A metabolomic protocol for investigating the gut microbiome

Shuo Han (shuo.han@duke.edu)
Duke University School of Medicine https://orcid.org/0000-0003-1181-351X
Emma R. Guiberson
Stanford University School of Medicine
Justin L. Sonnenburg (jsonnenburg@stanford.edu)
Stanford University School of Medicine

Method Article

Keywords: Metabolomics, Microbiome, Mass Spectrometry, Chemical Biology

Posted Date: October 11th, 2022

DOI: https://doi.org/10.21203/rs.3.pex-2055/v1

Abstract

A significant hurdle that has limited progress in microbiome science has been identifying and studying the diversity of metabolites produced by the gut microbes. Gut microbial metabolism produces thousands of difficult-to-identify metabolites, which present a challenge to study their roles in host biology. Over the recent years, mass spectrometry-based metabolomics has become one of the core technologies for identifying small metabolites. However, metabolomics expertise, ranging from sample preparation, instrument use, to data analysis, is often lacking in academic labs. Most targeted metabolomics methods provide high levels of sensitivity and quantification, while they are limited to a panel of predefined molecules that may not be informative to microbiome-focused studies. Here we have developed a gut microbe-focused, wide-spectrum, targeted metabolomic protocol using Liquid Chromatography-Mass Spectrometry (LC-MS) and bioinformatic analysis. This protocol enables users to carry out experiments from sample collection to data analysis, only requiring access to a LC-MS instrument, which is often available at local core facilities. By applying this protocol to samples containing human gut microbial metabolites, spanning from culture supernatant to human biospecimens, our approach enables high confidence identification of >800 metabolites. We expect this protocol will lower the barrier in tracking gut bacterial metabolism in vitro and in mammalian hosts, propelling hypothesis-driven mechanistic studies and accelerating our understanding of the gut microbiome at the chemical level.

Introduction

The human gut microbiota, consisting of trillions of microorganisms, encodes diverse metabolic pathways that impact host gastrointestinal health, systemic immunity, and numerous aspects of physiology. One key way in which gut bacteria interact with the host is through the production of small molecules from anaerobic metabolism[1–3]. Several works have characterized the roles of highly abundant microbiota-dependent metabolites (or MDMs, e.g., short chain fatty acids) that modulate aspects of host biology (e.g., via GPCR-binding[4], chromatin modification[5]). In contrast, the remaining thousands of uncharacterized MDMs from diverse precursors (e.g., dietary, pharmacological, host-derived) constitute the “dark matter” of the gut microbial metabolism, many of which are molecules whose chemical identities remain unknown[6]. The subset of this dark matter for which chemical identities are known are understudied MDMs: 1) not yet mapped to gut bacterial species or biochemical pathways, and 2) not represented in mass spectrometry-based databases. The inability to detect and track these molecules experimentally present a challenge to investigate their mechanistic roles in host phenotypes. Therefore, our limited knowledge of the diverse universe of MDMs and their bioactivities underscores the importance of mapping and characterizing them using *in vitro* culture, animal models, and human samples.

Over the recent years, mass spectrometry-based metabolomics has become one of the core technologies for characterizing MDMs. There are both targeted and untargeted approaches, where each brings unique insight into the diverse scope and specificity of compound identification. For example, targeted metabolomic methods in general provide high levels of sensitivity and quantification through

specific analyte monitoring and fragmentation that allows for higher order identifications, but they are limited to a predefined panel of molecules due to a need for a target *m/z* to monitor. In contrast, untargeted methods provide an expansive list of features, holding great discovery potential for novel compounds. However, the success of these data sets relies heavily on the ability to identify these peaks, which can be difficult without extensive samples or libraries to compare against. Additionally, identifications are often at the expense of quantification and accuracy. Here we present a gut microbe-focused, wide-spectrum, targeted metabolomic protocol to enable users to carry out their experiments from sample collection to data analysis. This tool provides an important resource to the field and accelerates our functional understanding of the gut microbiome at the chemical level.

Development of the protocol

We have created a gut microbe-focused, wide-spectrum, and targeted metabolomic protocol that can be set up in an academic lab with access to LC-MS instrumentation. First, we constructed a mass spectrometry-based reference library based on our curation of 800+ metabolite standards relevant to gut microbial and host metabolism. Second, we developed metabolite extraction protocols tailored to several sample types that are commonly used in microbiome studies (e.g., host feces, serum) (Fig. 1). Lastly, we built a bioinformatic workflow to integrate data from different experiments, sample types, and analytical methods (Fig. 2). Together, this protocol provides a detailed and open-source framework from constructing a custom mass spectrometry library to downstream data analyses, resulting in identification of metabolites that are present in, or differentiate, samples.

Applications of the method

This protocol enables detection and identification of a wide spectrum of metabolites in diverse bacterial and host sample types. Metabolomic data from biological samples often implicate specific molecules that may be involved in a biological process or disease for further study (e.g., inflammatory bowel disease). Using metabolomic profiles of bacteria grown in culture (we have previously reported the profiles of 170+ gut bacteria in culture), species capable of producing molecule(s) of interest can be identified. These bacterial species can then be studied 1) in bacterial culture (e.g., to study aspects of metabolite production and consumption) or 2) in gnotobiotic mice (e.g., to examine the biology of the metabolite in a host organism). For example, the production of a microbial metabolite can be enhanced in the gut and/or circulation by colonizing mice with a bacterial strain that produces high levels of a given metabolite; its production can be ablated by colonizing another group of mice with a mutant strain of bacteria deficient in metabolite production (when bacterial genetic tools are available). In this manner, mouse models can be leveraged to study host metabolic phenotypes (e.g., inflammatory bowel diseases) linked to elevated levels of a target metabolite[7,8]. In a second example, coupling strain-specific bacterial metabolomic and genomic data can reveal the genetic basis of the metabolic phenotypes among bacteria. Previously, we mapped the production patterns of several bioactive polyamines among 170+ sequenced human gut bacteria in individual *in vitro* cultures. Metabolite phenotypes were used as the

basis for comparative genomic analysis to define candidate sets of genes/enzymes responsible for metabolite production[7].

In addition, this protocol can be adapted and further enhanced by 1) the addition to the library of human dietary molecules (e.g., dietary plant natural products, fermented food metabolites), and 2) the application to characterizing MDMs from novel bacterial strains isolated from diverse human populations or patient cohorts. These broad applications will accelerate discoveries of potential mediators of microbe-host interactions.

Alternative methods

In this protocol, we have developed a wide-spectrum, targeted metabolomic workflow to inform microbiome research. Previously, many targeted analyses have been developed for metabolomics studies leveraging different instrumentations and approaches (e.g., NMR[9], GC-MS[10], and LC-MS[11,12]). For example, most fecal LC-MS metabolomics studies have been performed through targeted analyses of a small set of known MDMs, such as short-chain fatty acids[13,14], tryptophan metabolites[15,16], and secondary bile acids[17–19]. While targeted analyses enable higher level of sensitivity and quantification, many molecules of potential interest go unidentified. We overcome this limitation by constructing a microbiome-focused mass spectrometry reference library, based on our collection of 800+ metabolites spanning diverse chemical classes. Leveraging this reference library, we have 1) mapped MDMs to prevalent human bacterial species in culture, and 2) examined bacterial and host metabolism in gnotobiotic mouse models. The protocol enables users to identify MDMs in biological samples and address the critical question of which metabolites are associated with which species in the human gut microbiome, enabling future mechanistic studies of microbe-microbe and microbe-host interactions.

Limitations

While a shared instrument eliminates the high expense of purchasing one's own, how well the instrument is maintained over time determines the instrument's reliability, including fluctuations in sensitivity and frequency of repairs. We use an internal standard-based approach to correct for changes in both sensitivity and retention time within an experiment and between different experiments. While our internal standard collection represents diverse compound classes, they cannot fully recapitulate changes in all compound classes. Therefore, investigators should select a set of internal standards that are representative in chemical structures to those present in their reference library.

Because we use a reference library for identification, we are limited to the 800+ metabolites detected in our library. In a typical experiment, we observed an average of 1000-2000 unidentified molecular features that may be exciting candidates for further investigation. The users should decide between 1) a targeted approach using a standard collection to achieve high-confidence identification while limiting the overall scope, 2) an untargeted approach relying on public spectral databases for identifications, providing a broader scope with less identification confidence, or 3) a semi-targeted approach that is a hybrid of both 1) and 2). A semi-targeted approach further investigates the detected

but unknown molecular structures by untargeted methods, increasing the total number of identified compounds. Additional factors, such as how MS/MS data collection for untargeted analysis may impact scanning time and detection sensitivity[20], may be important considerations that require consultation with someone with mass-spectrometry expertise. Here we present this protocol as a beginning step for the community to update and expand.

Reagents

Chemicals (store at room temperature unless otherwise noted)

- Water, LC/MS grade (Fisher, cat. no. W6-4)
- Methanol, LC/MS grade (Fisher, cat. no. A456-4)
- Formic acid, LC/MS grade (Fisher, cat. no. A117-50), store at 4 ^{o}C
- Ammonium bicarbonate (Sigma, cat. no. 09830-500G)
- Ammonium formate (Sigma, cat. no. 70221-25G-F)
- 4-Chloro-phenylalaninne (Carbosynth, cat. no. FC13398-2G)
- Tridecanoic acid (Sigma, cat. no. T0502-1G), store at -20 ^{o}C
- 2-Fluorophenylglycine (Santa Cruz Biotechnology, cat. no. sc-259898)
- Phenylalanine-2,3,4,5,6-d5 (Cambridge Isotope Laboratories, or CIL, cat. no. DLM-1258-1)
- Glucose-1,2,3,4,5,6,6-d7 (CIL, cat. no. DLM-2062-0.5)
- Methionine-methyl-d3 (CIL, cat. no. DLM-431-1)
- 4-Hydroxyphenyl-d4-alanine (CDN Isotopes, cat. no. D-2599-0.25G)
- Tryptophan-2,4,5,6,7-d5 (CDN Isotopes, cat. no. D-1522-0.1G)
- Leucine-5,5,5-d3 (CDN Isotopes, cat. no. D-1973-0.5G)
- N-Benzoyl-d5-glycine (CDN Isotopes, cat. no. D-5588)
- 4-Bromophenylalanine (Sigma, cat. no. 18055-500MG)
- Progesterone-d9 (CIL, cat. no. DLM79530.01)
- Di-N-octyl phthalate-3,4,5,6-d4 (CDN Isotopes, cat. no. D-2291-0.1G)

