thermoscientific

An Untargeted Metabolomics Approach to Using High Resolution Mass Spectrometry for Identifying Disease Biomarkers

Gina Tan[1], Svetlana Rezinciuc[2], Heather Smallwood[2], Andreas FR Hühmer[1], [1]Thermo Fisher Scientific, San Jose, CA, USA, [2]University of Tennessee Health Science Center, Memphis, TN, USA

ABSTRACT

Purpose: Untargeted metabolomics approach using an integrated MS platform and software solution to potentially identify disease biomarkers and demonstrate the possible applicability of this workflow to metabolite profiling studies on clinical research samples.

Methods: Donor nasal washes were analyzed using an LC-MS approach on a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer and the data generated were processed on a single software application that performed metabolite identification and differential analysis.

Results: The proposed workflow allowed the confident identification of over 600 compounds and the observation of sample variation in this study, enabling us to analyze statistically significant metabolites that could potentially be biomarkers for the disease research. All of this was achieved by using a suite of statistical tools within a single software application.

INTRODUCTION

Metabolomics has an increasingly important role in biomarker discovery as detection and quantification of metabolites has complementary value to genomics and proteomics studies. Observing metabolic changes occurring alongside gene alterations or protein activity related to a disease can provide deeper understanding of disease occurrence. Metabolite identification and profiling in biological samples are challenging tasks due to their diverse physiochemical nature and presence at different abundance levels. To address these challenges and gain better metabolite coverage, we pursued an LC-MS analytical approach which took advantage of high resolution mass spectrometry and comprehensive software tools. Greater accuracy and identification confidence can be achieved, and here we demonstrate the potential application of this untargeted metabolomics workflow to actual clinical research biological samples.

MATERIALS AND METHODS

Sample Preparation

Cells from nasal washes were collected from 7 pediatric donors diagnosed with respiratory viral infections by PCR. Briefly, cells were washed twice with ice-cold buffered saline solution to remove medium components and then rinsed with ice-cold Milli-Q® water (MilliporeSigma). After removal of the water, tubes with cells were frozen with liquid nitrogen, and 1 mL of ice-cold 9MeOH:1$CHCl_3$ was added to each tube and cells were vortexed vigorously. Extracts were pelleted at 4°C for 4 minutes at 16 000g. Supernatants were then passed through Millipore® filters (molecular weight cutoff MWCO 3 K) to remove cell debris and transferred to new micro centrifuge tubes. Metabolite extraction using 3-fold of methanol was performed on all samples, and these extracts were further filtered through a 3,500 MWCO filter to remove any salt content before LC-MS analysis.

Liquid Chromatography

LC separation was carried out with the following conditions:

Table 1. UHPLC and column information.

LC system	Thermo Scientific™ Dionex™ UltiMate™ 3000 RS system
Column	Thermo Scientific™ Hypersil GOLD™ aQ analytical column
Column dimensions	C18, 100 x 3 mm, 1.9µm particle size, reverse-phase
Column temperature	55ºC
Mobile phases	A: 0.1% formic acid in water; B: 0.1% formic acid in methanol
Sample injection volume	4µl

Table 2. UHPLC gradient used.

Time (min)	%A	%B	Flow rate (µl/min)
0.0	99.5	0.5	300
5.5	50.0	50.0	300
6.0	2.0	98.0	300
12.0	2.0	98.0	300
13.0	99.5	0.5	300
15.0	99.5	0.5	300

Mass Spectrometry

All MS data were acquired on the Orbitrap Fusion mass spectrometer, equipped with a HESI source. The samples were analyzed in ESI positive polarity at 120,000 resolution using a fine-isotopic determination data-dependent MS^2 acquisition method. Specifically, a full MS (120K resolution) followed by a data-dependent MS^2 (15K resolution) scan mode was used. Scan range was set to 70 – 1050 m/z and default charge state was 1. AGC target and maximum injection time was 1e6 and 100ms respectively for full scan, 5e4 and 50ms respectively for data-dependent MS^2 scan.

Data Analysis

All acquired MS data were processed using the Thermo Scientific™ Compound Discoverer™ 2.1 software to carry out metabolite identification in the nasal wash samples. The untargeted metabolomics workflow was selected to perform retention time alignment, component detection and elemental composition prediction on the data files followed by identification using spectral matching on the mzCloud™ database and chemical formula matching against the ChemSpider™ database, an online chemical information resource (Royal Society of Chemistry). Differential analysis was done on the sample set to observe sample variance (Figure 2).

Figure 1. Experimental workflow.

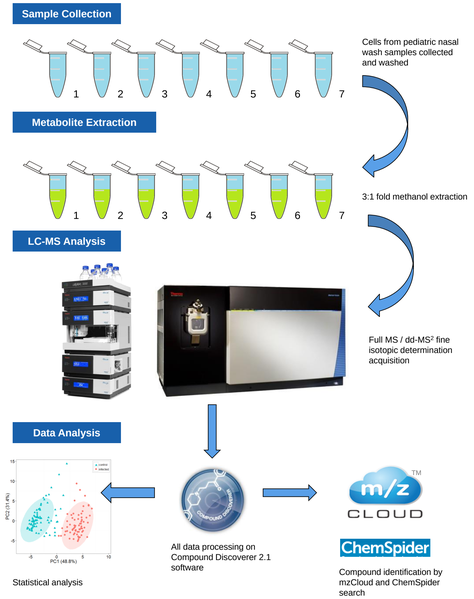
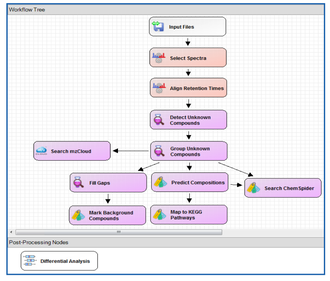


Figure 2. Untargeted metabolomics identification workflow for processing the .RAW input files on Compound Discoverer 2.1 software.



