Data generation and analysis with SIRIUS 4 on two biological case studies

Pieter Dorrestein (pdorrestein@ucsd.edu)
Skaggs School Of Pharmacy, UCSD
Alexey Melnik
Skaggs School Of Pharmacy, UCSD
Alexander Aksenov
Skaggs School Of Pharmacy, UCSD

Method Article

Keywords: mass spectrometry, sirius, in-silico ms-ms

Posted Date: September 20th, 2019

DOI: https://doi.org/10.1038/protex.2018.133

Abstract

The procedures for the metabolomics analysis performed in the present work has been described previously for skin\[1] and fecal material\[2]. The detailed step by step instructions to reproduce the biological case studies presented in this work are given below.

Introduction

This protocol summarizes the workflow on how to get from sample to the results described in the paper Sirius 4

Reagents

 Ethanol of LC-MS \(Optima) grade \(ThermoFisher Scienti c) for swabbing and sample extraction CAUTION: ethanol is highly ammable and irritating to eyes and skin. Water of LC-MS \(Optima) grade \(ThermoFisher Scienti c). ESI-L Low Concentration Tuning Mix \(Agilent Technologies, CA, USA) for external calibration of the MaXis II QTOF mass spectrometer Hexakis\(1H,1H,3H-tetra uoropropoxy)phosphazene \(Synquest Laboratories, FL, USA), m/z 922.009798, for internal calibration \(lockmass) of the MaXis II QTOF mass spectrometer. CAUTION: This compound is irritating to eyes and skin. Acetonitrile LC-MS \(Optima) grade \(ThermoFisher Scienti c). CAUTION: Acetonitrile is highly ammable. Formic acid of LC-MS grade, Optima grade \(ThermoFisher Scienti c). CAUTION: Formic acid is highly corrosive.

Equipment

 \(Skin sampling): Cotton swabs, model Puritan® \(Spectrum, CA, USA), reference 150-533353 or 150-2971696. \(Faeces sampling): BBL culture swabs \(Becton, Dickinson and Company, Sparks, MD). Deep-wells microtiter plate, 2 mL, polypropylene, model NUNC \(ThermoFisher Scienti c, MA, USA) reference 278753 or an equivalent generic 96 wells microtiter plate, 400 µL, polystyrene. Reusable lid for a 96 wells plate \(ThermoFisher Scienti c), reference 351190 or equivalent. Stainless steel tweezers. Single or multi channel 20-200 µL micropipette. Plate mat for storage, model Corning Storage Mat III \(ThermoFisher Scienti c, MA, USA). Benchtop vacuum concentrator compatible with 96-well plates evaporation \(Centrivap Labconco, MO, USA). Reverse phase C18 LC column 1.7 µm particle size, 50 2.1 mm \(Phenomenex, CA, USA) or equivalent. Ultra high-performance liquid chromatography \(UHPLC) system coupled to a tandem mass spectrometer with an ESI source; for the biological studies described here we have used UltiMate 3000 UHPLC \(Dionex, Idstein, Germany) coupled to MaXis II QTOF \(Bruker Daltonics, Bremen, Germany).