- d19-Decanoic acid (CDN Isotopes, cat. no. D-1616-0.5G)
- d15-Octanoic acid (CDN Isotopes, cat. no. D-1948-0.5G)
- Indole-2,4,5,6,7-d5-3-acetic acid (CDN Isotopes, cat. no. D-2203-0.05G)
- Carnitine-trimethyl-d9 (CDN Isotopes, cat. no. D-5780-0.05G)
- d27-Tetradecanoic acid (CDN Isotopes, cat. no. D-1711-0.5G)

Commercial mass spectrometry standard libraries

- IROA Mass Spectrometry Metabolite Library of Standards
- Metasci Organic Acids Library

Plasticware and consumables

- 2 mL 96-Well polypropylene microplates (Fisher, cat. no. 12566121)
- Silicone closing mats for 96-well microplate (Agilent, cat. no. 5042-1389)
- 96-Well filter plates, 0.2 µm PVDF filter (Corning, cat. no. CLS3508-50EA)
- 1 mL 96-Well polypropylene microplates (Agilent, cat. no. 5043-9305)
- Glass beads, acid-washed, 150-212 µm (Sigma, cat. no. G1145-10)
- 2 mL Microcentrifuge tubes with screw caps (Fisher, cat. no. 02-682-558
- 1 L glass bottles (Pyrex, cat. no. 13-700-447)
- Acquity UPLC BEH C18 Column, 130 Å, 1.7 µm in particle size, 2.1 mm x 100 mm (Waters, cat. no. 186002352)
- Acquity UPLC BEH C18 VanGuard Pre-column, 130 Å, 1.7 µm in particle size, 2.1 mm x 5 mm (Waters, cat. no. 186003975)
- Acquity UPLC BEH Amide, 130 Å, 1.7 µm particle size, 2.1 mm x 150 mm (Waters, cat. no. 186004802)
- Acquity UPLC BEH Amide VanGuard Pre-column, 130 Å, 1.7 µm particle size, 2.1 mm x 5 mm (Waters, cat. no. 186004799)

Buffer and solvent preparation

Extraction buffer (Supplementary Table 1)

- Use LC/MS grade methanol.
- Make stock solution for each internal standard at 1000x concentration in 100% methanol (e.g., 6.8 mM for 4-Chloro-phenylalaninne in a 40 mL stock solution). Store the stock solutions in glass vials at 4^{o}C.
- Prepare buffer by adding 1 mL of each internal standard stock solution (1000x) into 1 L of 100% methanol in a glass bottle. Store the extraction buffer at 4^{o}C for repeated use in the next six months.

Reconstitution buffer (Supplementary Table 1)

- Use LC/MS grade methanol and water.
- Make stock solution for each internal standard at 1000x concentration in 50% methanol (e.g., 12.5 mM for Phenylalanine-2,3,4,5,6-d5 in a 40 mL stock solution). Store the individual stock solutions in glass vials at 4^{o}C.
- Prepare buffer by adding 1 mL of each internal standard stock solution (1000x) into a 1 L of 50% methanol in a glass bottle. Store the reconstitution buffer at 4^{o}C for repeated use in the next six months.

Mobile phase solvents for the three analytical methods (Supplementary Table 1)

- Use LC/MS-grade formic acid, water, methanol, and acetonitrile.
- Make each mobile phase solvent fresh in 1 L volume in a 1 L glass bottle.
- For C18 negative mode solvent B, first dissolve 514 mg of ammonium bicarbonate in 50 mL of water, then add solution to 950 mL of methanol.
- For HILIC positive mode solvent B, first dissolve 631 mg of ammonium formate in 50 mL of water, then add solution to 950 mL of acetonitrile. Sonicate the final 1 L solvent in a heated sonicating water bath at 50^{o}C for 20 minutes to ensure full dissolution.

Equipment

Equipment

- Agilent 6545 LC/Q-TOF with dual jet stream ESI source (Agilent, cat. no. G6545BA)
- Agilent 1290 Infinity II LC (Agilent, cat. no. G7120A)
- TurboVap 96 (Biotage, cat. no. C103263)
- Mini-BeadBeater 24 Cell Disrupter (BioSpec Products, cat. no. 112011)
- Sorvall Legend XT large benchtop centrifuge (Thermo Scientific, cat. no. 75063839)
- Small benchtop centrifuge (Eppendorf, cat. no. 5424)

Software

- Agilent MassHunter Qualitative Data Analysis software is available for purchase from https://www.agilent.com/en/product/software-informatics/mass-spectrometry-software/data-analysis/qualitative-analysis
- Abf Converter software is available for free from https://www.reifycs.com/AbfConverter/
- MS-DIAL software is available for free from http://prime.psc.riken.jp/compms/msdial/main.html

The version 3.82 that we use in this protocol can be downloaded from the archive folder http://prime.psc.riken.jp/compms/msdial/download/repository/

- An example bioinformatic analysis workflow and its associated scripts are available from this protocol (https://github.com/the-han-lab/Han_et_al_Metabolomics_Protocol_2022). For installation and configuration, follow the instructions in the repository.

Equipment setup

Set up analytical methods on the Agilent 6545 LC/Q-TOF instrument (Supplementary Table 2)

LC-MS gradient for both the C18 positive and C18 negative methods (Data are collected from 0.5 min to 10 min).

LC-MS gradient for the HILIC positive method (Data are collected from 2 min to 10 min).

Mass spectrometer key parameters for three methods (see more method-specific details in Fig. 3-5).

Procedure

Build the chemical standard collection

Timing varies depending on the sources of the standard collection

1. Curate a library of metabolite standards that are commercially available as a collection (e.g., IROA Mass Spectrometry Metabolite Library of Standards, Metasci Organic Acids Library) and/or are individually purchased in manufacturer's stock bottles (e.g., Sigma, Acros).

2. For commercial metabolite standard collections (e.g., IROA, Metasci), follow the manufacturer's instructions on drying, reconstituting, and pooling. Whenever possible, reconstitute metabolite standards in 50% methanol (LC-MS grade) to reach a stock concentration of 10 mM.

3. For each individually curated metabolite in manufacturer's stock tubes, weigh at least 3 mg, transfer to a 2 mL Eppendorf tube. The metabolite should be reconstituted in 50% methanol to reach a stock concentration of 10 mM.

4. Generate individual pools of 20-30 metabolite standards by combining stocks and diluting with 50% methanol. Each metabolite in the pool should reach a working concentration of 200 uM.

CRITICAL STEP The metabolite standards in each pool should not share the same molecular mass. Two metabolite standards that share the same molecular mass will confound the RT assignment to each metabolite standard in the pool (see Step 6 below).

5. Analyze individual pools by one or more of the three analytical methods (Steps 7-11 below).

(Optional) If users wish to construct a linear range for each metabolite to enable quantification, individual pools of metabolite standards can be serially diluted in 50% methanol or a biological matrix where the quantification will be done (e.g., serum). If the matrix already contains the metabolite, in some cases there are commercially available versions of the matrix readily depleted of small molecules. The dilution series can be analyzed using each of the three analytical methods.

Construct the *m/z*-RT reference library

Timing varies depending on the size of the standard collection

6. Construct the following library values in an excel file spreadsheet: Calculate the *m/z* of each metabolite standard by 1) retrieving its monoisotopic mass from the PubChem database and 2) adding or subtracting the mass of a proton (1.007276 Da) depending on the default adduct ion type. For example,

the adduct ion $[M+H]^+$ is the default adduct ion for electrospray ionization in the positive mode (ESI+) and $[M-H]^-$ is the default adduct ion for electrospray ionization in the negative mode (ESI-).

7. Run individual pools of metabolite standards for each analytical method and save all raw file output from the LC-MS instrument.

8. Open in the Agilent Qual (Agilent MassHunter Qualitative Data Analysis software, v.B.07.00) a raw file containing a known pool of metabolite standards (e.g., .d file produced by an Agilent Q-TOF). The steps below are carried out in Agilent Qual.

9. Right click on the “TIC scan” option in the left panel and select “extract chromatogram” in the dropdown menu to open the Extract Chromatogram window.

10. Select the settings as described (Fig. 6a-c). The *m/z* input here should be the comma-separated list of predicted *m/z* for all metabolite standards contained in this pool (see Step 6 above).

11. Inspect each EIC-scan chromatogram corresponding to each *m/z* value in the left panel. Record RT in the reference library for a compound only when there is a single, strong peak (e.g., greater than 10^4 in peak height) for the corresponding *m/z* value (Fig. 6d). Alternatively, there can be two or more strong peaks or no peak detected for a given *m/z* value (Fig. 6e-f). Save the library file as .txt after all compounds have either retention times associated with *m/z* values, or have been removed from the original list of compounds. This will be your library file that you specify when analyzing sample data below.

? TROUBLESHOOT (see Supplementary Table 3)

(Optional) For metabolites run in dilution series, RTs of the same metabolite at several concentrations can be used to produce an averaged RT in the reference library. This averaged RT value (1) increases the accuracy by averaging sample-to-sample variations; and (2) distinguishes the true signal from background noise by validating the peaks for which the ion counts proportionally increase with the concentration.

Prepare samples for LC-MS analysis

Timing 0.5 d

We have developed a metabolite extraction method with specific sections tailored to different sample types used in microbiome studies. Once metabolites are extracted from different sample types, they can either be analyzed immediately or stored at -80 °C for future analysis (Fig. 1).

Homogenize and precipitate samples (steps vary depending on each sample type)

Sample type 1: Bacterial supernatant

12. Remove frozen bacterial supernatant from -80 °C and thaw on ice prior to extraction.

13. Pipette 200 µL of each sample into a 2 mL 96-well microplate well.

14. Add 1 mL of extraction buffer (see 'Buffer and solvent preparation') to each well to precipitate proteins.

15. Seal the microplate with a silicone mat and vortex the sealed plate at the highest speed for 5 seconds.

16. Incubate samples on the benchtop at room temperature for 5 minutes.

17. Centrifuge samples for 10 minutes at 5,000 x g.

Sample type 2: Serum

12. Remove frozen serum samples from -80 °C and thaw on ice prior to extraction.

13. Pipette 200 µL of sample into a 2 mL 96-well microplate well.

14. Add 1 mL of extraction buffer (see 'Buffer and solvent preparation) to each well to precipitate proteins.

15. Seal the microplate with a silicone mat and vortex the sealed plate at the highest speed for 5 seconds.

16. Incubate samples on the benchtop at room temperature for 5 minutes.

17. Centrifuge samples for 10 minutes at 5,000 x g.

Sample type 3: Urine

12. Remove frozen urine samples from -80 °C and thaw on ice prior to extraction.

13. Dilute samples 1:20 in water to reach a total volume of 200 µL (10 µL urine sample in 190 µL of water) in a 2 mL 96-well microplate well.

