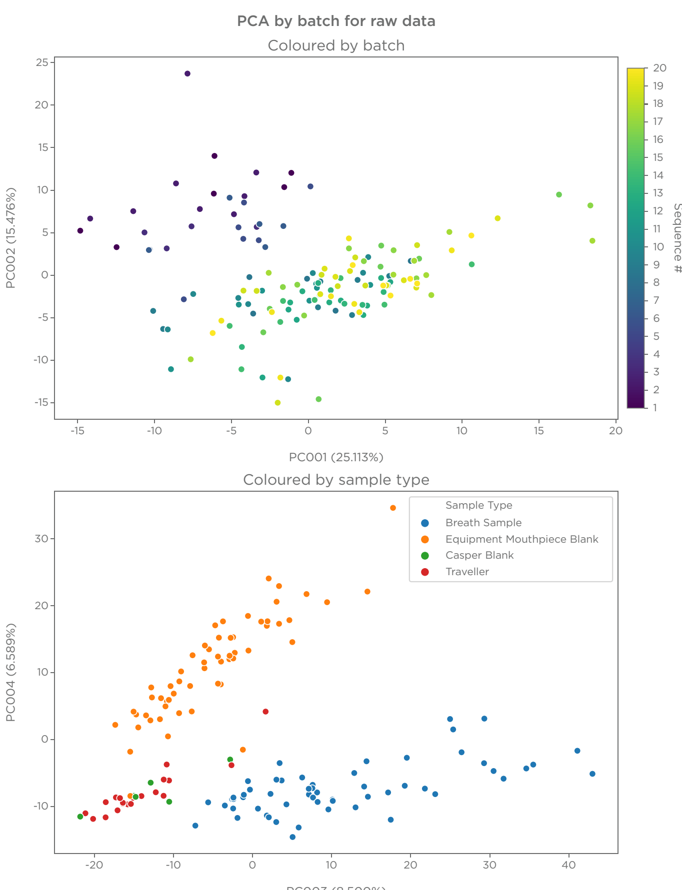


Figure 1: A PCA decomposition of an example dataset, colored by analytical sequence (above) and sample type (below).

These PCAs are inspected once again post normalization to ensure that undesirable analytical variability does not account for a large portion of the variability in the data. Figure 2 shows the same example dataset post normalization. The trends in the data for both batch and sample type appear comparable, but upon further inspection we can see the PCs attributed to these trends have changed, with sample type now clustering in the primary sources of variation (PC001 and PC002), and clustering by batch falling to PC004 and PC005, which contribute much less to the total variance of our data. The earlier batches also overlap more with the later ones when compared to pre-normalized data.

This is a positive outcome for the data set, as being sure that the variation in the data is associated with the primary outcome provides a greater probability of finding statistically significant results when conducting statistical analysis and reduces the chances of false positive results.

