# An automated workflow for rapid alignment and identification of lipid biomarkers obtained from chip-based direct infusion nanoelectrospray tandem mass spectrometry

**Jens Hoefkens (1), Tobias Kind (2), Kent Pinkerton (3), Oliver Fiehn (2)**

(1) Genedata Inc, Waltham, MA; (2) UC Davis - Metabolomics, Davis, CA; (3) UC Davis Center for Health and the Environment, Davis, CA

## Introduction

Polar lipids (glycerophospholipids, saccharolipids) are routinely analyzed with chromatographic and mass spectrometric techniques. Chip-based direct infusion nanoelectrospray tandem mass spectrometry can provide lipid profiles or lipid fingerprints with short infusion times (one to two minutes). Using high-resolution FT-ICR-MS or low-resolution linear ion trap mass spectra in positive and negative mode together with data dependent tandem mass spectrometry, polar lipids can be semi-quantified and identified. The major challenge is however not the analysis itself but automated data handling and data evaluation of the obtained results. We present efforts which include a reusable workflow environment for rapid alignment and identification of lipids from infusion tandem mass spectra. Results from an environmental tobacco smoke study and plasma and lung tissue are shown.

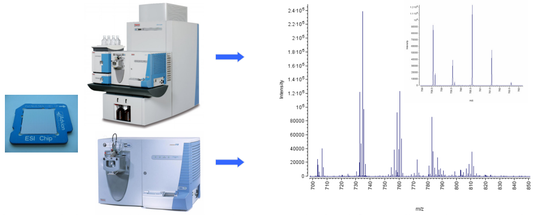
## Methods

To 30 µL plasma or 10 mg cryogenically ground lung samples 1.0 mL of chilled (isopropanol:acetonitrile:water (3:3:2; v/v/v; -20°C) solvent was added. The solution was shaken for 5 minutes in a vortexer and centrifuged 2.5 minutes at 13,000 rpm. After adding 10 µL of 100 mM ammonium acetate to 90 µL of the supernatant the lipid extracts were infused into a linear ion trap mass spectrometer coupled to an FT-ICR-MS using a Advion Nanomate chip-based infusion system. Mass spectra were acquired from 400-1000 m/z with positive ionization and negative ionization and 100,000 resolving power and 0V CID voltage. Data dependent MS2 scans were performed on the 40 highest peaks in the full scan spectrum and scanned in low-resolution mode with 30V CID voltage. The data alignment of the high-resolution and low-resolution mass spectra was performed using Genedata Expressionist Refiner MS. The statistical data evaluation was performed with Genedata Analyst using univariate and multivariate methods. Unique lipid species were identified by matching tandem mass spectra against a database of known fragment ions and validating results with LipidMaps tools and LipidMaps web services.

## Results

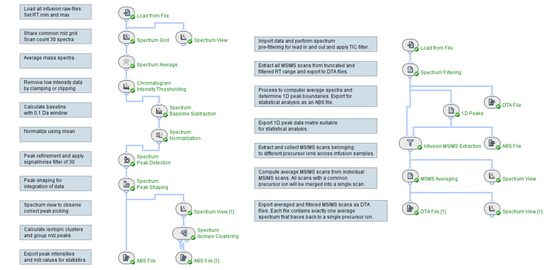
Lipid extracts from heart, plasma and lungs from rats which were exposed to environmental tobacco smoke and clean air were analyzed. An alignment workflow for high-resolution and low-resolution mass spectra was built using Expressionist Refiner MS. It included spectrum grid alignment, spectrum averaging, intensity thresholding, peak detection, peak shaping and alignment export. The post-processing time for a sample setup of 70 infusion experiments took around 20 seconds. Additionally MS/MS spectra were extracted from data dependent scans. A matrix containing all samples and all aligned ions above a certain signal noise threshold was obtained as a result file. Identification of lipid species was performed by matching specific sn1 and sn2 fatty acid chain fragmentation from MS/MS data of the product ions. 50 lipid species were identified in plasma and 50 remained unknown. Different lipid patterns revealed alterations during exposure to tobacco smoke. This research project was supported by R01 ES013932 NIH/NIEHS.

## Platform – nanoESI infusion with iontrap



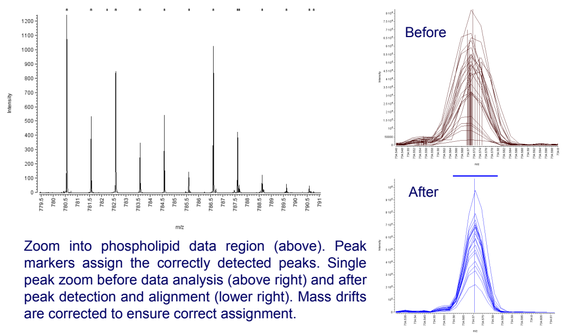

Low-resolution and high-resolution data were obtained from a linear ion trap and a linear iontrap coupled to an Fourier transform (FT-ICR-MS) mass spectrometer. Infusion of full extracts from lung, plasma and heart samples was performed with a chip-based nanoelectrospray robot (Advion Nanomate). Infusion time per sample with data dependent MS/MS scans is two minutes to obtain about 100 mass spectra.