14. Add 1 mL of extraction buffer (see 'Buffer and solvent preparation') to each well to precipitate proteins.

15. Seal the microplate with a silicone mat and vortex the sealed plate at maximum speed for 5 seconds.

16. Incubate samples on the benchtop at room temperature for 5 minutes.

17. Centrifuge samples for 10 minutes at 5,000 x g.

Sample type 4: Feces and/or intestinal contents

12. Remove frozen samples from -80 °C and thaw on ice prior to extraction.

13. Add ~30 mg of sample to a 2 mL screw cap vial containing previously added ~30 mg of acid-washed glass beads.

14. Add 600 µL of water and 600 µL of extraction buffer (see 'Buffer and solvent preparation') to each vial to precipitate proteins.

15. Homogenize vials using a mini bead beater at 4 °C for 5 minutes at 3500 oscillations per minute.

16. Incubate samples on the benchtop at room temperature for 5 minutes.

17. Centrifuge samples for 10 minutes at 5,000 x g.

Extract metabolites from all sample types above

18. Aliquot 440 µL of samples into wells of a new 2 mL 96-well microplate. Every sample should be aliquoted into duplicate plates. One plate will be used for LC-MS analysis, and the other will be archived at -80 °C as a backup. A total of 880 µL of supernatant should be collected from each sample between the duplicate 2 mL 96-well microplates.

19. Dry both plates of samples under air at 65 psi at 37 °C in a Turbovap evaporator. Continue this step until the solvent in all wells is completely evaporated. Seal one plate with a silicone mat fitted for the 96-well plate and store at -80 °C for archives as backup in case of re-analysis (e.g., instrument error, sample loss during the first analysis).

20. Reconstitute the remaining samples in 200 µL of reconstitution buffer (see 'Buffer and solvent preparation') and seal the plate with a silicone mat.

21. Vortex the sample plate at the maximum speed for 5 seconds.

22. Centrifuge the sample plate for 1 minute at 2,000 x g to spin down the residual sample stuck on the side of the well and/or on the silicone mat.

23. Transfer all 200 µL of samples into a 0.22 µm 96-well filter plate that is placed on top of a new 1 mL 96-well microplate, and centrifuge this plate stack for 10 min at 2,000 x g. For fecal or intestinal samples, centrifuge at5,000 x g for 10 minutes to fully pass the samples through the filter into the bottom plate.

24. Seal the sample plate with a silicone mat and store at -80 °C until LC-MS analysis.

PAUSE POINT Samples can be stored at -80 °C for up to 3 months prior to analysis.

25. Prior to LC-MS analysis, bring plate with reconstituted samples from -80 °C to room temperature.

26. Mix 4 µL of each sample contained in the same experiment into a user-designated well location on the same plate reserved for quality control (QC). This QC mix consists of a pool of all samples in the same experiment to provide a representation of all metabolomic features.

LC-MS analysis on the Agilent 6545 LC/Q-TOF instrument

Timing varies depending on the number of samples to run

The user operation instructions vary between instruments (e.g., Agilent Q-TOF vs. Thermo Orbitrap). The instructions below are provided for the use of Agilent Q-TOF 6545. The user should be trained by the professional staff at the core facility before operating the instrument independently.

27. Open the Agilent MassHunter Workstation Data Acquisition software (v.10.1), which is the user platform for implementing the subsequent steps.

28. Assemble a set of column and guard column of the same composition (e.g., both are C18 or Amide) to each analytical method, following manufacturer's instructions.

29. Load a column securely attached to a guard column on the column compartment.

CRITICAL STEP Ensure the column is connected properly and check for leaks from both ends of the column once connected. Leaky column reduces the column pressure and often terminates the run as leaked solvent accumulates in the column compartment.

30. Connect the mobile phase solvent bottles to the corresponding "A" and "B" solvent positions of the LC unit.

31. Select "Tune" in "Context" at the top left corner to calibrate the instrument at either the positive or the negative mode, depending on the method to run next (Fig. 7a-b for example tune settings). The Q-TOF instrument uses the Agilent tuning mix, which contains a collection of reference ions for tuning both positive and negative modes.

32. Select “Acquisition” in “Context” to return to the home page to set up the analysis run.

33. Create a worklist including the following required information: Sample Name (e.g., “mouse_serum_1”), Sample Position (e.g., “P1-A01”), Method (file path to the method file), Data File (file path to the folder storing the new data), Sample Type (e.g., “Sample” by default), and Injection Volume (e.g., “As Method”).

34. Introduce five blanks at the beginning of each worklist. The blank sample containing 50% methanol can be either stored in a 2 mL Agilent glass vial or an empty well on the same plate as the biological samples. The first three blanks are used for removing the residual contaminants in the system while equilibrating the column with fresh solvents. While the first three blanks often contain contaminants from the previously injections, the fourth and fifth blanks can be used to assess background noise intrinsic to the solvent mix.

35. Add one or more QC samples at the beginning, middle, and the end of the full worklist. This is to assess any intra-experimental shift in retention time.

36. Load each 96-well 1 mL sample plate onto the autosampler. The autosampler should be kept at 4^{o}C throughout the run to ensure sample integrity and to minimize sample evaporation.

37. Select an analytical method (see “Equipment setup”) and click “run” to initiate the worklist.

38. As the instrument progresses through the worklist, each newly generated .d raw file populates the folder previously designated in the “Data File” column of the worklist.

Data analysis using the MS-DIAL software

Timing 0.5 d

39. We follow the workflow as described in the MS-DIAL documentation (https://mtbinfo-team.github.io/mtbinfo.github.io/) with the following specifications. We use the MS-DIAL version 3.83 that is downloadable via the archive folder.

40. Convert raw data files to Analysis Base File (.abf) format using an open source Abf Converter (https://www.reifycs.com/AbfConverter/). Depending on the mass spectrometry instrument, the raw data file types may vary. For example, Agilent Q-TOF generates .d files, and Thermo Orbitrap generates .raw files. Regardless of instruments, the final converted file(s) should have the .abf file extension before subsequent analysis in MS-DIAL.

41. Create a new project in MS-DIAL in “New Project Window” by specifying the file path containing all ABF files, and selecting the following options (Fig. 8a):

a. Ionization type: “Soft ionization”

b. Separation type: “Chromatography”

c. MS method type: “Conventional LC/MS or data dependent MS/MS”

d. Data type (MS1): “Centroid data” or “Profile” (depending on how data are collected on the mass spectrometry instrument)

e. Ion mode: “Positive ion mode” or “Negative ion mode” (depending on the analytical method used for data collection)

f. Target omics: Metabolomics

42. Designate each file in the “Type” column in “New Project Window” as “Sample”, “QC”, or “Blank”. “QC” or quality control samples contain a pool of all samples used in an experiment (Fig. 8b), as described above. Select the specific set of files to analyze, and modify the column “Type” to define each file as “sample”, “blank”, or “QC”. The “QC” files will be used to generate a master list of molecular features for peak alignment of all samples. The “blank” files will be used to provide a list of molecular features present in the solvent background. Do not load the first three blanks at the beginning of an experiment, as they are used for cleaning purposes.

43. Load a parameter settings file (.med file extension) containing recommended parameters (source data: c18positive_settings.med, c18negative_settings.med, or hilicpositive_settings.med) by clicking on the “Load” button (Fig. 9a). This will enable MS-DIAL to retrieve and auto-populate our recommended parameters specific to each method (Fig. 9a-g), using the C18 positive parameters as an example). The parameter settings provided in this protocol have been used in analyzing data generated by an Agilent Q-TOF instrument and are broadly applicable to data analysis generated via other instruments.

44. In the “Identification” tab, specify a reference library used for compound annotation based on the *m/z* and RT information. Upload a library file (e.g., the one created at the end of step 11). in the “Identification” tab, under “Advanced: Text file and post identification (retention time and accurate mass based) setting”.

CRITICAL STEP The reference library file needs to be reloaded for each new analysis and is not included in the recommended setting file. The choice for the library file depends on the analytical method used to collect the data.

45. Prior to analyzing the full dataset using the full library, carry out a preliminary analysis using a small number of samples from the full experiment (e.g., several QCs collected throughout an experiment). This preliminary analysis gauges the retention time shifts of the internal standards detected in the current experiment compared to those reported in the reference library. For the reference library, use one of the internal standard libraries specific to each analytical method. These libraries will instruct the software to only annotate peaks for the matching internal standards. If all internal standards of the experiment fall within $\pm$0.2 minute from the library’s retention time, proceed to analyze the full data using the full library.

? TROUBLESHOOT (see Supplementary Table 3)

46. Export data by selecting “Alignment result export” (Fig. 9h). Users can select different output files based on their preferences. The most common output is “Raw data matrix (Area),” where the area under the curve of each peak is used to quantify ion abundance of each metabolite. Export format as a .txt file, which can be opened as an excel spreadsheet. The MS-DIAL analysis generates a list of *m/z*, RT, and ion counts (area under the curve) for 1) annotated metabolites (matched to the reference library based on based on *m/z* and RT) and 2) detected but unknown molecular features (those not matched to the reference library).

47. Manually filter each set of aligned peaks corresponding to each annotated metabolite in MS-DIAL (Fig. 10a). Select the following settings in the graphic user interface: 1) the analysis file name in “Alignment navigator,” 2) “Identified” option in the “Peak spot navigator”, 3) “EIC of aligned spot” in the top middle window, and 4) “Alignment spot viewer” in the bottom middle window.

48. Visually inspect each aligned peak shape and the area under the curve. Remove the entire metabolite data row from the alignment result excel spreadsheet (see Step 45) for the problematic peaks: 1) odd peak curvature resulting in only a subset of the peak being counted for the area under the curve (Fig. 10b), or 2) the peak is only detected in the blank controls but not in the samples.

Data analysis using the bioinformatics workflow

One code repository (https://github.com/the-han-lab/Han_et_al_Metabolomics_Protocol_2022) can be downloaded to provide an example workflow (Fig. 2) for integrating MS-DIAL data output with sample metadata within and across experiments. Detailed documentation for input files and code logic are explained in detail in data_analysis.ipynb.

Troubleshooting

Supplementary Table 3. Troubleshooting Table

Time Taken

Anticipated Results

The output from the bioinformatics analysis workflow contains two types of files: 1) a final fold-change data matrix containing all data from different experiment, sample types, and analytical methods, and 2) a metadata file describing properties associated with each sample used in this analysis. In addition, a built-in statistical analysis in the example workflow enables quick identification of metabolites with significantly altered levels (e.g., produced or consumed) between samples and their controls (e.g., mono-

colonized vs. germ-free mouse serum). All example input and output files are provided as a part of the code and its associated documentation (https://github.com/the-han-lab/Han_et_al_Metabolomics_Protocol_2022).

