# Non-targeted differential screening using LC-HRAM-MS as a tool for evaluating differences in chemical composition between samples and groups

Daniel Arndt, Arno Knorr , Mark Bentley

Philip Morris International R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland (part of Philip Morris International group of companies)

## Introduction

Mass-spectrometry based non-targeted screening (NTS) is a key methodology for characterizing the chemical composition of complex matrices using an unbiased approach. Very often the focus is upon the comparison of two or more samples, to evaluate any significant differences in chemical composition between samples in an unsupervised way or if group related pre-knowledge is available between sample groups. For this comparative approach, non-targeted differential screening (NTDS) is performed, where relative differences in abundance for constituents, as well as semi-quantitative estimates of absolute abundance, are considered. Due to the highly complex nature in chemical composition of biological samples including tobacco aerosol samples containing several thousand constituents, the extraction of significant differences can be very challenging. A complementary differential screening approach using liquid chromatography coupled to high resolution accurate mass spectrometry (LC-HRAM-MS) in parallel with two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (see also Poster TP 550, Almstetter et al., 64th ASMS 2016) has been developed in order to ensure comprehensive analytical coverage for identifying the most relevant differences in composition between diverse samples. The use of full scan LC-HRAM-MS combined with first order fragmentation ($MS^2$) data assists with structural elucidation of unknown compounds. Statistical models are used to filter significantly different compounds, followed by a ranking procedure which considers the relative differences in abundance of each compound as well as the absolute abundance. The findings are then passed on for toxicological assessment or used to assist product development.

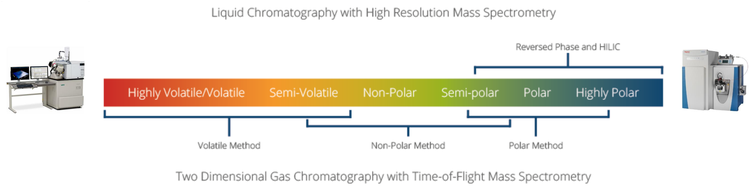


*Figure 1: Analytical Coverage of Non-Targeted Differential Screening (NTDS)*

## Methods

LC-HRAM-MS in full scan mode combined with high-energy collision dissociation (HCD) using stepped normalized collision energy (NCE) was applied using a Q Exactive™ (Thermo Fischer Scientific, Bremen, Germany) in both reversed phase (RP) and hydrophilic interaction chromatography (HILIC) modes, applying heated electrospray (HESI) and atmospheric pressure chemical (APCI) ionization modes (RP HESI positive, RP HESI negative, RP APCI positive and HILIC HESI positive) to cover the broadest possible range of compounds. Sample replicates ($n$=5) of different tobacco product aerosols were fortified with a set of stable isotope labeled internal standards (*Table 1*) to enable semi-quantification. Raw data analysis was followed by advanced data processing using metabolomics software tools for generic peak finding, deconvolution and peak alignment. Compound identification was performed semi-automatically using database confirmation of $MS^2$ fragments and retention times. In addition, experimental $MS^2$ fragments were compared with *in-silico* predicted fragments derived from different databases. Significant differences were filtered using Student's t-test and ranked using an in-house developed procedure considering the relative differences in abundance of each compound as well as a semi-quantitative estimate of absolute abundance (NTDS RANK)[1].

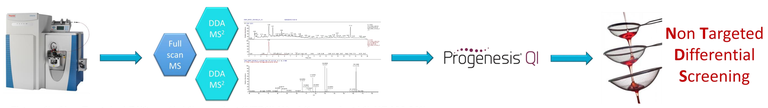


*Figure 2 : Non Targeted Differential Screening (NTDS) Workflow using LC-HRAM-MS*

**Reversed Phase Modes**

LC separation was performed using a Hypersil GOLD™ column (150 x 2.1mm, 1.9µm; Thermo Scientific, Waltham, MA, USA) with a UHPLC guard filter cartridge (10 x 2.1mm, 0.2 µm; Thermo Scientific, Waltham, MA, USA) operating at 50 °C with a 1.5 µL injection volume. For RP HESI positive and RP APCI positive chromatographic separation the mobile phase consisted of A: 10mM ammonium acetate and B: 1mM ammonium acetate in methanol. For RP HESI negative chromatographic separation the mobile phase consisted of A: 1mM ammonium fluoride and B: methanol. Gradient elution was performed with a constant flow rate of 400 µL $min^{-1}$ according to the program described in *Table 2*.