Procedure

Sampling of skin 1. Prepare the swabs by disposing of the ethanol:water 1:1 \(v/v) from the swab container and by replacing it with fresh ethanol:water 1:1 \(v/v). 2. Prepare a 96-well plate \(volume 2 mL) by adding 500 µL of the solvent mix into each well, and placing it on an ice bed. 3. For each sampling spot: 4. Take one swab, and swab vigorously the spot of interest with the cotton bud side of the swab for at least five seconds. CRITICAL STEP: The swabbing procedure must be as consistent as possible across spots. 5. Place the swab into the corresponding well of the 96-well plate and cut it with a pair of scissors in order to leave only the cotton bud in the well. Dispose the swab wood stick. CRITICAL STEP: During the placing of the swab and the cutting of the cotton bud, use aluminium fold ﬁlm to cover the rest of the plate to prevent cross-contamination of other wells. 6. Close the 96 well plate with a polypropylene reusable cover. CAUTION: All samples must be prepared using the same extraction protocol, using the same reagents, and ideally by the same operator. **Sampling of stool** The sampling is carried out by individuals as described in \[2]. Samples are collected using swabs and returned by mail. Samples collected outside of the United States are shipped using domestic post within each country to an aggregation site and stored at -80° at the aggregation site until shipment to the United States. Shipment to the United States is done on dry ice using a certiﬁed shipping service. **Metabolite extraction of skin samples** 1. Leave the 96-well plate with soaked cotton buds at a temperature of 4° for 24 to 48 hours. CRITICAL STEP: Note that the extraction yield can be increased with higher extraction temperature and longer extraction time. However, this can lead to the degradation of some molecules. 2. Remove each cotton bud with a pair of tweezers, and retrieve as much solvent as possible by pressing the bud on the side of the well. Dispose of the cotton bud. CRITICAL STEP: During this step prevent contamination of other wells by covering them with aluminium ﬁlm. 3. Evaporate the 96-well plate solvent content using the benchtop vacuum centrifuge concentrator. CAUTION: The use of centrifugation device requires speciﬁc training, in particular to ensure that the rotor is properly equilibrated before spinning. 4. Redissolve the plate content in 120 µL of 1:1 \(v/v) acetonitrile:water at ambient temperature. 5. Cover the plate with a lid. Agitate gently the plate and use ultrasonic bath for 5 min to ensure the appropriate dissolution of the sample. 6. Centrifugate the 96-well plate at 2000 rpm for 15 minutes at ambient temperature. 7. Transfer 100 µL of the plate content with multi-channel pipette into a new 96-well plate \(well volume 400 µL). CRITICAL STEP: Make sure to reproduce the order of the wells and avoid cross contamination of the plate by covering it with aluminium foil, and avoid undissolved material upon sample transfer. 8. \(Optional, but recommended). Pool 10 µL from each of twelve representative samples to create a pooled QC sample. 9. Store the plate with extracts at -20° until further analysis. Note that for longer storage, storing the samples at -80° is recommended, followed by sonication of extracts to redissolve any precipitated material. **Metabolite extraction of stool samples** 1. Remove the swab tubes scheduled for analysis from the -80° freezer and place on dry ice for the duration of sample processing. 2. Place each swab onto a Thermo Fisher Scientific \(Waltham, MA) 2-ml deep-well 96-well plate set on top of dry ice coolant. CRITICAL STEP: Make sure to reproduce the order of the wells and avoid cross contamination of the plate by covering it with aluminium foil, and avoid undissolved material upon sample transfer. 4. Snap the top part of each swab stick o° and discard. 5. Add 200 µl of HPLC-grade 90% \(vol/vol) ethanol-water solvent to each well using a multichannel pipette immediately after ﬁlling all wells. It is recommended to include four blanks of unused swabs and extraction solvent onto each plate. 6. Seal with a 96-well plate lid,

sonicate for 10 min 7. Place into the refrigerator at 2□ to extract samples overnight. 8. Remove and discard swabs. 9. Place the plates into a lyophilizer, and dry down the entire sample. 10. Add 200 µl 90% \(vol/vol) ethanol-water for resuspension. 11. Resealed the plates and centrifuge at 2,000 rpm for 10 min. CAUTION: The use of centrifugation device requires speci□c training, in particular to ensure that the rotor is properly equilibrated before spinning. 12. Transfer 100-µl aliquots of sample onto a Falcon 96-well MS plate using a multichannel pipette, seal plate immediately with sealing □lm. 13. Centrifuge sealed plates at 2,000 rpm for 10 min and analyze immediately or store at 2□ until analysis. CAUTION: The use of centrifugation device requires speci□c training, in particular to ensure that the rotor is properly equilibrated before spinning. **Mass spectrometry analysis** 1. Prepare the LC mobile phase A of 100:0.1 \(v/v) water:formic acid. 2. Prepare the LC mobile phase B of 100:0.1 \(v/v) acetonitrile:formic acid. 3. Prime the C18 HPLC column according to the manufacturer's guidelines. 4. Set up the LC linear gradient method as follows: 0-0.5 min 2% B, 0.5-2 min 2-20% B, 2-8 min 20-98% B, 8-9 min 98-98% B, and 9-10 min 2% B with a constant of flow rate of 500 µL/min. 5. Set up the LC parameters as following: column temperature \(40□), loop factor wash \(3) \(Dionex UltiMate 3000, or as appropriate for other systems), injection volume \(10 µL for skin samples, 5 µL for fecal samples). Prime the injection system multiple time. 6. \(Optional) Clean-up the ESI source of the MS instrument according to the manufacturer's guidelines. Calibrate instruments using the calibration solution according to manufacturer's guideline. For the MaXis II QTOF, add the solution containing internal lockmass \(m/z 922.009) in the ESI source in the positive ion mode, following the the manufacturer's guideline. 7. Set up the MS method as follows: for the MaXis II QTOF MS, acquire spectra in positive ion mode in the mass range of m/z 80□2,000 with the following settings: capillary voltage, 4,500 V; ion source temperature, 180 □; dry gas □ow, 9 L/min; spectra rate acquisition, 3 spectra/s. Perform MS/MS fragmentation of the seven most intense selected ions per spectrum \(TOP7) using ramped collision-induced dissociation \(CID) energy, ranging from 16 to 48 eV, to get diverse fragmentation patterns. Set MS/MS exclusion after 3 spectra to be released after 30 s. Set an MS/MS exclusion list for the mass range of m/z 921.5□924.5 to exclude the lockmass, and if needed include other adducts of the lockmass if their intensities are in the range of MS/MS threshold. 8. Run a blank sample and verify that no signi□cant contamination from source is impacting the system by assessing abundances of observed ions and ensuring that they do not exceed thresholds appropriate for the instrument \(e.g. below few thousand counts for the MaXis II QTOF instrument). If high abundance contaminants are present, consider additional source cleaning, replacement of mobile phases, glassware, LC column or tubing, as needed. 9. \(Optional, but recommended): Start by running the QC sample, followed by a blank, and repeat it at least two times. Check for satisfactory reproducibility by comparing multiple QC samples \(coefficient of variation, CV, below 0.1-0.3). Check for carryover by inspecting the blank that followed a QC sample. If needed, adjust the injection volume to prevent chromatographic saturation \(peak broadening), and mass spectrometer detector saturation \(peak flattening). If needed, optimize the LC gradient to ensure appropriate separation of representative features, and the number of needle/injector washes to reduce carry-over below 1%. 10. Prepare the LC-MS sequence for the samples injections. Include the analysis of the blank and the QC sample every 12 samples or less, depending on total number of samples. If possible, the samples order should be randomized. 11. \(Optional): Include a clean-up sample every 12 samples to limit column saturation. This is recommended if sample contains

components that are strongly retained by the column and are not eluted completely by the mobile phase. Typically, this clean-up run is composed of multiple mobile phase composition variation according to the manufacturer's guidelines. 12. Run the LC-MS sequence, and monitor regularly that \(i) elution pro□le of every QC sample is reproducible \(5-10 sec window maximum for UHPLC, or a coefficient of variation of 0.1-0.3), and \(ii) that the calibration is stable \(< 10 ppm for MaXis II QTOF). If needed, stop the LC-MS sequence, and calibrate the instrument according to manufacturer guidelines. For the MaXis II QTOF, lock mass calibrant must be added every 12 hours. If the LC-MS sequence was stopped, run a QC sample and a blank prior continuing the rest of your analysis. 13. After □nishing the LC-MS experiments, open representative □les, from both the beginning and the end of the sequence, with the MS data viewing software \(Data Analysis for MaXis II QTOF). Overlap features across samples, such as the internal standard, and assess the stability of the chromatography \(retention shift in seconds), and the MS calibration \(m/z window in ppm). 14. Export all LC-MS/MS □les into the centroided .mzXML format. For the MaXis II QTOF MS, use CompassXport to apply calibration, and convert the data from .d format to .mzXML format in centroid mode \(select Filters: Peak Picking, MS-Levels 1-2). **Data analysis** Creating feature tables. Multiple software packages for MS feature extraction exist and can be used; the present protocol is given for the use of the open-source OpenMS 2.0\[3] software utilized for feature detection in the fecal sampling biological study example and MZmine2\[4] used in the skin sampling biological study. The recommended settings were found to be appropriate for the data obtained in our experiments; however, experienced users may □nd changes in parameters necessary, especially if di□erent instrumentation is used. **Creating feature tables with OpenMS**. 1. Prior to feature extraction, the collected HPLC-MS raw data □les are converted from Bruker's .d to .mzXML format 2. Peaks across di□erent chromatograms deconvolved and aligned \(feature detection). The alignment window is set at 0.5 min, the noise threshold is set at 1,000 counts, the chromatographic peak full width at half-maximum \(FWHM) value is set at 20, and the mass error is set at 30 ppm. 3. An output table with detected MS features across all samples and their corresponding abundances is obtained. 4. \(Recommended): All of the peaks that are present in any of the blanks with a signal-to-noise ratio \(S/N) below 10:1 are removed from the □final feature table. The threshold can be adjusted based on speci□cs of each experiment. **Creating feature tables with MZmine2**. 1. Prior to feature extraction, the collected HPLC-MS raw data □files are converted from Bruker's .d to .mzXML format. 2. The experimental fi□les are loaded and batch-processed with the following settings for each step in the batch: _Mass detection_ * Noise level 1000 * Chromatogram builder * Minimum time span 0.01 min * Minimum peak height 3000 * m/z tolerance 0.1 m/z or 20 ppm _Chromatogram deconvolution - Baseline cutoff * Minimum peak height 3000 * Peak duration range \(min) 0.01 - 3.00 * Baseline level 300 _Deisotopisation - Isotopic peak grouper_ * m/z tolerance 0.1 m/z or 20 ppm * RT tolerance 0.1 min * Maximum charge 4 _Peak alignment - Join aligner_ * m/z tolerance 0.1 m/z or 20 ppm * Weight for m/z 75 * Weight for RT 25 * RT tolerance \(min) 0.1 _Peak □ltering - Peak list raw □lter_ * Minimum peak in a row 3 * Minimum peak in an isotope pattern 2 3. An output table with detected MS features across all samples and their corresponding abundances is obtained. 4. \(Recommended): All of the peaks that are present in any of the blanks with a signal-to-noise ratio \(S/N) below 10:1 are removed from the □nal feature table. The threshold can be adjusted based on speci□cs of each experiment. **Molecular networking** 1. Raw mass spectrometry data □files in the

.mzXML format are uploaded MassIVE mass spectrometry database \(https://massive.ucsd.edu/). 2. Molecular networking\[5] is performed to identify spectra shared between di□erent sample types and to identify known molecules in the data set. 3. MS/MS spectra are window □ltered by choosing only the top 6 peaks in the □50-Da window throughout the spectrum. 4. The MS spectra were are clustered with a parent mass tolerance of 0.02 Da and an MS/MS fragment ion tolerance of 0.02; consensus spectra that contained fewer than 4 spectra are discarded. 5. A network is created where with edges □ltered to have a cosine score above 0.65 and more than 5 matched peaks. The edges between two nodes are kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. 6. Set required library matches to have a score above 0.7 and at least 6 matched peaks, and search the spectra in the network against GNPS spectral libraries\[6]. All resulting annotations are at level 2/3 according to the proposed minimum standards in metabolomics\[7]. 7. Molecular networks are visualized and using Cytoscape software\[8]: download the required fi□les from the generated GNPS job by clicking on the □Download GraphML for Cytoscape□ link on the status page. 8. Open the Cytoscape software and import the downloaded □file: File \! Import \! Network \! File. . . and then select the .graphml fi□le in the downloaded folder. 9. Select appropriate labels for the network nodes and edges under the □Style□ tab. 10. The structure of molecules interest that are found within the cluster with existing annotations may be postulated using annotation propagation from adjacent annotated nodes in the cluster as described in reference \[6] by assessing di□erences in parent mass and fragmentation patterns. 11. \(Optional, if possible): For level 1 identi□cations, authentic standards for postulated compounds need to be procured and analyzed under identical experimental conditions. The retention times \(RTs) and MS/MS spectra need to be compared to those experimentally obtained for the putatively annotated compounds. **Determining the molecular structure with SIRIUS 4** 1. Extract the MS/MS spectrum of the compound of interest; this can be achieved in several ways: _Extract the MS/MS using MSConvert software \(ProteoWizard package)9_: 2. Launch MSConvert GUI 3. Click "Browse" for the "File" □field to select the experimental .mzXML □file that contains MS/MS spectrum of interest. 4. Click "Add" to include the fi□le into the analysis window. 5. Click "Browse" for the "Output Directory" □field to select the output directory. 6. Select "Output format": mgf 7. Select "Binary encoding precision": 32 bit. 8. Uncheck "Use zlib compression" box 9. Under "Filters", select "Subset" and enter the scan number of the MS/MS spectrum of interest. 10. Click "Add". 11. Click "Start". 12. The .mgf □file containing the single MS/MS will be saved in the selected output folder. _Extract the MS/MS peak list directly from GNPS job_: 13. When viewing the in-browser network cluster, select the node of interest by left clicking on it. The top right window will display the MS/MS of the consensus spectrum for this node. 14. Click on the "Download MS/MS Peaks" icon in the top right corner of the MS/MS spectrum window. The browser will download the .txt fi□le. CAUTION: The downloaded MS/MS peaks will belong to the consensus spectrum obtained from clustering multiple experimental spectra and thus may di□er from MS/MS patterns in individual □files. Caution must be used to ensure that the appropriate spectrum is used for further analysis. 15. Launch the GUI of the SIRIUS software. 16. Import the MS/MS spectrum \(spectra) for the analysis: * Click "Import" icon * Click "+" icon * Select the □file with MS/MS of interest \(enter the parent m/z if not recognized from the fi□le) * Alternatively: Click "Batch Import" icon * Select the □file with MS/MS of interest \(Note: for batch import, multiple MS/MS spectra can be loaded at once from a single □file. The .txt □files cannot be used for batch

import). 17. For single spectrum import, select the ionization mode \(\[M+H]+ for the present case studies). 18. Click "Compute All" 19. In the pop-up window, select: * "Elements beside CHNOP": include sulfur for the present biological case studies; for the range of expected number of atoms use up to "+2" instead of "auto" * Instrument: "Q-TOF" for the present biological case studies * "ppm": "20" for the present biological case studies * "candidates": default of 10 * In the drop down menu, select "formulas from biological databases" \(in the present case studies) or other option if appropriate \(Note: selecting "all possible molecular formulas" would result in longer computing time) * Click the "CSI:FingerID" button * Under "Search in" select "biological database" * Select possible adducts \(\[M+H]+ is selected by default) 20. Click "Compute". 21. Once the "Jobs" icon is static, it indicates that the computation is completed and the results can be explored. Under the "Compounds" tab \(left hand side), the currently active job is selected. 22. In the "Sirius Overview" tab \(main window), browse the results by clicking the suggested formulas listed on top of the window. The explained fragments and calculated fragmentation tree are displayed below for each formula or could be viewed individually in the corresponding tabs. 23. Select "CSI:FingerID Details" tab for possible structures for each formula. Speci c libraries for possible candidate molecules could be selected. If no entries are displayed, the molecule with the experimental spectrum is not present in selected source databases. 24. In both of the biological case studies, the correct structures were found to have the highest scores for both the molecular formulas and predicted structures; however, oftentimes that may not be the case and researcher's discretion and expertise are essential for selecting the potential candidate molecules for further consideration and interpretation.

Timing

The sampling time for skin swabbing and faeces sampling is generally under one minute per sample. The total time necessary for the samples preparation and data analysis depends on the number of samples. Typical timing is approximately one hour for placing swabs into 96-well plate, 6 hours to 12 hours for metabolite extraction, 24 hours for LC-MS analysis, 2 hours for MS feature extraction \(strongly depends on the computational power of the system that runs the software), 6 hours for preparing and running GNPS job. The time for network-based annotation propagation can vary widely depending on the analysis circumstances. The time for a single SIRIUS/CSI:FingerID job for a single spectrum processing is typically few seconds on a personal computer. A detailed breakdown of timing for speci c experimental steps is given in publication of Sirius 4.

Troubleshooting

The users should follow general good laboratory practices for the analysis to avoid issues such as contamination. For detailed troubleshooting guidance speci c to the LC-MS/MS analysis, the user should refer to the manuals of the instrumentation that is used in the experiment. Additional troubleshooting guidelines speci c to the described work ow are given in \[1]. The troubleshooting tips for the GNPS-based molecular networking and annotation could be accessed at: https:// ccms-ucsd.github.io/GNPSDocumentation/troubleshooting/

Anticipated Results

The anticipated results should be comparable to those described in the given examples for the two biological case studies.

References

1. Protsyuk, I. et al. 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. Nature protocols 13, 134–154 \(2018). 2. McDonald, D. et al. American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems 3 \(2018). 3. Röst, H. L. et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods 13, 741–748 \(2016). 4. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf 11, 395 \(2010). 5. Watrous, J. et al. Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A 109, E1743–E1752 \(2012). 6. Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34, 828–837 \(2016). 7. Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221 \(2007). 8. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498–2504 \(2003). 9. Adusumilli, R. & Mallick, P. Data Conversion with ProteoWizard msConvert. Methods in molecular biology \(Clifton, N.J.) 1550, 339–368 \(2017).