References

1. Koppel, N., Rekdal, V. M. & Balskus, E. P. Chemical transformation of xenobiotics by the human gut microbiota. *Science.* **356**, 1246–1257 (2017).

2. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).

3. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science.* **349**, (2015).

4. Husted, A. S., Trauelsen, M., Rudenko, O., Hjorth, S. A. & Schwartz, T. W. GPCR-Mediated Signaling of Metabolites. *Cell Metab.* **25**, 777–796 (2017).

5. Krautkramer, K. A., Fan, J. & Bäckhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nat. Rev. Microbiol.* **19**, 77–94 (2021).

6. Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A. K. & Dorrestein, P. C. Mass spectrometry-based metabolomics in microbiome investigations. *Nat. Rev. Microbiol.* **20**, 143–160 (2022).

7. Han, S. *et al.* A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* **595**, 415-420 (2021).

8. Mars, R. A. T. *et al.* Longitudinal Multi-omics Reveals Subset-Specific Mechanisms Underlying Irritable Bowel Syndrome. *Cell* **182**, 1460-1473.e17 (2020).

9. Holmes, E., Wijeyesekera, A., Taylor-Robinson, S. D. & Nicholson, J. K. The promise of metabolic phenotyping in gastroenterology and hepatology. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 458–471 (2015).

10. Fiehn, O. Metabolomics by gas chromatography-mass spectrometry: Combined targeted and untargeted profiling. *Current Protocols in Molecular Biology* **2016**, (2016).

11. Rinschen, M. M., Ivanisevic, J., Giera, M. & Siuzdak, G. Identification of bioactive metabolites using activity metabolomics. *Nat. Rev. Mol. Cell Biol.* **20**, 353–367 (2019).

12. Melnik, A. V. *et al.* Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. *Anal. Chem.* **89**, 7549–7559 (2017).

13. Liebisch, G. *et al.* Quantification of fecal short chain fatty acids by liquid chromatography tandem mass spectrometry—investigation of pre-analytic stability. *Biomolecules* **9**, 1–13 (2019).

14. Song, H. E., Lee, H. Y., Kim, S. J., Back, S. H. & Yoo, H. J. A facile profiling method of short chain fatty acids using liquid chromatography-mass spectrometry. *Metabolites* **9**, (2019).

15. Boulet, L., Faure, P., Flore, P., Montérémal, J. & Ducros, V. Simultaneous determination of tryptophan and 8 metabolites in human plasma by liquid chromatography/tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1054**, 36–43 (2017).

16. Yan, J., Kuzhiumparambil, U., Bandodkar, S., Solowij, N. & Fu, S. Development and validation of a simple, rapid and sensitive LC-MS/MS method for the measurement of urinary neurotransmitters and their metabolites. *Anal. Bioanal. Chem.* **409**, 7191–7199 (2017).

17. Scherer, M., Gnewuch, C., Schmitz, G. & Liebisch, G. Rapid quantification of bile acids and their conjugates in serum by liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.***877**, 3920–3925 (2009).

18. Bobeldijk, I. *et al.* Quantitative profiling of bile acids in biofluids and tissues based on accurate mass high resolution LC-FT-MS: Compound class targeting in a metabolomics workflow. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **871**, 306–313 (2008).

19. Wegner, K. *et al.* Rapid analysis of bile acids in different biological matrices using LC-ESI-MS/MS for the investigation of bile acid transformation by mammalian gut bacteria. *Anal. Bioanal. Chem.* **409**, 1231–1245 (2017).

20. Kind, T. *et al.* Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **37**, 513–532 (2018).

21. Cajka, T., Smilowitz, J. T. & Fiehn, O. Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography-High-Resolution Mass Spectrometry Platforms. *Anal. Chem.* **89**, 12360–12368 (2017).

Acknowledgements

The authors thank Stanford ChEM-H for use of the LC/MS qTOF instrument; W. Van Treuren, D. Dodd, M.A. Fischbach, C.R. Fischer, B.C. DeFelice, and J.M. Sanchez for helpful input on method development and optimization. This work is supported by R01-DK085025, DP1-AT009892, R01-DK101674, and a Stanford Discovery Innovation Fund Award (J.L.S.), and NRSA F32AG062119, and Duke Science and Technology Initiative (S.H.).

Figures

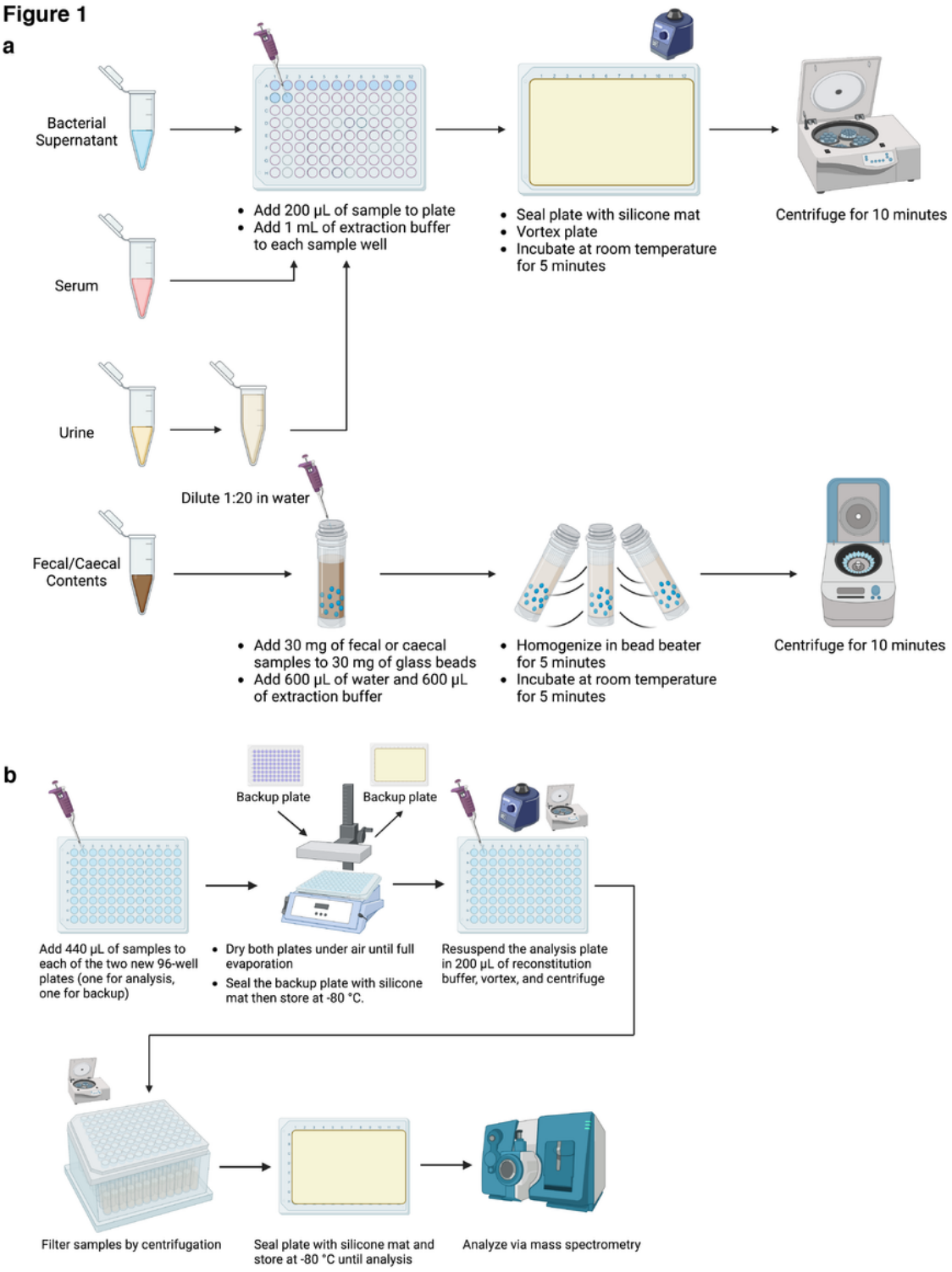


Figure 1

Schematic of metabolite extraction from diverse sample types used in microbiome studies. a, Steps for homogenizing samples followed by protein precipitation. Four main sample types are provided as examples: bacterial supernatant, host serum, urine, and feces/caecal contents. **b,** Steps for extracting metabolites from all sample types listed above. Duplicate plates are generated from extraction of a single

sample plate: one is used for downstream mass spectrometry analysis, and the other is saved as a backup plate for future analyses.