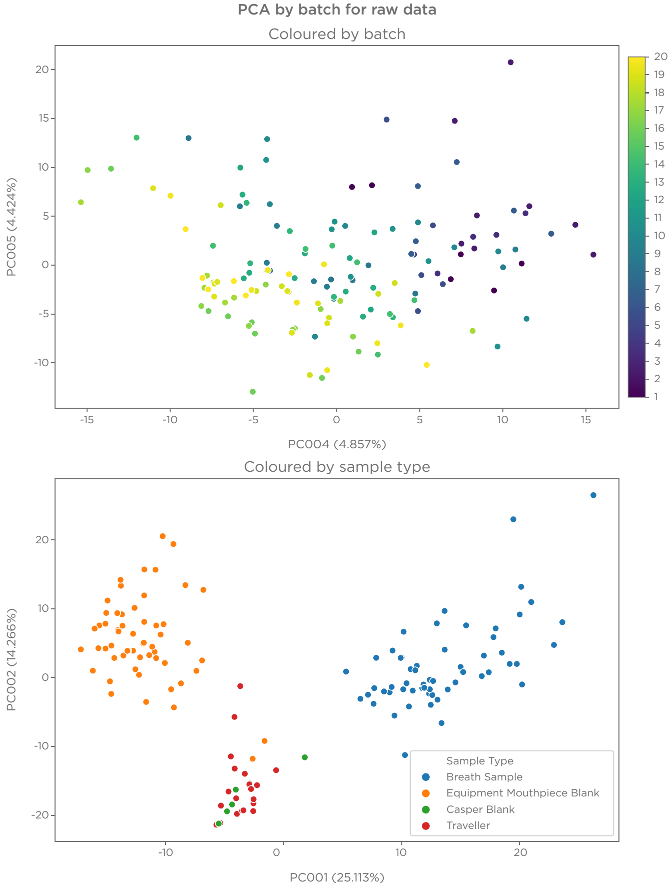


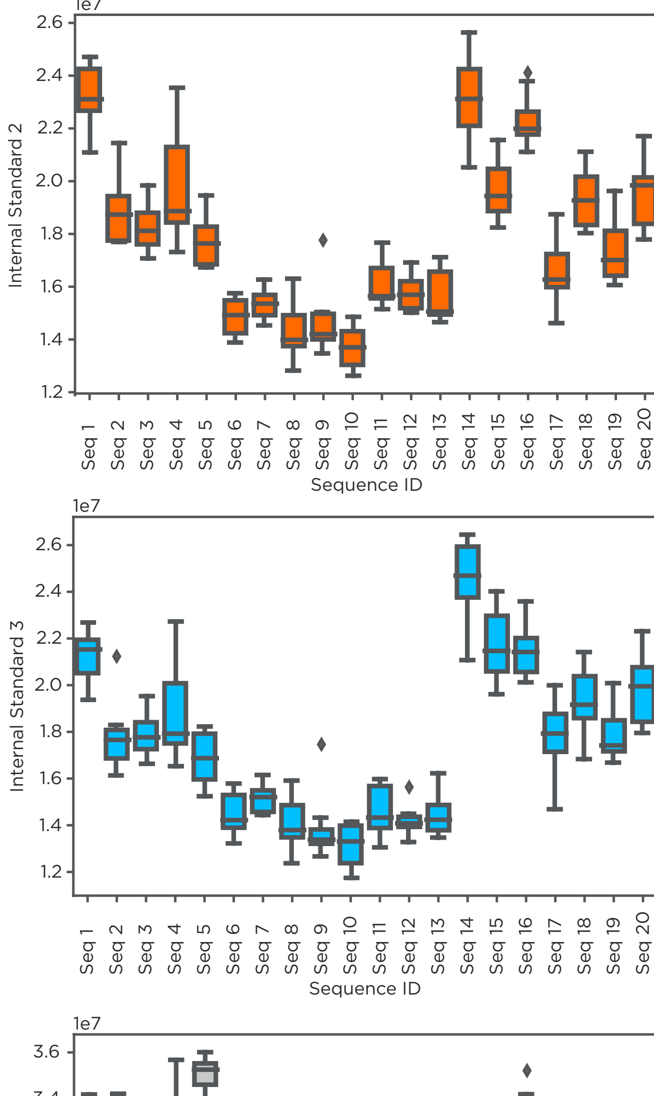
Figure 2: A PCA decomposition of the same example dataset post normalization colored by analytical sequence (above) and sample type (below).

Pitfalls to be avoided

Whilst calculating the median of IS peak areas is a reasonable approach to capture the analytical variability in the data, there are a number of inherent issues.

Firstly, the magnitude of IS peak areas may not be consistent across different compounds; one IS may consistently produce a peak area orders of magnitude higher than another and consequently, the variability of the largest peak area ISs will dominate any calculations of analytical variability. Moreover, the ranking of ISs by peak area may be unaffected by platform variability.

Utilizing median peak areas to derive scale factors necessitates the use of only one or two peaks in the final calculation, and while the other peaks may influence the selection of those areas, fluctuations in the other peaks are not captured in the scaling factors. For example, in the OMNI example dataset above, normalizing with the median of eight IS peak areas led to the same two ISs dictating the scaling factor for every VOC in each sample of the study. It is therefore unlikely that this method considers all the available information when correcting for the analytical variability of the platform.



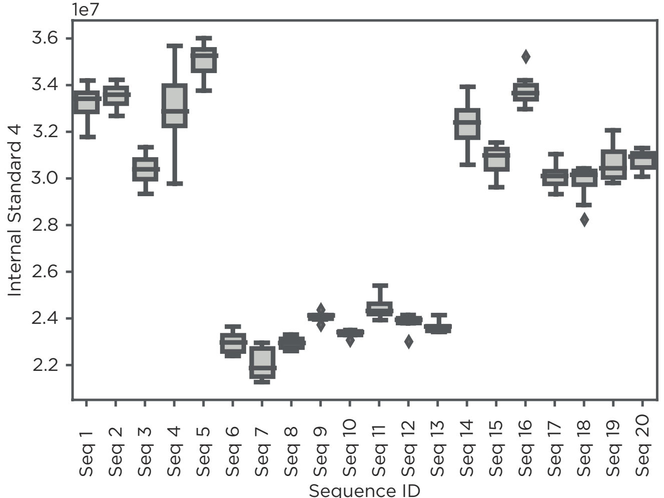


Figure 3: Internal standards over time in the OMNI example dataset.

Finally, both VOCs and ISs alike have a wide range of differing chemical traits and properties. It is unlikely that every VOC will have an identical response to analytical changes on the platform. This can be observed when comparing how ISs peak areas change over time during a study.

While the general trends and changes in response are comparable, there are clear differences in how these compounds behave. If these differences are due to chemical characteristics shared with other VOCs, we can much more accurately infer the effects of analytical variability on the response of those VOCs. This information can be captured in an improved normalization protocol.

A more sophisticated approach is needed

A wide range of methods have been developed for tackling the removal of analytical variability in mass spectrometry data, however, performances can vary greatly depending on the nature of the data being normalized and the general complexity of metabolomics data. Therefore, it is unlikely that a single, visual criteria (such as the PCA inspection discussed above) would accurately evaluate the success of normalization.

This initiated the Data Normalization Project to increase the normalization capability of the OMNI platform whilst creating the capacity to remove the variability associated with major analytical events, such as ion source changes. The project aimed to pull from existing literature and create a panel of normalization metrics specific to the OMNI platform, which can quantitatively assess the success of normalization with a greater degree of confidence.

Broad, commonly accepted criteria for assessing normalization performance have been laid out by Li et al. [1], and include:

- Reduction of intragroup variation among samples,
- Effect on differential metabolomic analysis,
- Consistency of identified metabolomic markers among partitions of data,
- Influence on classification accuracy, and
- Correspondence between normalized data and some reference.

The expansive panel of metrics generated by the Normalization Project spans all five of these categories. One such metric for measuring the reduction of intragroup variation is the reduction in variability of QC samples [2]. At Owlstone, check standards are evenly distributed throughout batches to ensure the accurate tracking of analytical drift. These contain ISs in fixed concentrations so successful normalization methods should greatly reduce the variability observed in these samples, both within and between batches.

Calibration curves are run at the start of every OMNI batch to ensure linear response, which gives more data with a known input to track analytical variability. A typical quality measure for calibration curves is the R-squared of the linear fit made using your calibration samples. After normalization, when truth should have been restored in your data, this should increase. In addition, analytical variability between batches should be removed by normalization, so a calibration curve created from randomly sampled calibration samples from all batches should also show an increase in R-squared compared to unnormalized data.

Further metrics examine the mutual information between analytical batch and sources of variation in the data, ensuring that any variations of interest (such as biological) are at the forefront of the data.