**HILIC Mode**

LC separation was performed using an Accucore HILIC™ column (150 x 2.1mm, 2.6µm; Thermo Scientific, Waltham, MA, USA) with a HILIC Defender guard cartridge (10 x 2.1mm, 2.6 µm; Thermo Scientific, Waltham, MA, USA) operating at 50 °C with a 1.5 µL injection volume. For HILIC HESI positive chromatographic separation the mobile phase consisted of A: 10mM ammonium acetate and B: 10mM ammonium acetate in acetonitrile. Gradient elution was performed with a constant flow rate of 500 µL $min^{-1}$ according to the program described in *Table 3*.

| Method | ISTD | Formula |
|---|---|---|
| RP HESI pos. | d8-Isophorone | $C_9H_6D_8O$ |
| RP HESI neg. | Decanoic-d19 acid | $C_{10}HD_{19}O_2$ |
| RP APCI pos. | d8-Isophorone | $C_9H_6D_8O$ |
| HILIC HESI pos. | Myosmine-d4 | $C_9H_6D_4N_2$ |

*Table 1: Internal Standards (ISTDs)*

| Time [min] | A [%] | B [%] |
|---|---|---|
| 0 | 85 | 15 |
| 7.00 | 10 | 90 |
| 12.80 | 0 | 100 |
| 18.00 | 0 | 100 |
| 18.10 | 85 | 15 |
| 20.00 | 85 | 15 |

*Table 2: Gradient RP Modes*

| Time [min] | A [%] | B [%] |
|---|---|---|
| 0 | 2 | 98 |
| 7.00 | 25 | 75 |
| 8.00 | 2 | 98 |
| 15.00 | 2 | 98 |

*Table 3: Gradient HILIC Mode*

**HRAM Detection**

Ionization was performed using HESI in both positive and negative (RP HESI positive, RP HESI negative, HILIC HESI positive) mode and atmospheric pressure chemical ionization (APCI) in positive mode (RP APCI positive). HRAM Full scan MS was performed at a resolution of 70.000 (FWHM), acquiring a mass range of $m/z$ 80 – 800 in combination with a data dependent $MS^2$ Top3 of each scan at a resolution of 17.500 (FWHM) and applied stepped normalized collision energies of 25, 50 and 75 eV and automated gain control of 1 x $10e^5$ in order to generate HCD first order fragmentation (TopN = 3, loop count = 3, dynamic exclusion = 10 s). Vaporizer heater temperature, capillary temperature, spray voltage, sheath gas and auxiliary gas were set at 350 °C, 380 °C, ±3.00 kV, 60 and 20 arbitrary units respectively for HESI modes. Vaporizer heater temperature, capillary temperature, discharge current, sheath gas and auxiliary gas were set at 450 °C, 380 °C, 5.0 µA, 50 and 5 arbitrary units respectively for APCI Mode.

## Results

**Non-targeted Screening & Compound Identification**

Unbiased non-targeted screening of combined full scan ($MS^1$) and data dependent Top3 ($MS^2$) data was performed using Progenesis QI™ (Nonlinear Dynamics, Newcastle upon Tyne, UK) comprising raw data import, selection of possible adducts, alignment, peak picking, compound identification and normalization with ISTDs. 5 replicates of each sample were injected (Aerosol of Tobacco Product 1, Tobacco Product 2, Blank, Pool Sample). Non-targeted screening resulted in >15,000 detected compounds.

*Figure 3: Progenesis QI™ Data Evaluation*

Compound identification was performed using a semi-automatic stepwise approach employing an experimental $MS^2$ fragmentation database and *in-silico* predicted fragmentation of chemicals from public databases. In Step 1 all detected constituents were matched and assigned against an in-house database comprising experimental data for approximately 400 reference compounds with accurate mass data, stepped NCE $MS^2$ first order fragmentation and retention times (precursor and fragment tolerance 5ppm, retention time tolerance 0.5 min). In Step 2 the fragmentation patterns for all detected constituents were compared with *in-silico* predicted fragments for UCSD (Unique Compounds & Spectra Database, PMI, Neuchâtel, Switzerland)[2], HMDB 3.6 (Human Metabolome Database, University of Alberta, Edmonton, Canada)[3,4,5] and via Chemspider search plugin with data sources of ChemIDplus (ChemIDplus, SIS, NLM, NIH, Bethesda, MD, USA) and FDA (U.S. Food and Drug Administration, Silver Spring, MD, USA) (precursor and fragment tolerance 5ppm). In Step 3 fragmentation spectra for detected constituents were compared with experimental fragmentation spectra of NIST14 MS/MS library (precursor and fragment tolerance 5ppm) (U.S. National Institute of Standards and Technology, Gaithersburg, MD, USA). All putative hits were scored using Progenesis QI™ algorithms, which considered mass similarity, isotope similarity and fragmentation score.

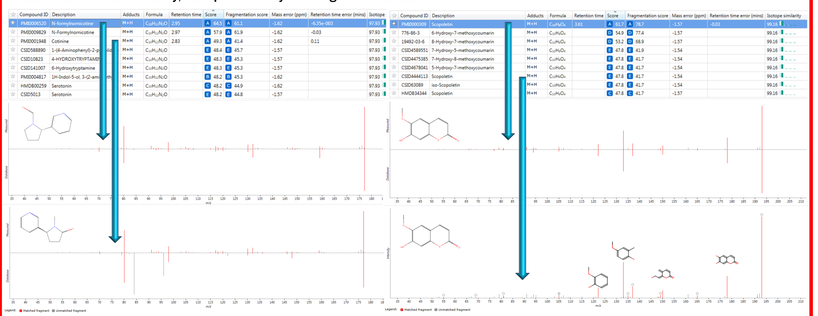

*Figure 4: Discriminatory Power of Experimental Fragment Database*



*Figure 5: Experimental vs. in-silico Fragmentation*

**Non-targeted Differential Screening (NTDS)**

In order to identify significant differences (compounds of interest) between independent data sets aligned and normalized results from non-targeted screening are processed by applying a two-tailed distributed heteroscedastic Student's t-test (2 groups, 5 replicates = 10 observations). Results that yield p values > 0.05 were not considered statistically different and were therefore excluded from further analysis.

To consider the relevance of each finding, significant different compounds are ranked according to the relative difference in abundance (x-fold change) and the semi-quantitatively estimated absolute abundance based on peak area ratios between analyte and the assigned internal standard with known concentration (i.e., the greater the difference and absolute abundance, the greater the relevance). The sorting of obviously different compounds (variables) by their relevance is done by applying an empirically developed formula (RANK) on the t-test filtered data sets.

The RANK formula mathematically combines two criteria: 1. Abundance of the variable ("Average Concentration for a pre-defined group [µg/sample]) and 2. Relative difference of the variable ("Effect" [%]). $L_x$ is the measured concentration value for sample x to be compared with sample y and $L_y$ is the measured concentration value for sample y to be compared with sample x.

$$\%Effect = \frac{Ly-Lx}{Ly+Lx} * 100, \qquad Index = \frac{\%Effect^3}{1000}, \qquad RANK = \frac{Index \times (Lx+Ly)}{2}$$

The data set is sorted in the order of decreasing RANK-values resulting in a table with negative RANK values (compounds in sample group X elevated compared to sample group Y) and positive RANK-values in increasing order (compounds in sample group Y elevated compared to sample group X). A fused data table is created showing the proposed compound name, structure, formula, $m/z$, compound identifier, retention time, abundance, semi-quantitative concentration, x-fold change, detected adducts, p-value, %-Effect, RANK value and the analytical method. The accumulated RANK value gives a hint about the overall differences between the sample groups.