Figure 2

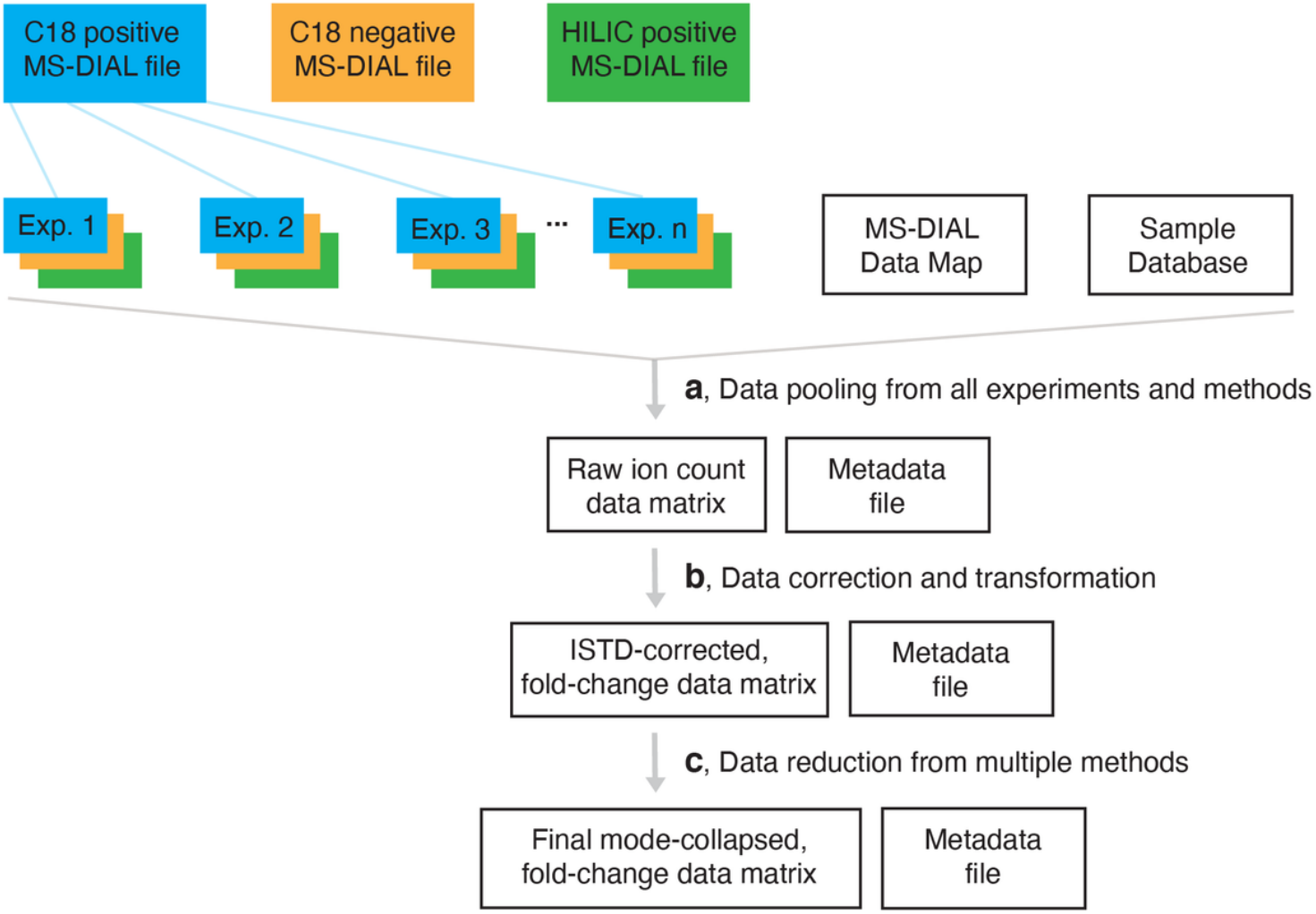


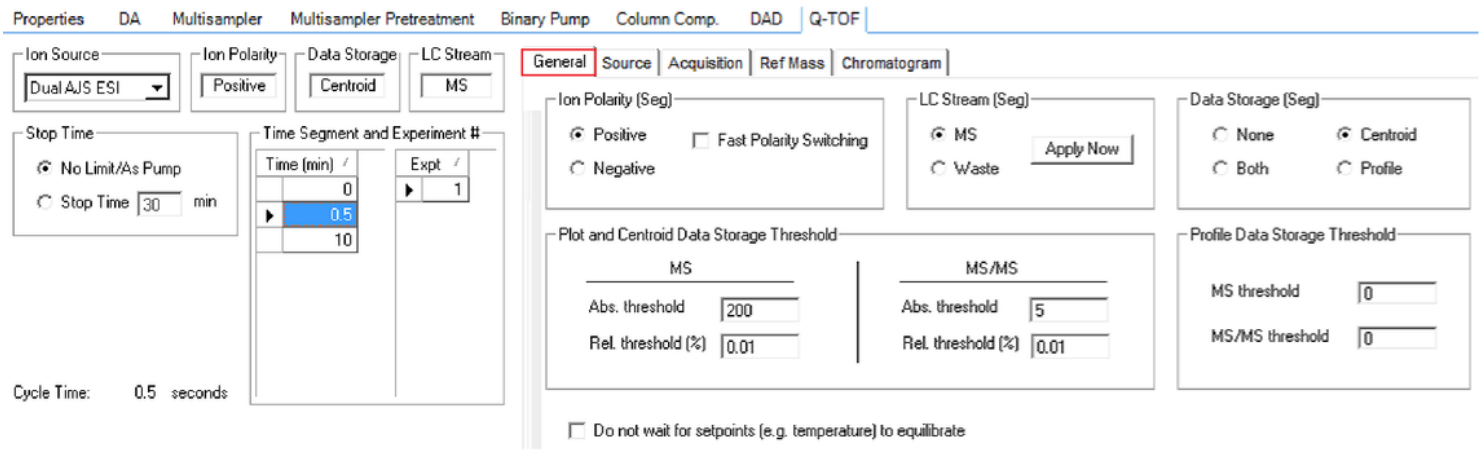
Figure 2

Example bioinformatics workflow for integrating data from multiple experiments, sample types, and analytical methods. a, MS-DIAL data files, MS-DIAL data map, and sample database are integrated into a single data matrix of raw ion counts across all experiments from three analytical methods. Blue rectangles, C18 positive method. Orange rectangles, C18 negative methods. Green rectangles, HILIC positive methods. **b**, A shared set of robustly detected internal standards (ISTDs) among samples from one or multiple experiments is used to correct for intra- and inter-experimental variations in instrumental sensitivity. The ISTD-corrected matrix is generated by dividing the raw ion count of each metabolite by the sum of ISTDs specific to each experiment type (e.g., mono-colonization) and sample type (e.g., serum). In addition to the ISTD-corrected matrix, a fold-change data matrix is generated by calculating the relative fold change in ISTD-corrected ion count between the experimental samples (e.g., colonized mice) vs. controls (e.g., germ-free mice). The relative fold change calculation is carried out on a per-experiment type and per-sample type basis. **c**, When a metabolite is detected by multiple analytical methods, its fold-

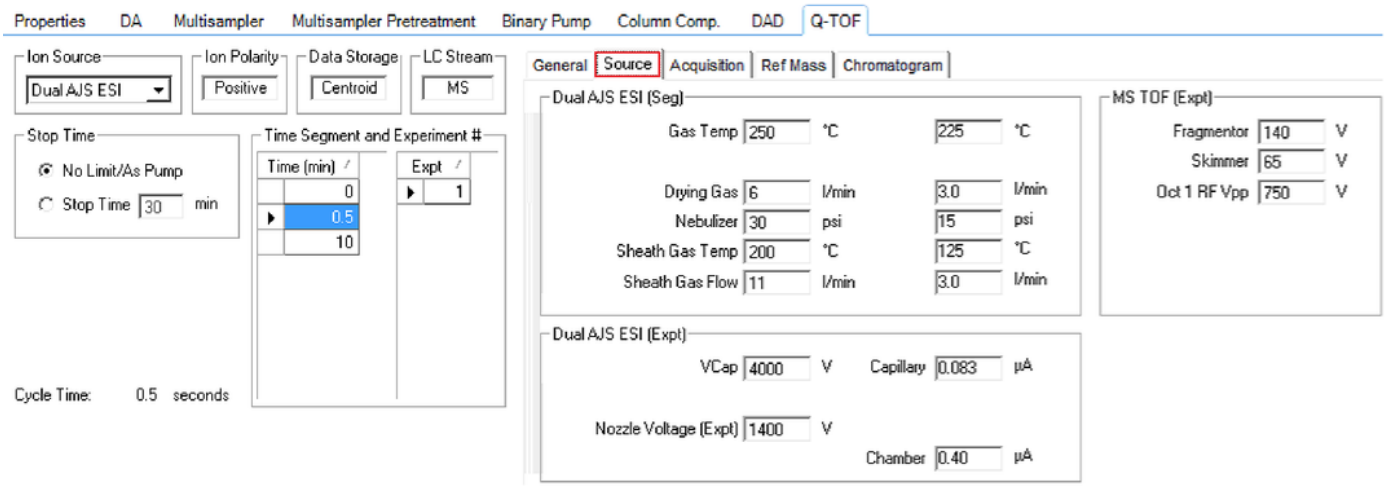
change values are averaged among the preferred methods. Preferred methods for each metabolite are selected based on the consistency of detection among replicates for each method. Throughout this analysis, a separate metadata file is generated to detail the associated properties of each sample analyzed in each method. Lastly, this example workflow is created based on the common host sample types in microbiome studies, and it can be adapted to bacterial culture analysis.

Figure 3

a



b



c

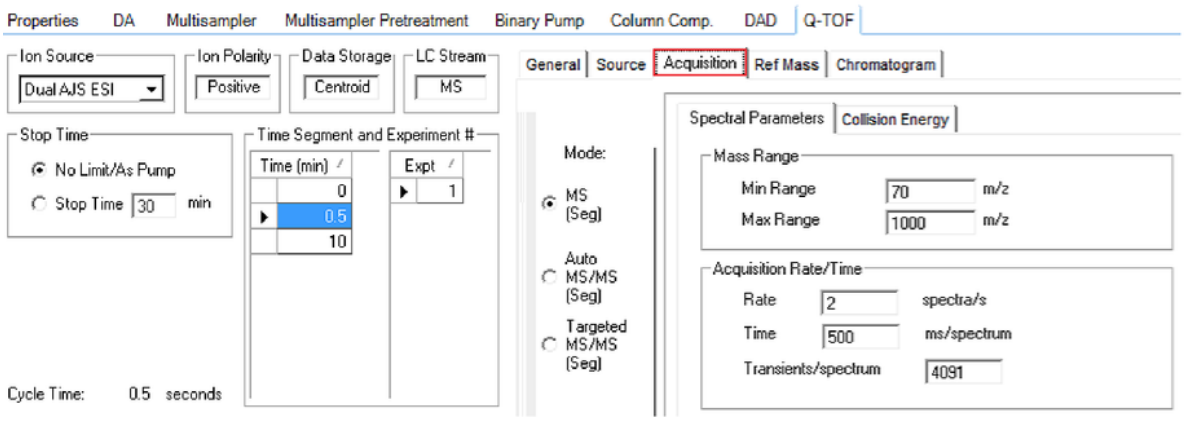
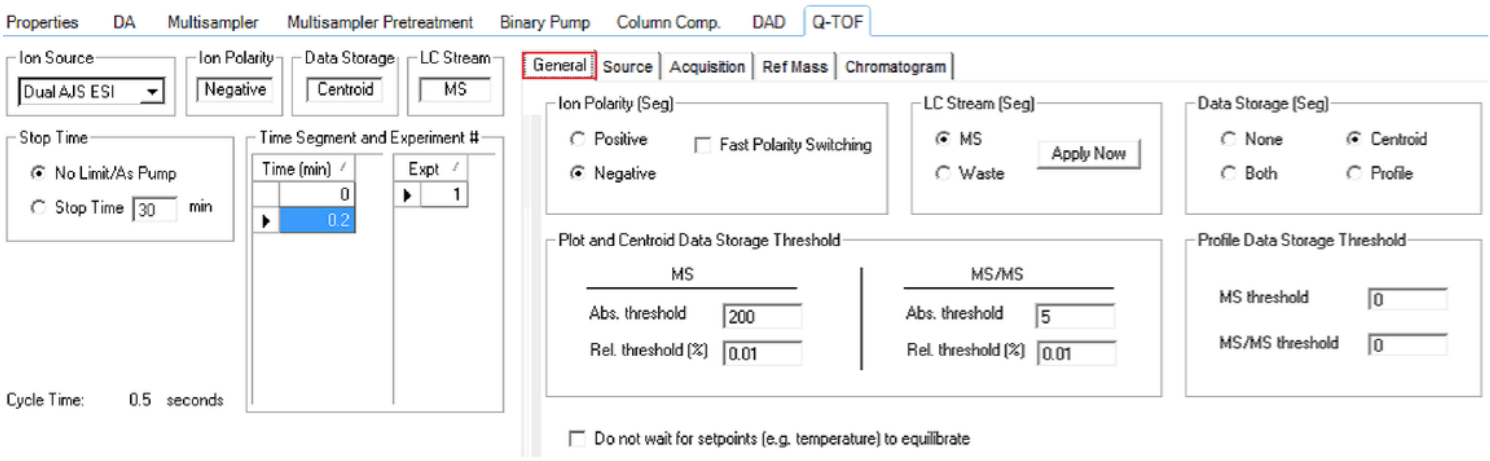