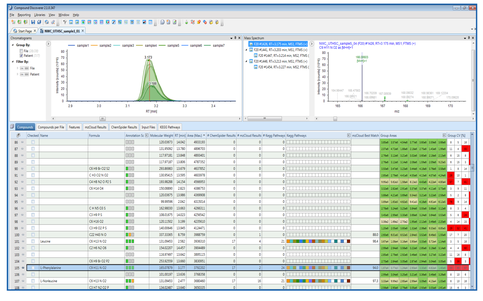
RESULTS

Metabolite Identification with Compound Discoverer Software

A table of results was generated which included assigned elemental composition, annotation source, results scores, p-values, etc. which are essential and useful for interpretation and identification (Figure 3). From the 7 nasal wash samples, 604 compounds were detected, of which 528 had positive assignments from ChemSpider search and 225 were positively identified via mzCloud (Figure 4). By searching against a highly curated spectral repository such as mzCloud, greater identification confidence is achieved. The inclusion of an automated mzCloud similarity search functionality in the processing step provided more mzCloud IDs as potential metabolites of modified compounds were identified as well. Over 60% of the compounds identified had a mzCloud match score of over 80, suggesting the possible presence of these metabolites in the samples.

Initial analysis of the chromatographic profile for the 7 samples showed distinct chromatographic peak features with the exception of sample 5 which gave a different separation profile, suggesting the need to further investigate patient 5's sample to determine what contributed to this difference and/or if there were significant differences in the metabolite profile as well (Figure 5) .

Figure 3. Overall results table of processed data on Compound Discoverer 2.1 software, displaying chromatographic overlay and mass spectra of a selected potential metabolite present in the nasal wash samples.



Statistical Analysis

The PCA plot showed a clustering of samples 1, 2, 3 and another set of cluster for samples 4, 6, 7. Sample 5 was clearly separated from the other samples and aligned with our previous observation that this sample showed variation and behaved differently to the rest (Figure 6). We further investigated this statistical behavior by observing the S-plot to determine which metabolites contributed to a higher covariance intensity and high correlation reliability (Figure 7), as it would indicate these identified metabolites give a higher possibility to be putative biomarkers for the respiratory virus infected samples and could potentially be selected for monitoring in a targeted metabolite assay.

Figure 4. Representation of number of metabolite compounds identified from a Compound Discoverer search.

Figure 5. Chromatographic profiles of 7 nasal wash samples with sample 5 showing a different chromatographic separation behavior from the other samples.

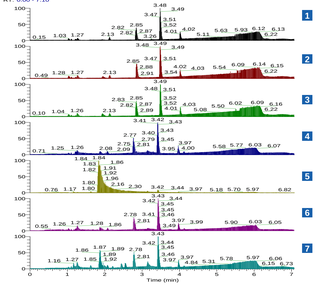
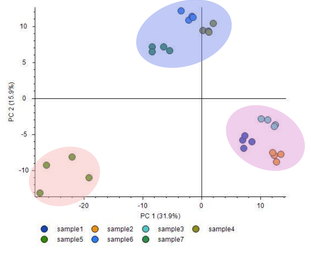


Figure 6. PCA plot showing a clear divergence of donor sample 5 with the remaining samples.



Potential Metabolites as Putative Biomarkers

From the S-plot, 8 metabolites that showed significant covariance and correlation were selected and further analyzed by observing the trend plot for any significant behavior to give us insight on these patient samples. We noticed a consistent trend for all 8 metabolites: a huge spike in abundance of the metabolite present only in sample 5 and reliable amounts present in sample 7 (Figure 8). Further comparison experiments will be required to determine if this set of metabolites are indeed indicative markers for respiratory infected virus.

The use of several available statistical tools in one software package allowed us to seamlessly review the results and determine patient sample behavior by observing any statistical differences within the sample group. Identified metabolites of statistical significance were shortlisted as potential metabolite biomarkers from the list of compounds detected. This serves as a starting point to evaluate the eligibility of this set of metabolites to be developed into a targeted assay relative to the disease of interest, in subsequent experiments.

Figure 7. S-plot generated from Compound Discoverer software highlighting the potential metabolites (blue dots) with statistical significance selected for further analysis.

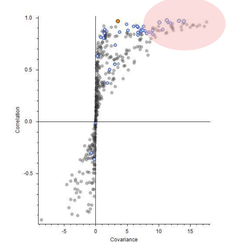
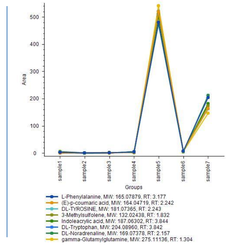


Figure 8. Trend plot of selected 8 metabolites that shows substantial amounts present consistently in samples 5 and 7.



CONCLUSIONS

For Research Use Only. Not for use in diagnostic procedures.

- Integrated instrument platform and intuitive software solution in a single workflow facilitates ease of data acquisition, data processing and interpretation.
- Metabolites with good intensities and statistical significance were identified as potential biomarkers that could be further monitored for the study.
- A larger cohort study with greater sample diversity, disease and treatment variation will be required next to draw more conclusive results for determining disease disparity and more accurately identify putative biomarkers for the disease of interest.

ACKNOWLEDGEMENTS

We would like to thank the University of Tennessee Health Science Center for providing the pediatric nasal wash samples.

TRADEMARKS/LICENSING