| # | Proposed Compound Name | Structure | Formula | m/z (M + H / M - H) | Compound Identifier | Retention time (min) | Tobacco Product 1 mean conc. (µg/item) | Tobacco Product 1 Abundance | Tobacco Product 2 mean conc. (µg/item) | Tobacco Product 2 Abundance | TP 1 % of TP 2 (TP1>100%) | X-fold change TP 2 > TP 1 | Adducts | T-Test | % Effect | RANK Value | Analytical Method |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Lanost-8-en-3-ol, 24-methylene-, (3β) | | C31H52O | 441.40904 | PMY0006771 | 13.74 | 255197 | 6.30 | 65187 | 1.61 | 391 | 3.91 | M+H, M+NH4 | 1.08E-09 | 208.60 | 825.11 | RP ESI pos |
| 2 | (3,14-Labadiene-7,8-diol, (7R,8a,13Z) | | C20H34O2 | 307.26300 | PMY0005787 | 10.64 | 129496 | 1.43 | 5753 | 0.06 | 2194 | 22.34 | M+H | 7.36E-13 | 764.25 | 176.66 | RP APCI pos |
| 3 | Isolinderanolide | | C21H30O3 | 337.27343 | HMDB38105 | 9.63 | 202184 | 4.99 | 74968 | 1.85 | 270 | 2.70 | M+H | 3.55E-10 | 96.71 | 330.91 | RP ESI pos |
| 4 | Ethyl 2,4-dioxohexanoate | | C8H12O4 | 173.08080 | PMY0019568 | 1.40 | 277702 | 6.73 | 144434 | 3.57 | 189 | 1.89 | M+H | 1.26E-06 | 29.09 | 149.74 | RP ESI pos |
| 5 | Benzoic acid, 2,5-dihydroxy-, methyl | | C9H10O4 | 183.06512 | PMY0004649 | 2.05 | 184165 | 4.55 | 88089 | 2.18 | 209 | 2.09 | M+H | 5.38E-08 | 43.95 | 147.71 | RP ESI pos |
| 6 | Ergosterol | | C28H44O | 397.34660 | PMY0006710 | 11.13 | 128837 | 3.18 | 63869 | 1.58 | 202 | 2.02 | M+H | 1.80E-07 | 38.32 | 91.33 | RP ESI pos |
| 7 | 13-Labdene-8,15-diol, (8a,13E)-form | | C20H36O2 | 309.27835 | PMY0001846 | 11.20 | 72158 | 0.14 | 719 | 0.01 | 1691 | 16.91 | M+H | 1.09E-12 | 701.09 | 50.21 | RP APCI pos |
| 8 | Labdane-8,15-diol, (13S) | | C20H38O2 | 311.29403 | PMY0008387 | 11.83 | 12889 | 0.14 | 1322 | 0.01 | 975 | 9.75 | M+H | 8.84E-11 | 539.27 | 42.62 | RP APCI pos |
| 9 | Glycidyl acetate | | C5H8O3 | 117.05494 | PMY0007698 | 1.15 | 182090 | 4.45 | 125807 | 3.11 | 143 | 1.43 | M+H | 6.62E-06 | 5.66 | 21.43 | RP ESI pos |
| 10 | Pyranone | | C6H8O4 | 145.04935 | PMY0002228 | 1.49 | 588240 | 6.54 | 455387 | 5.07 | 129 | 1.29 | M+H-H2O, M+H | 2.29E-05 | 2.66 | 11.97 | RP APCI pos |
| 11 | 5-Methylfurfural | | C6H6O2 | 111.04435 | PMY0000001 | 3.08 | 89474 | 1.00 | 56799 | 0.63 | 158 | 1.58 | M+H | 1.47E-07 | 11.15 | 9.07 | RP APCI pos |
| 12 | Isoquinoline, 3-methyl | | C10H9N | 144.08074 | PMY0003968 | 2.39 | 254738 | 6.29 | 203912 | 4.99 | 126 | 1.26 | M+H | 3.35E-05 | 1.92 | 8.73 | RP ESI pos |
| 13 | Pyridoxin | | C8H11NO3 | 170.08308 | PMY0002009 | 2.05 | 404234 | 0.70 | 304053 | 0.53 | 133 | 1.33 | M+H | 1.12E-05 | 2.82 | 1.73 | HILIC ESI pos |

*Figure 6: Example for a Fused Data Result Table*

## Conclusion

An unbiased non-targeted screening approach for assessing significant differences in chemical composition of sample groups has been developed, which considers the changes of relative differences in abundance of each compound as well as the absolute abundance (i.e., the greater the difference and absolute abundance, the greater the relevance) enabled by the application of the NTDS RANK formula. The use of LC-HRAM-$MS^n$ enables semi-automatic identification of compounds and comparison with *in-silico* fragmentation strengthens the ability for structural elucidation of unknown compounds. Finally the implementation of accurate mass databases simplifies compound identification in non-targeted screening methods.

NTDS using full scan LC-HRAM-$MS^n$ is a powerful non-targeted screening tool for the elucidation of novel compounds. In combination with the complementary NTDS GCxGC-TOF-MS workflow this methodology demonstrates a powerful comparative technique to determine whether any new or unexpected chemical compounds are present in developmental products.

## References

1.Knorr A, A method of analyzing data from chemical analysis, International Patent WO 2013098169 A1, PCT/EP2012/076244, 2013. Jul 4
2.Martin E, Monge A, et al., Building an R&D chemical registration system, Journal of Cheminformatics 2012 4:11, DOI: 10.1186/1758-2946-4-11
3.Wishart DS, Tzur D, Knox C, et al., HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007 Jan;35(Database issue):D521-6
4.Wishart DS, Knox C, Guo AC, et al., HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009 37(Database issue):D603-610.
5.Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al., HMDB 3.0 — The Human Metabolome Database in 2013. Nucleic Acids Res. 2013. Jan 1;41(D1):D801-7.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL
www.pmiscience.com

64th ASMS Conference on Mass Spectrometry and Allied Topics
San Antonio, TX, USA
June 5 – 9, 2016

**Competing Financial Interest**

The research described in this poster was funded by Philip Morris International