Figure 3

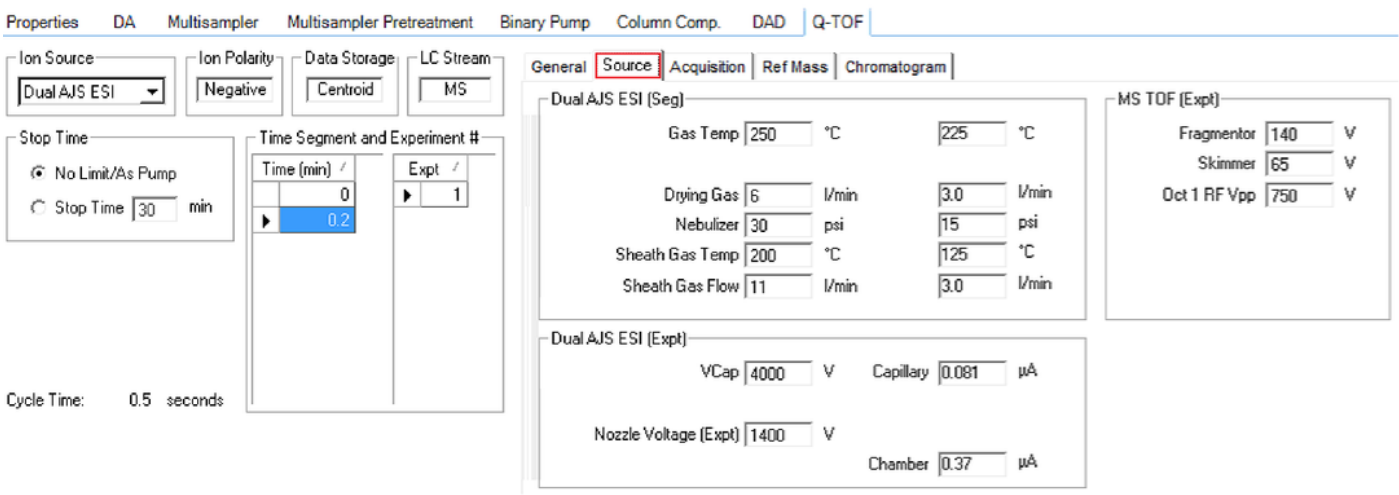
Agilent Q-TOF example settings for the C18 positive analytical method. These settings are created using the Agilent MassHunter Workstation Data Acquisition software (Agilent MassHunter in short), which communicates directly with the Q-TOF instrument. The example shown is for the time segment (0.5-10 minutes) where data are collected. **a**, General settings. **b**, Ion source settings. **c**, Acquisition settings. Red boxes highlight each tab in the software that corresponds to each type of settings.

Figure 4

a



b



c

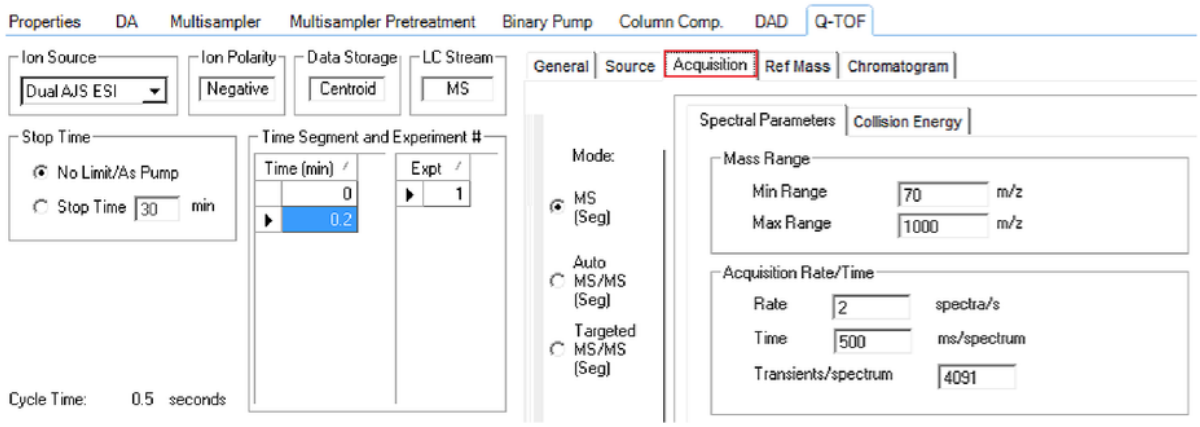
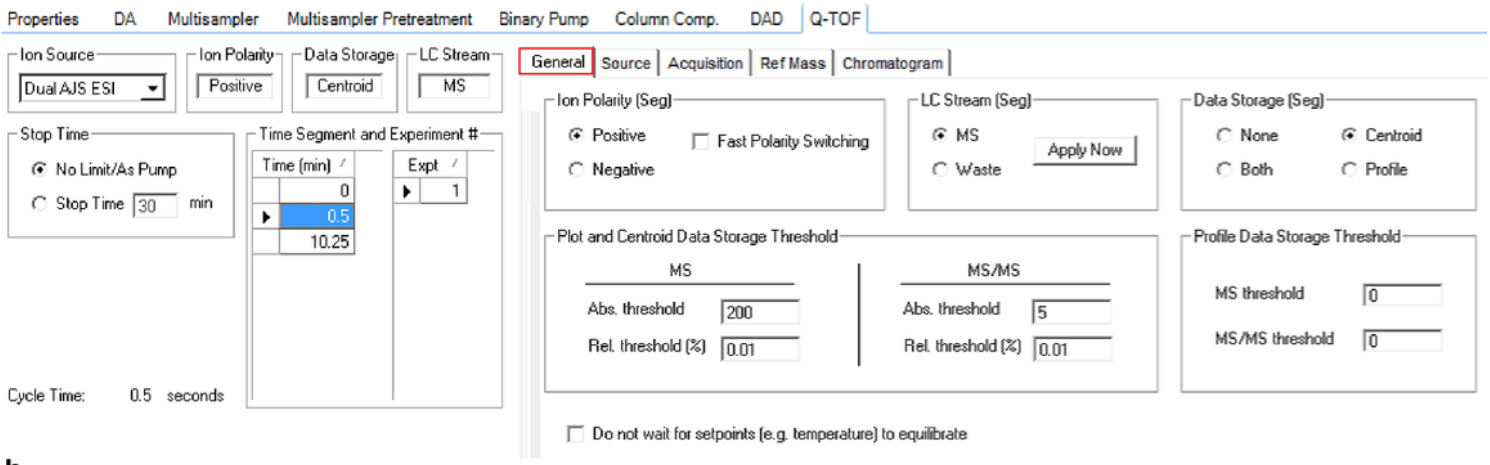


Figure 4

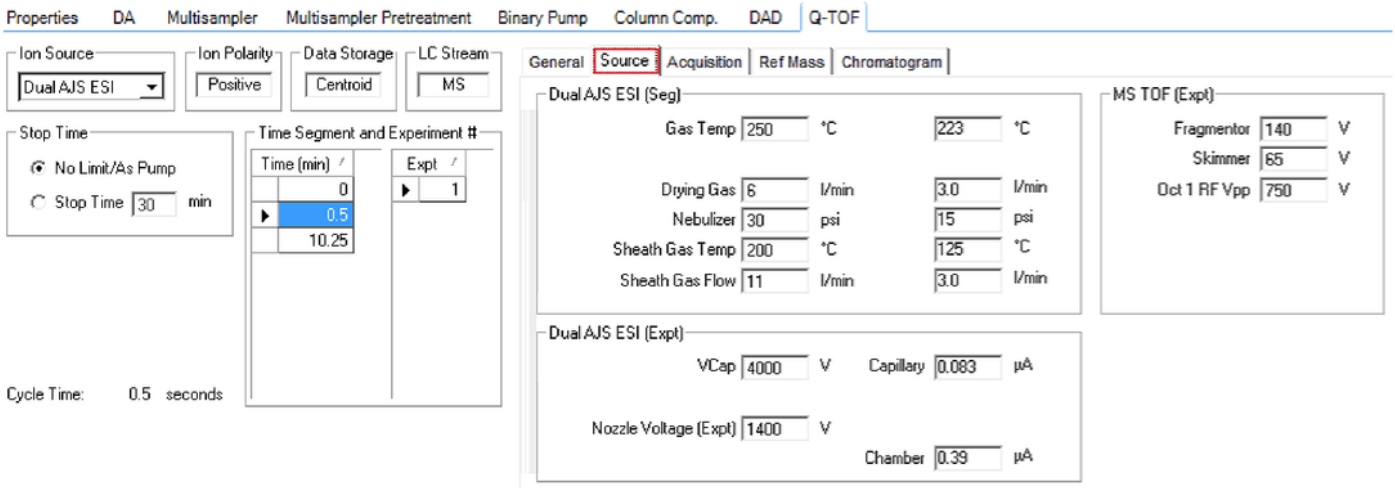
Agilent Q-TOF example settings for the C18 negative analytical method on Agilent MassHunter. The example shown is for the time segment (0.2-10 minutes) where data are collected. **a**, General settings. **b**, Ion source. **c**, Acquisition. Red boxes highlight each tab in the software that corresponds to each type of settings.