ISs are a strong candidate for the tracking of analytical variability. However, the 8 ISs respond differently to sources of analytical variation (as seen in Figure 3). The different responses might be due to the differing chemical properties of the ISs, and this association is not taken advantage of in the current normalization method; a VOC with similar chemical properties to IS1 is hypothesized to respond to analytical changes in the same way as IS1. Methods that can use any extra information about features (e.g., chemical properties) to select which ISs they should be normalized with are likely to perform better than methods which do not consider any additional properties.

A better way forward

In developing a new normalization method, both historical OMNI datasets and a specific experiment were used. The normalization experiment used both the primary and backup samples (2 pairs of sorbent tubes, collected during a single ReCIVA® sample), which should contain near identical concentrations of VOCs, to show that their similarity can be restored through normalization, even if a major analytical event has significantly changed the raw peak areas. This was done for both breath and ambient air samples, to ensure all potential matrices being run on the OMNI platform can be normalized.

Normalization Experiment Design:

Sequence 1

Sequence 2

Sequence 3

Run empty tubes

Sequence 4

Cold trap change

Sequence 5

Ion source change

Sequence 6

Figure 4: Process used in the normalization experiment. Sequence 1, 2, and 3 were taken under best case conditions and 4, 5, and 6 under worst case conditions. Sequences were normalized with the aim of restoring truth.

The data from this experiment was used to assess a wide range of normalization methods including some found in literature, and some developed in house. Using the criteria for normalization performance listed above, our bespoke panel of normalization metrics was applied to these methods and compared to the benchmark performance set by our previous protocol. The best normalization method, which will be used on future OMNI datasets going forward, gave a significant performance increase in every metric, validated over several datasets.

The new approach allows for more specific scaling factors by considering each VOC separately rather than one for a whole sample, and weighting the contribution of each IS to the scaling factors by the likelihood that the analytical variability of a VOC is captured by that IS. It calculates these weightings by considering statistical similarity and chemical similarity, i.e. chemical traits common between the two compounds.

When using the new method to normalize data, on average the variability (RSD%) of our check standards is reduced from 18.7% in unnormalized data to 7.9% between all analytical batches, and from 4.8% to 2.6% within batch. The R-squared of calibration curves ran within each batch, and calibration curves generated from randomly sampling calibration samples from all batches of a study, both increase (exceeding average R-squareds of 0.975 and 0.94 respectively).

Overall, the protocol generated by the Normalization Project successfully restores the truth in OMNI data, by more accurately capturing and removing analytical variability using our automated pipeline of methods and metrics.

Our constant innovation to further improve our platform is what makes OMNI the leading VOC biomarker discovery solution.

We look forward to sharing further advancements in the future and invite you to get in touch about incorporating Breath Biopsy into your research.

References

[1] Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., ... & Zhu, F. (2017). NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic acids research*, 45(W1), W162-W170.

[2] Drotleff, B., & Lämmerhofer, M. (2019). Guidelines for selection of internal standard-based normalization strategies in untargeted lipidomic profiling by LC-HR-MS/MS. *Analytical chemistry*, 91(15), 9836-9843.

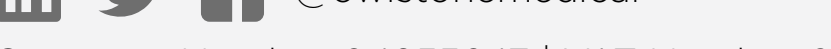
Contact us to find out more about breath biomarkers relevant to your area of interest and to discuss adding Breath Biopsy to your research.

breathbiopsy@owlstone.co.uk

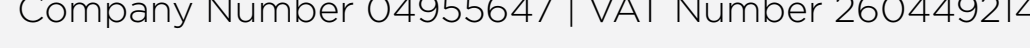
owlstonemedical.com

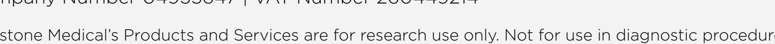
Owlstone Medical Ltd, 183 Cambridge Science Park,
Milton Road, Cambridge, CB4 0GJ, UK

@owlstonemedical



Company Number 04955647 | VAT Number 260449214





Owlstone Medical's Products and Services are for research use only. Not for use in diagnostic procedures.