## Mass spectral alignment with Refiner MS



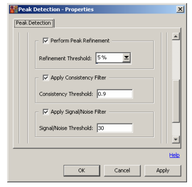

Alignment of low-resolution and high-resolution mass spectral infusion data was implemented within Expressionist Refiner MS. The workflow performs peak detection, noise analysis, peak integration, isotope grouping and alignment of large datasets. The workflow runtime for 70 infusion samples is 20 seconds, allowing fast parameter optimizations. MS/MS data extraction is performed to obtain tandem mass spectra for assignment of phospholipid structures including head group and neutral loss analysis.

## Alignment of infusion mass spectra




Zoom into phospholipid data region (above). Peak markers assign the correctly detected peaks. Single peak zoom before data analysis (above right) and after peak detection and alignment (lower right). Mass drifts are corrected to ensure correct assignment.

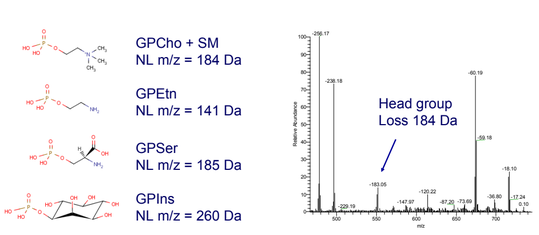
## Parameter optimization and peak refinement




| Name | Initial Peaks | After Peak Refinement | After Consistency Filter | After S/N Filter |
|---|---|---|---|---|
| K7823 | 526 | 546 | 239 | 177 |
| K7824 | 645 | 662 | 304 | 195 |
| K7825 | 415 | 423 | 234 | 174 |
| K7826 | 743 | 761 | 321 | 207 |
| K7827 | 884 | 893 | 403 | 258 |
| K7828 | 323 | 326 | 200 | 155 |
| K7829 | 756 | 769 | 485 | 325 |
| K7830 | 587 | 601 | 292 | 197 |
| K7831 | 782 | 803 | 408 | 247 |
| K7832 | 687 | 698 | 318 | 235 |
| K7833 | 577 | 588 | 303 | 197 |
| K7834 | 836 | 848 | 387 | 255 |
| K7835 | 363 | 374 | 177 | 136 |

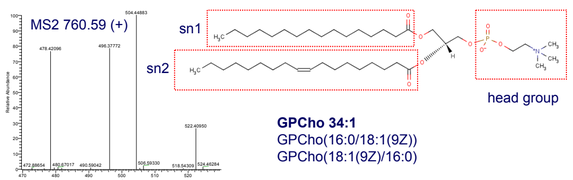
Depending on the mass spectral data acquisition mode, certain workflow parameters can be optimized to avoid peak detection errors or to adjust sensitivity if needed. The table shows a very high number of peaks after the first initial step. The recursive consistency filter and the signal to noise filter ensure to retain peaks which have a very low abundance but good signal-to-noise (S/N) ratio.

## Data dependent MS/MS for class analysis




The analysis of the different phospholipid classes including Glycerophosphocholines (GPCho), Sphingomyelines (SM), Glycerophosphoethanolamines (GPEtn), Glycero-phosphoinositols (GPIns), Glycerophosphoserines (GPSer) and others can be performed by subtraction of specific neutral losses from the precursor ion in positive and negative mode. The precursor mass is then used to compute or lookup only species from a specific lipid class. In case of overlaps or unclean spectra multiple species can occur.

## sn1 and sn2 fatty acid analysis from MS/MS




**GPCho 34:1**
GPCho(16:0/18:1(9Z))
GPCho(18:1(9Z)/16:0)

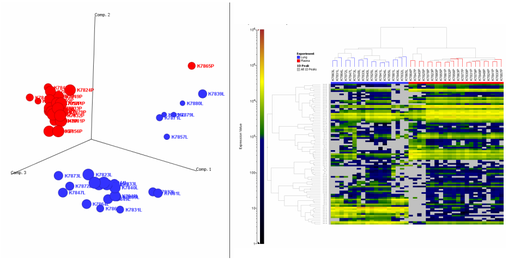
Zoom into the MS/MS data region at 35 eV CID. The four peaks refer to the loss of the sn1 alkyl or acyl group and sn2 group loss. The peaks with delta -18 Da are due to loss of water. The MS/MS analysis additionally can be used to check for correct assignment of possible adducts (+H) (+Na) which would result in different fragmentation patterns. A further assignment of the sn1 and sn2 group would require a MS3 step. In many cases the data dependent MS2 scans are not clean due to selected isolation width of 2 Da and overlap from different lipid species. In this case a quality factor is assigned (low/high).

## Identified phospholipids from rat lung

| # | m/z | Name | # | m/z | Name |
|---|---|---|---|---|---|
| 1 | 706.540 | GPCho 30:0 | 26 | 746.570 | GPEtn 36:1 |
| 2 | 734.569 | GPCho 32:0 | 27 | 746.607 | GPEtn 37:0 |
| 3 | 732.555 | GPCho 32:1 | 28 | 772.526 | GPEtn 38:2 |
| 4 | 730.539 | GPCho 32:2 | 29 | 768.555 | GPEtn 38:4 |
| 5 | 728.522 | GPCho 32:3 | 30 | 766.536 | GPEtn 38:5 |
| 6 | 724.528 | GPCho 32:5 | 31 | 788.619 | GPEtn 39:1 |
| 7 | 762.599 | GPCho 34:0 | 32 | 786.602 | GPEtn 39:2 |
| 8 | 760.586 | GPCho 34:1 | 33 | 794.607 | GPEtn 40:5 |
| 9 | 758.571 | GPCho 34:2 | 34 | 790.561 | GPEtn 40:7 |
| 10 | 756.553 | GPCho 34:3 | 35 | 746.607 | GPEtn(O) 37:1 |
| 11 | 788.619 | GPCho 36:1 | 36 | 752.560 | GPEtn(O) 36:4 |
| 12 | 786.602 | GPCho 36:2 | 37 | 754.537 | GPEtn(O) 38:4 |
| 13 | 784.587 | GPCho 36:3 | 38 | 680.482 | GPSer 28:0 |
| 14 | 782.570 | GPCho 36:4 | 39 | 790.561 | GPSer 36:1 |
| 15 | 780.552 | GPCho 36:5 | 40 | 812.544 | GPSer 38:4 |
| 16 | 810.602 | GPCho 38:4 | 41 | 622.409 | PC 24:0 |
| 17 | 808.585 | GPCho 38:5 | 42 | 610.542 | PC 24:6 |
| 18 | 806.569 | GPCho 38:6 | 43 | 636.558 | PC 25:0 |
| 19 | 836.619 | GPCho 40:5 | 44 | 650.441 | PC 26:0 |
| 20 | 834.603 | GPCho 40:6 | 45 | 638.573 | PC 26:6 |
| 21 | 832.586 | GPCho 40:7 | 46 | 703.577 | SM 16:0 |
| 22 | 830.566 | GPCho 40:8 | 47 | 759.630 | SM 20:0 |
| 23 | 720.555 | GPEtn 34:0 | 48 | 787.671 | SM 22:0 |
| 24 | 718.576 | GPEtn 34:1 | 49 | 813.686 | SM 24:1 |
| 25 | 748.586 | GPEtn 36:0 | 50 | 815.702 | SM 24:1/SM 24:0 |

Phospholipids were identified by accurate mass to obtain possible phospholipid species and head group analysis from MS/MS scans. If the MS/MS spectra were clean to allow sn1 and sn2 fatty acid side chain analysis unique names were assigned (not shown here). A quality index was assigned to each lipid to specify high quality species. Data were validated with LipidMaps online and SDF tools, the Scionics LipidInspector and the LipidMaps MS Prediction tools. A data table of 45,000 phospholipids was created for internal data matching of lipids.

## Analyst performs multivariate statistics




Multivariate statistics including Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) and univariate statistics are performed with the integrated Analyst platform. The left picture shows the PCA scores and loadings plot of rat plasma (red) and lung samples (blue). The single components in the upper right corner are outliers which need special attention during data evaluation. The right picture (HCA) shows the different phospholipid patterns from lung (blue) and plasma (red).

## Conclusion

- Lipid fingerprinting can be used for fast hypothesis generation
- Chip-based nanoESI infusion can be used for cross-contamination free screening of plasma or tissue full extracts for phospholipids
- Chip-based nanoESI infusion allows fast acquisition (one minute per sample) in high/low resolution mode
- Iontrap technology together with MS/MS scans used for phospholipid identification via head group and neutral loss analysis of fatty acids identified around 50 lipid components
- Expressionist Refiner MS workflow performs full data analysis in 20 seconds
- Expressionist Analyst platform performs multivariate and univariate statistics
- Perturbations in the phospholipid contents for rats exposed to environmental tobacco smoke (ETS) were observed