Figure 5

a



b



c

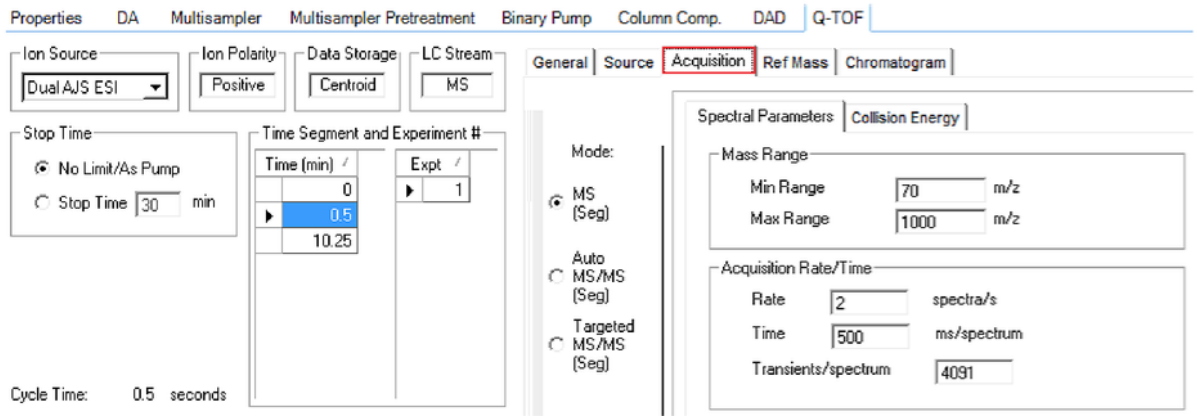


Figure 5

Agilent Q-TOF example settings for the HILIC positive analytical method on Agilent MassHunter. The example shown is for the time segment (0.5-10.25 minutes) where data are collected. **a**, General settings. **b**, Ion source. **c**, Acquisition. Red boxes highlight each tab in the software that corresponds to each type of settings.

Figure 6

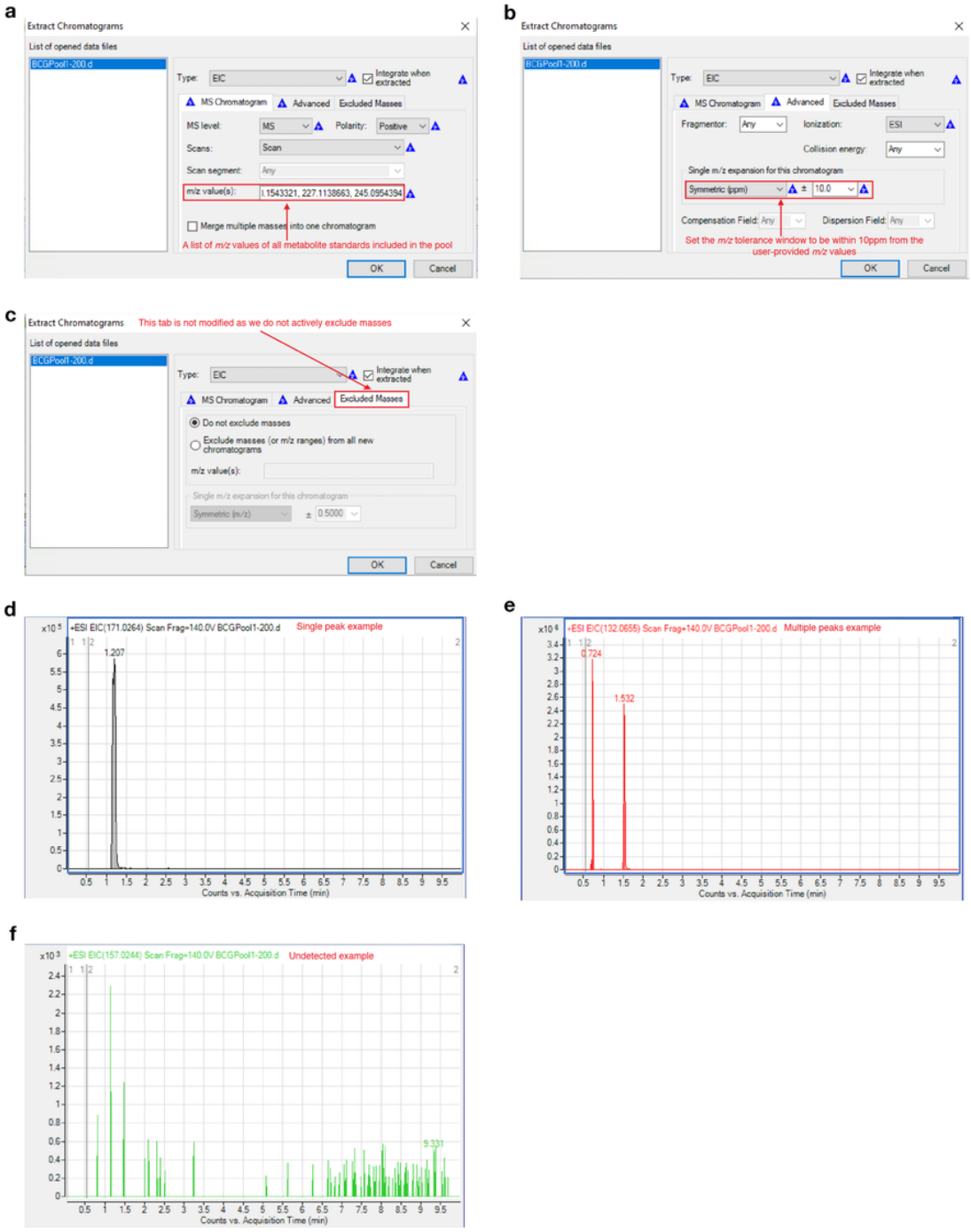


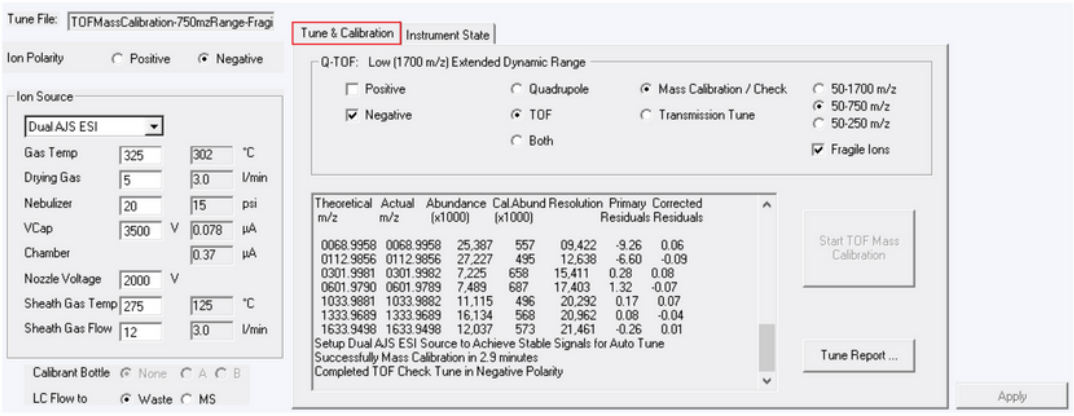
Figure 6

Extract chromatogram settings and example images of extracted ion chromatograms. These settings are created using the Agilent MassHunter Qualitative Data Analysis software (Agilent Qual in short), which is used for RT identification based on user provided *m/z* values. **a**, MS Chromatogram settings. **b**, Advanced

settings. **c**, Excluded masses (this tab is not used in this protocol). **d**, Example of a single peak identified from extracting chromatogram from a single *m/z* value. **e**, Example of multiple peaks identified from extracting chromatogram based on a single *m/z* value. **f**, Example of no peak identified from extracting chromatogram based on a single *m/z* value. Red boxes and associated arrows highlight explanations for specific input.

Figure 7

a



b

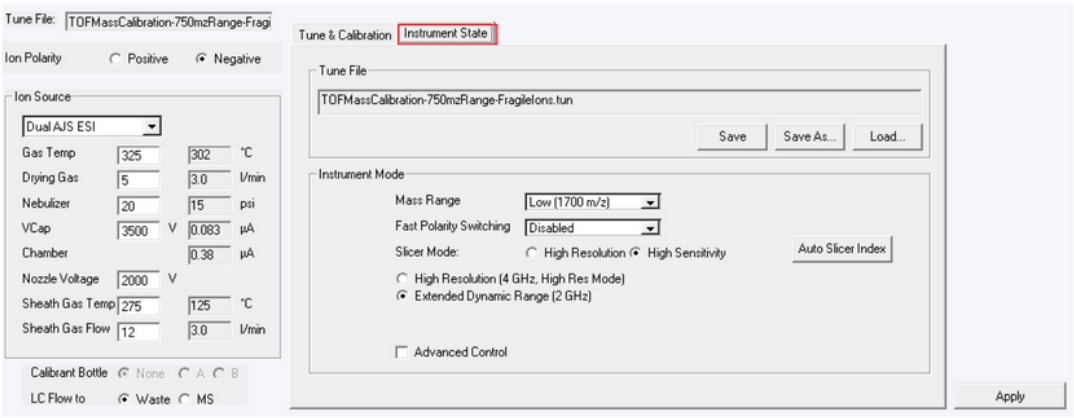


Figure 7

Agilent Q-TOF mass calibration settings on Agilent MassHunter. Example calibration shown is in the negative ionization mode (e.g., C18 negative method). For the positive ionization methods (C18 positive, HILIC positive), these calibration settings remain identical with the exception of shifting ion polarity from “negative” to “positive” in both “Ion polarity” and “Q-TOF” sections. **a**, Tune and calibration. **b**, Instrument state. Red boxes highlight each tab in the software that corresponds to each type of settings.

Figure 8

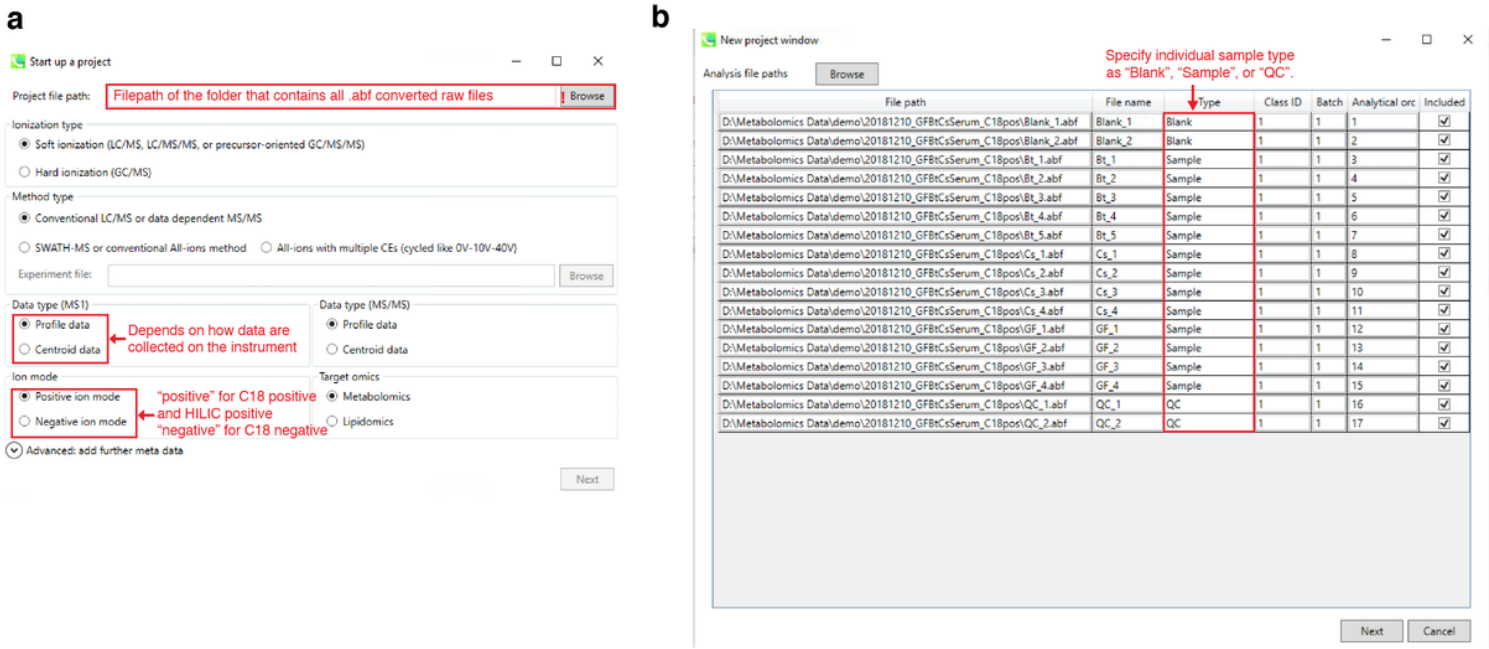


Figure 8

Project settings using the MS-DIAL software. **a**, General project information. **b**, Project file information. Red boxes and associated arrows highlight explanations for specific input.

Figure 9

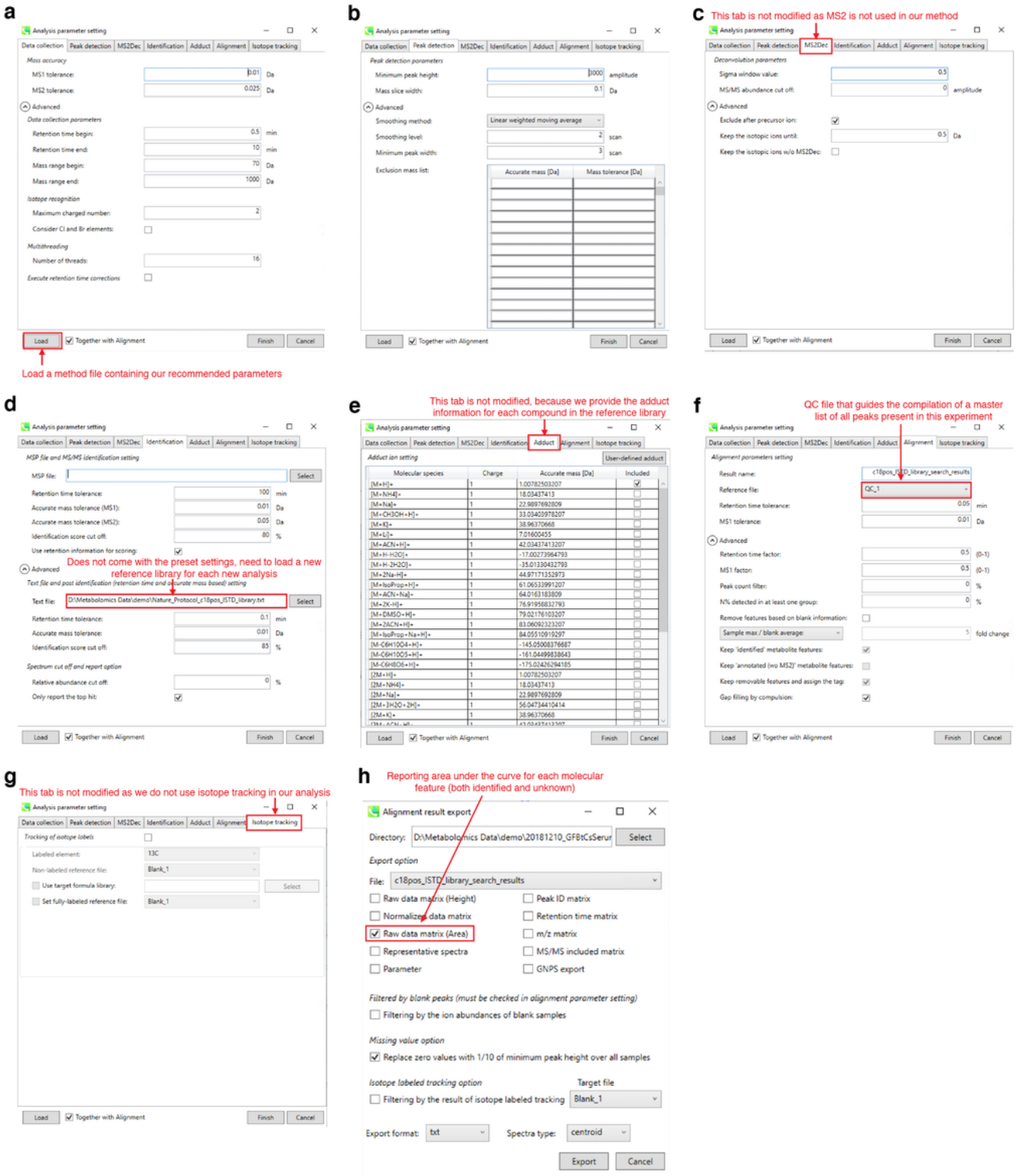


Figure 9

MS-DIAL analysis parameters and alignment result export settings. **a**, Data collection. **b**, Peak detection. **c**, MS2 deconvolution (not used in this protocol). **d**, Identification. **e**, Adduct (not modified here, but information is provided in the reference library input file in "Identification." **f**, Alignment. **g**, Isotope tracking (not used in this protocol). **h**, Alignment result export. Red boxes and associated arrows highlight explanations for specific input.

Figure 10

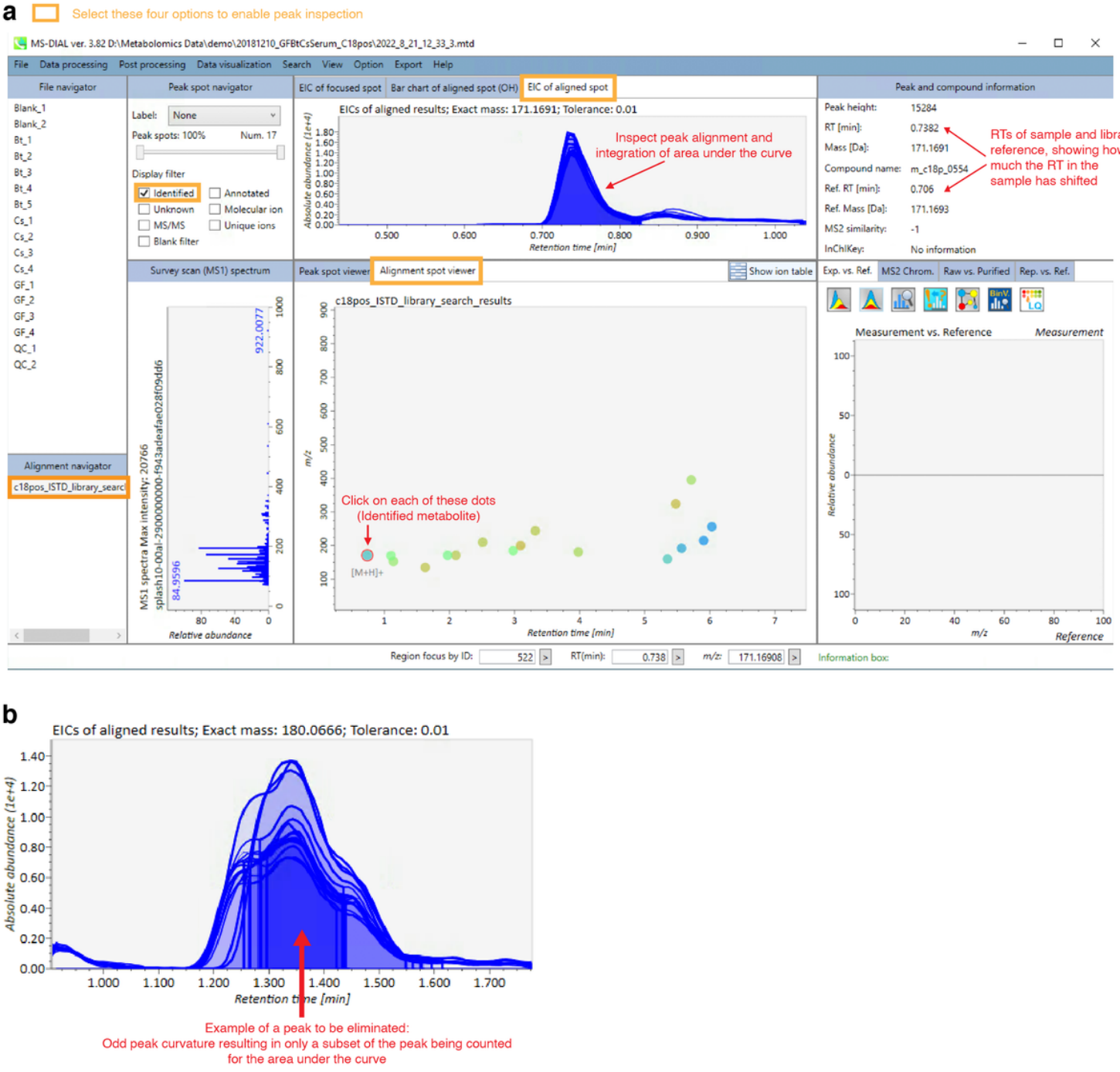


Figure 10

MS-DIAL user interface for checking the quality of annotated peaks. **a**, Example of visualizing an analyzed dataset on MS-DIAL. **b**, Example of a low-quality peak to be eliminated from the output result file. Orange boxes highlight specific items to select for visualization. Red arrows highlight specific explanations or instructions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx