Chapter 19

Metabolomics Analyses from Tissues in Parkinson's Disease

Fatema Bhinderwala, Shulei Lei, Jade Woods, Jordan Rose, Darrell D. Marshall, Eli Riekeberg, Aline De Lima Leite, Martha Morton, Eric D. Dodds, Rodrigo Franco, and Robert Powers

Abstract

Metabolomics has been successfully applied to study neurological and neurodegenerative disorders including Parkinson's disease for (1) the identification of potential biomarkers of onset and disease progression; (2) the identification of novel mechanisms of disease progression; and (3) the assessment of treatment prognosis and outcome. Reproducible and efficient extraction of metabolites is imperative to the success of any metabolomics investigation. Unlike other omics techniques, the composition of the metabolome can be negatively impacted by the preparation, processing, and handling of these samples. The proper choice of data collection, preprocessing, and processing protocols is similarly important to the design of an effective metabolomics experiment. Likewise, the correct application of univariate and multivariate statistical methods is essential for providing biologically relevant insights. In this chapter, we have outlined a detailed metabolomics workflow that addresses all of these issues. A step-by-step protocol from the preparation of neuronal cells and metabolomic tissue samples to their metabolic analyses using nuclear magnetic resonance, mass spectrometry, and chemometrics is presented.

Key words Parkinson's disease, Neurodegeneration, Metabolomics, Chemometrics, NMR, Mass spectrometry

1 Introduction

Parkinson's disease (PD), the second most common neurodegenerative disorder worldwide, is characterized by the selective loss of dopaminergic neurons of the substantia nigra pars compacta (SNpc) [1]. There is no current treatment to stop neuronal cell death progression or to cure PD. Thus, to find neuroprotective strategies, a clear understanding of the mechanism(s) involved in dopaminergic cell death is needed. Mitochondrial dysfunction and

Fatema Bhinderwala, Shulei Lei and Jade Woods contributed equally to this chapter.

Sanjoy K. Bhattacharya (ed.), *Metabolomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1996, https://doi.org/10.1007/978-1-4939-9488-5_19,

the concomitant alterations in redox homeostasis and bioenergetics (energy failure) are thought to be a central component of PD [2–4]. One means of analyzing the state of a biological system is by monitoring the metabolome, i.e., all the metabolites present in a cell, biofluid, tissue, organ, or organism [5, 6]. In this regard, metabolomics is the study of the changes in the concentration and the identity of these metabolites that result from environmental or genetic stress or from a disease state or drug treatment. A better understanding of the biological phenotype during disease development and progression may be achieved by identifying and quantifying variations in metabolite levels. In essence, metabolomics provides a top-down view of complex biological systems. Accordingly, metabolomics has evolved to become an important resource for systems biology and a valuable tool to study disease states [7]. Metabolomics has been successfully applied to study neurological and neurodegenerative disorders [8]. Indeed, previous studies have demonstrated the applicability of metabolomics in (1) the identification of potential biomarkers of PD diagnosis, onset, and progression [9–11]; (2) the identification of novel mechanisms of disease progression [12–15]; and (3) the assessment of treatment prognosis and outcome [16]. Using metabolomics, we and others have established a link between the alterations in central carbon metabolism induced by PD risk factors, redox homeostasis, and bioenergetics and their contribution to the survival or death of dopaminergic cells [2].

Unlike other omics techniques, the composition of the metabolome can easily change from the processing, handling, and storage of samples [17]. Metabolites may chemically transform or degrade due to residual enzymatic activity, from oxidation, from low chemical stability, or from other chemical activity. Thus, robust and reproducible isolation of metabolites is a key step in the metabolomics workflow. Univariate and multivariate statistical analyses are also an important aspect of a metabolomics study [18]. But the incorrect application of statistical techniques, the insufficient preprocessing, the lack of proper model validation, and the overinterpretation of models and outcomes are all common concerns that often lead to erroneous or misleading biological insights from metabolomics data [18]. Metabolomics has commonly relied on mass spectrometry (MS) [19] or nuclear magnetic resonance (NMR) [20] as the primary analytic source for sample analysis. Again, a successful metabolomics investigation is dependent on appropriate protocols for data collection, processing, and analysis. To address these issues, we have provided a detailed, step-by-step description of a metabolomics workflow specifically applicable to the analysis of brain cell cultures and tissues used in our research using PD experimental models (*see* Fig. 1). We describe methods to assist in the efficient cell culturing, metabolite extraction, and data collection and analyses. Alongside, we discuss a combined NMR

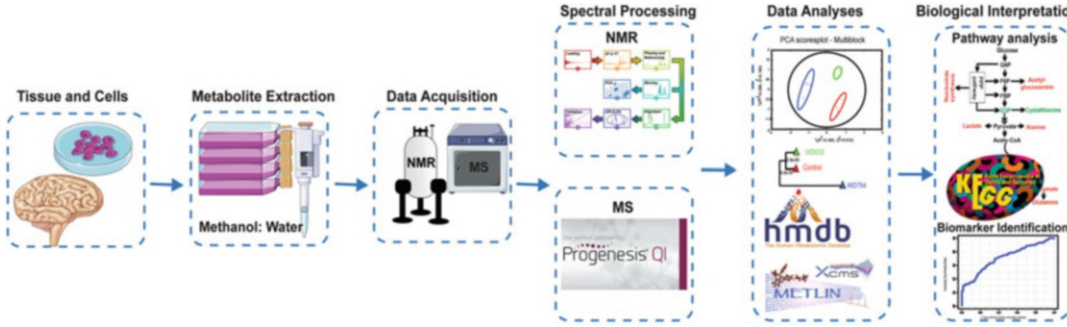

Fig. 1 A schematic diagram is shown that outlines the overall metabolic workflow used in the analysis of brain cell cultures and tissues from experimental PD models. Only the major protocol steps are highlighted in the flow diagram. The figure was generated using free medical images from Servier Medical Art (https:/smart.servier.com/) under the Creative Commons License Attribution 3.0 Unported (CC BY 3.0)

and MS approach to improve metabolome coverage, which allows for the identification of key neurological metabolites. While the protocols outlined in this chapter have been developed using PD experimental models, most of the methodology may be universally applied to any metabolomics study.

2 Materials

Prepare all aqueous solutions and buffers with either Nanopure H_2O or deuterated water (D_2O). Please follow all safety regulations in regard to handling biological samples and the disposal of both chemical and biological waste. A valuable rule of thumb in the handling of all tissues, biofluids (e.g., blood, urine, etc.), and cell lines is to assume a contamination with a virus, pathogen, or toxin and to handle the samples accordingly.

2.1 Laboratory Equipment

1. Bruker AVANCE III HD 700 MHz NMR spectrometer equipped with a 5 mm quadruple resonance QCI-P cryoprobe (^{1}H, ^{13}C, ^{15}N, and ^{31}P) with z-axis gradients, an automatic tune and match system (ATM), and a SampleJet automated sample changer system with Bruker ICONNMR software (Bruker BioSpin, Billerica, MA).
2. Synapt G2 HDMS quadrupole time-of-flight (TOF) MS instrument equipped with an ESI source (Waters, Milford, MA).
3. Waters ACQUITY M-class Xevo G2-XS QToF MS instrument equipped with an ESI source (Waters, Milford, MA).
4. BSL-2 biosafety level grade hood (e.g., Biological Safety Cabinet, LF BSC class 2 type A, Thermo Fisher Scientific, Waltham, MA).

5. Nanopure ultra water system (Barnstead Inc., Dubuque, IA).
6. Lab Armor bead bath (Chemglass Life Sciences, Vineland, NJ).
7. Incubator capable of maintaining physiological temperature and proper carbon dioxide levels (e.g., Heracell VIOS 250i CO_2 Incubator, Thermo Fisher Scientific, Waltham, MA).
8. pH meter and probe.
9. Refrigerated centrifuge capable of speeds up to 21,100 $\times g$ (e.g., Sorvall micro 21R centrifuge, Thermo Fisher Scientific, Waltham, MA).
10. SpeedVac for solvent removal (e.g., Savant SC210A SpeedVac concentrator, Thermo Fisher Scientific, Waltham, MA).
11. Freeze dryer to remove water (e.g., FreeZone 4.5, Labconco, Kansas City, MO).
12. 1000 μL to 1 μL pipettes.
13. FastPrep-96 homogenizer (MP Biomedicals, Santa Ana, CA) for brain tissue analysis, which uses Lysing Matrix D.
14. Accu-Scope 3030 ph microscope (Commack, NY).
15. Cryogenic storage container (Taylor Wharton, Theodore, Al).
16. −80 °C freezer.

2.2 Disposable Supplies

1. 1 mL to 1 μL pipette tips.
2. 10 mL aspirating pipettes.
3. 15 mL Falcon tubes.
4. 2 mL Eppendorf tubes.
5. 1 mL screw-cap microcentrifuge tubes.
6. LC-MS certified total recovery vial (Waters, Milford, MA).

2.3 Isotopically Labeled Solvents and Reagents (See Notes 1 and 2)

1. Deuterium oxide (D_2O, 99.8 atom %D).
2. 3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt (TMSP-d_4, 99.8 atom % D).
3. Dimethyl sulfoxide-d_6 (DMSO-d6, 99.8 atom %D).
4. $^{13}C_6$-glucose (99% ^{13}C).
5. $^{13}C_2$-acetate (99% ^{13}C).
6. Other potential ^{13}C-carbon-labeled or ^{15}N-nitrogen-labeled reagents.

2.4 Buffers

1. Wash buffer, phosphate-buffered saline (PBS) at pH 7.4: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4. To prepare 1 L PBS buffer at pH 7.4, add 8.0 g of NaCl, 0.2 g of KCl, 2.68 g of $Na_2HPO_4{\cdot}7H_20$, and 0.24 g of KH_2PO4 to a final volume of 1 L of Nanopure water.

2. NMR buffer: 50 mM phosphate buffer at pH 7.2 (uncorrected, *see* **Note 3**) in 600 μL of 99.8% D_2O. Add 50 μM (one-dimensional [1D] NMR experiment) or 500 μM (two-dimensional [2D] NMR experiment) TMSP-d_4 as an internal chemical shift reference.
3. MS extraction buffer: Mix 20 mL LC-MS grade water with 80 mL LC-MS grade methanol. Store at −40 °C.
4. MS reconstitution solution: LC-MS grade water with 0.1% LC-MS grade formic acid.
5. LC mobile phase A: LC-MS grade water with 0.1% LC-MS grade formic acid.
6. LC mobile phase B: LC-MS grade acetonitrile/methanol with 0.1% LC-MS grade formic acid.

2.5 Cell Lines and Media

1. For cell cultures, we have used human dopaminergic neuroblastoma cell lines such as SK-N-SH (HTB-11, ATCC, Manassas, VA) [15], SH-SY5Y (CRL-2266, ATCC), N27 immortalized rat dopaminergic cells (SCC048, EMD Millipore, Temecula, CA) [21], human immortalized midbrain neuronal precursors LUHMES (CRL-2927, ATCC), and primary rat/mouse astrocytes [22] following the specifications of the commercial providers or published protocols.
2. Cell culture media and supplements are obtained from commercial vendors such as GIBCO/Life Technologies (Grand Island, NY), Fisher Scientific, Hyclone (GE Healthcare, Logan, UT), and Atlanta Biologicals (Flowery Branch, GA).

2.6 Software and Databases

1. Bruker ICONNMR software for automated NMR data acquisition (Bruker BioSpin, Billerica, MA).
2. MVAPACK metabolomics toolkit for processing and analyzing chemometric data (http://bionmr.unl.edu/mvapack.php) [23].
3. PCA/PLS-DA utilities for quantifying separation in PCA, PLS-DA, and OPLS-DA scores plots (http://bionmr.unl.edu/pca-utils.php) [24].
4. NMRPipe software for processing and visualizing NMR data (https://www.ibbr.umd.edu/nmrpipe/install.html) [25].
5. NMRViewJ software for processing and visualizing NMR data (One Moon Scientific, Inc. Westfield, NJ; https://nmrfx.org/) [26].
6. MassLynx V4.1 (Waters Corp., Milford, MA) for mass spectral data processing (http://www.waters.com/waters/en_US/MassLynx-Mass-Spectrometry-Software-/nav.htm?locale=en_US&cid=513164).

7. Progenesis QI (version 2.0, Nonlinear Dynamics, Newcastle, UK) for processing and analysis of LC-MS data (http://www.nonlinear.com/progenesis/qi/).
8. R statistical package (https://www.r-project.org/) [27].
9. Chenomx (Chenomx, Inc., Edmonton, AB, Canada) software for automated metabolite assignment and quantification from 1D ^{1}H NMR spectra (https://www.chenomx.com/).
10. MZmine software (http://mzmine.github.io/download.html) for metabolite identification from MS data [28].
11. MetaboAnalyst software for the statistical, functional, and integrative analysis of metabolomics data (http://www.metaboanalyst.ca/) [29].
12. ChemSpider chemical structure database http://www.chemspider.com/ [30].
13. Human Metabolome Database (HMDB) of reference NMR and mass spectral data for known metabolites (http://www.hmdb.ca/) [31].
14. Biological Magnetic Resonance Data Bank (BMRB) of reference NMR data for known metabolites (http://www.bmrb.wisc.edu/metabolomics/) [32].
15. Nonuniform schedule (NUS) generator (http://bionmr.unl.edu/dgs-gensched.php) for NUS NMR data acquisition [33].

3 Methods

3.1 Experimental PD Models

The etiology of PD has yet to be clearly established. The major risk factor identified for PD is aging as its prevalence and incidence increases exponentially from ages 65 to 90 [21]. A fraction of PD occurrence (~10%) is related to mutations in genes such as those encoding α-synuclein (*SNCA/PARK1-4*), DJ-1 (*PARK7*), PTEN-induced putative kinase 1 (*PINK1/PARK6*), leucine-rich repeat kinase 2 (*LRRK2/PARK8*), and parkin (*PARK2*) [22, 34]. However, over 85% of PD occurs in a sporadic (idiopathic) form without a clearly defined genetic basis. Epidemiological studies suggest that lifestyle, occupational, and environmental exposures can increase the risk of developing PD [35–37]. Thus, it is thought that PD arises from the convergence of genetic susceptibility, environmental exposures, and aging.

Cellular and animal disease models based on both genetic-, toxin- or stress-induced neurodegeneration have been used to understand PD pathogenesis [36, 38] (*see* Fig. 2). However, not all experimental models recapitulate all PD hallmarks in their entirety. Genetically engineered PD mouse models have been developed for the overexpression of mutant genes

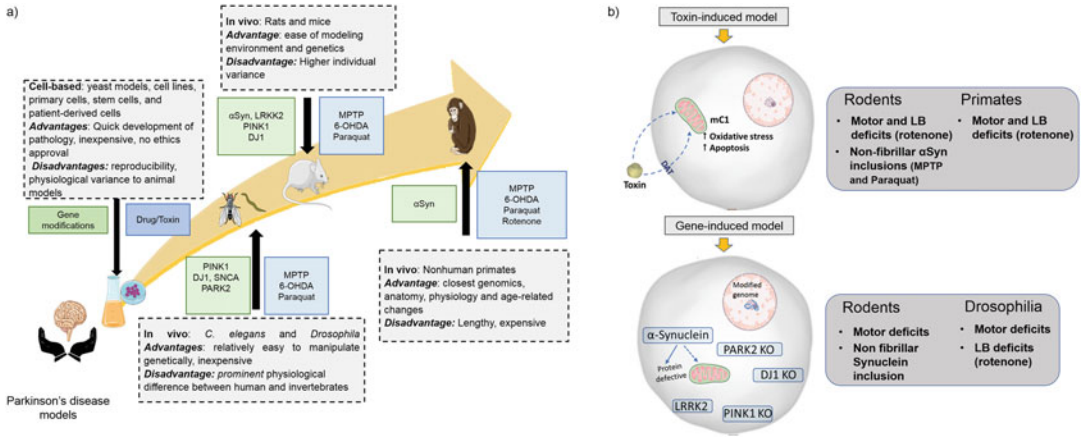

Fig. 2 Common models of PD. (**a**) A summary of advantages and disadvantages of common models of PD. (**b**) A list of some model-specific characteristics observed for different PD models. The figure was generated using free medical images from Servier Medical Art (https://smart.servier.com/) under the Creative Commons License Attribution 3.0 Unported (CC BY 3.0)

[34, 36]. However, only marginal or null dopaminergic cell death has been observed in genetic-based animal models. Recent advances in mammalian genome engineering technology have led to the generation of rat PD models that seem to better reproduce PD hallmarks including progressive loss of dopaminergic neurons, locomotor behavior deficits, and age-dependent formation of abnormal α-synuclein protein aggregates (Lewy bodies) [39].

On the other hand, the use of mitochondrial/environmental toxins such as 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP, or its active metabolite 1-methyl-4-phenylpyridine MPP^+) and the pesticides rotenone and paraquat that induce dopaminergic cell death in vitro and in vivo is supported by clinical and epidemiological studies [36]. Several other toxicants such as metals, diverse pesticides, polychlorinated biphenyls, diet, as well as inflammatory processes have been implicated as PD risk factors [40, 41]. However, it is clear that not a single environmental exposure is responsible for all PD cases nor are they the single cause for PD. Accordingly, new models studying gene-environment interactions have also emerged.

For the most part, experimental PD models are designed to reproduce one or more key aspects of PD pathogenesis including genetic modifications, mitochondrial dysfunction, oxidative stress, accumulation of misfolded aggregates and impaired proteostatic processes, alterations in dopamine metabolism, and inflammation [36]. Experimental PD models have helped to identify important mechanisms regulating dopaminergic cell death and survival, and they should continue to enhance our understanding of PD pathogenesis. In our metabolomics investigations, we have used

neuronal-like cell cultures of neuroblastoma cells and immortalized midbrain dopaminergic cells from rats and humans exposed to PD-related insults and gene-environment interactions. In addition, we have also evaluated changes in the metabolome of mice exposed to pesticides and heavy metals linked to PD or parkinsonism [15, 42]. The protocol described below is a general protocol for isolating and characterizing changes in the metabolome applicable to different types of cell cultures and brain.

3.2 Cell Culture

Cell culture procedures must follow published guidelines to avoid misidentification and contamination [43]. We recommend to start with one 100 mm^2 dish of 90% confluent cells per sample/replica, but if the metabolite is abundant enough, this can be reduced to a smaller sample size.

For PD-related insults, cells can be treated with mitochondrial toxins (MPP^+ or rotenone), pesticides (paraquat or dieldrin) or the overexpression of PD-related genes (WT or mutant forms of α-synuclein via viral vectors or conventional transfection procedures), as explained in our previous publications [15, 42]. The exact dose and time course must be determined empirically, but we recommend to work with a dose that will induce cell death of ~50% within no less than 48 h as neurodegeneration is a slow process and evaluate changes in the cellular metabolome prior to any detection of cell death (~24 h of treatment) (*see* **Notes 4–6** for considerations in regard to cell survivability, sample handling, and randomization).

3.3 Unlabeled Metabolomics Sample Number/Replicas

Use the maximal number of replicates per group that is possible (*see* **Note 7**). A typical number of replicate cultures per group is ten. Adjust the number of replicates given practical considerations, such as the number of groups, but the number of replicates per group should not be below six.

3.4 Isotopically Labeled Metabolomics Samples

Identify the ^{13}C-, ^{15}N-, or other isotopically labeled tracers. The tracer should be in accordance with the metabolic pathway of interest and expected to be affected by the experimental treatment. $^{13}C_6$-glucose is a common choice for a tracer since it highlights central carbon metabolism (glycolysis and TCA cycle), but a variety of other tracers may be used. Supplement the culture media with the appropriate ^{13}C-carbon-labeled source (*see* **Note 8**).

3.5 Extracting Water-Soluble Metabolites from PD Cell Cultures

All samples should be kept on ice or at 4 °C during sample preparation or handling. Samples should be stored at −80 °C, but, ideally, samples should be immediately analyzed. In addition to keeping samples cold, there are four other issues that are critical to the successful preparation of metabolomics samples: (1) speed, (2) consistency, (3) random processing of samples, and (4) the efficient removal of all biomolecules and cell debris [6]. The processing of all

metabolomics samples should proceed as quickly as possible while minimizing any loss in quality. Metabolites can chemically degrade or transform within milliseconds due to enzymatic activity, oxidation, chemical instability, or any number of other chemical processes [44]. Accordingly, rapidly inactivating and removing all biomolecules and cell debris (usually through methanol/ethanol precipitation) that may transform or bind a metabolite is a necessary step of the protocol (*see* **Note 9**).

1. Collect 1 mL of the media for metabolomics analysis. In addition to the cell extract, the media should also be analyzed for metabolomics changes as many metabolites get exchanged or effluxed outside of the cell. In this regard, the media is treated simply as another cell extract.
2. Wash the cells twice with 5 mL of PBS to remove debris. Discard the wash.
3. Lyse and quench cells with 1 mL of prechilled methanol at −20 °C. Incubate cells at −80 °C for 15 min.
4. Using a cell scraper, detach and collect cell debris and methanol in a 2 mL microcentrifuge tube. Confirm cell detachment using a microscope and repeat lyse and quenching if necessary.
5. Centrifuge the 2 mL microcentrifuge tube for 5 min at 15,000 × *g* and 4 °C to pellet the cell debris.
6. Collect the supernatant and transfer to a new 2 mL microcentrifuge tube.
7. Repeat the metabolome extraction by adding 0.5 mL of an 80%/20% mixture of methanol/water kept at −20 °C to the cell pellet.
8. Centrifuge the cell pellet with the extraction solvent for 5 min at 15,000 × *g* at 4 °C to pellet the cell debris.
9. Collect the supernatant, and transfer it to the 2 mL microcentrifuge tube containing the original methanol extract. Combine the two extraction supernatants into a single tube.
10. Repeat the metabolome extraction a third time by adding 0.5 mL of ice cold water to the cell pellet.
11. Centrifuge the cell pellet with the extraction solvent for 5 min at 15,000 × *g* at 4 °C to pellet the cell debris.
12. Collect the supernatant, and transfer it to the 2 mL microcentrifuge tube containing the two previous extraction supernatants. Combine the three extraction supernatants into a single tube.
13. Split the sample into two 2 mL Eppendorf tube. Aliquot 100 μL for MS analysis, and the remainder of the sample is used for NMR analysis.

14. Use a SpeedVac or a rotary evaporator to remove the methanol.
15. Flash-freeze the samples in liquid nitrogen.
16. Remove the water and bring to dryness using a lyophilizer.
17. Repeat **steps 1–16** for each replicate and for each group (*see* **Note 6**).
18. Store samples in a −80 °C freezer or proceed to preparing the NMR and/or MS samples (*see* Subheadings 3.7 and 3.8).

3.6 Extracting Water-Soluble Metabolites from Mouse Brain Tissue

1. Similar to cell culture treatments, a number of experimental paradigms have been used to model PD in vivo [27, 36]. We have used the subchronic exposure to pesticides and metals [15], but the protocol described can be applied to all murine animal models.
2. We have successfully used 200 mg/kg of $^{13}C_6$-glucose at a total volume of 100 μL administered to fasted mice (overnight) via intra-orbital injection to label metabolites extracted from mouse brain tissue (Fig. 3).

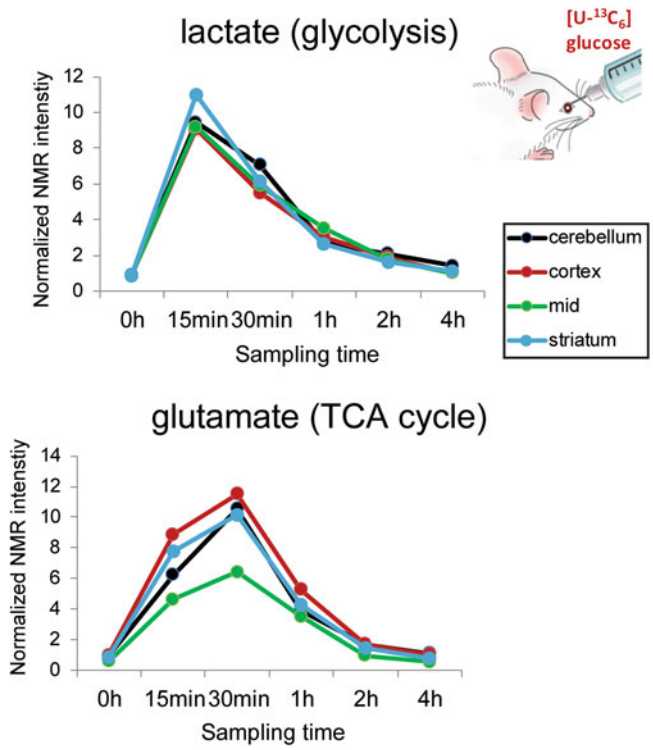

Fig. 3 In vivo evaluation $^{13}C_6$-glucose metabolism. Fasted mice (overnight) were administered ^{13}C-glucose (200 mg/kg body weight, 100 μL) via retro-orbital injection, and brain regions were dissected at the time indicated for NMR analysis

3. Harvest and dissect the mice brain tissue (15–20 min after the injection of ^{13}C-labeled tracer if used, *see* Fig. 3).
4. Transfer the tissue to a 2 mL microcentrifuge tube containing Lysing Matrix D, and weigh the amount of tissue harvested from the mice, and immediately freeze the tissue with liquid nitrogen.
5. Extract the tissue with a 1:1 mixture of methanol and water prechilled to −20 °C. The volume of the extraction solvent depends upon the weight of the tissue.
6. Homogenize the sample in a FastPrep with Lysing Matrix D at 1300 rpm for 20 s and for two cycles.
7. Incubate the tissue at −80 °C for 10 min to extract the metabolome.
8. Centrifuge at 1000 × *g* for 10 min at 4 °C to remove tissue debris.
9. Collect the supernatant and transfer to a new 2 mL microcentrifuge tube.
10. Repeat the metabolome extraction by adding 0.7 mL of 1:1 mixture of methanol and water prechilled to −20 °C to the tissue pellet.
11. Repeat **steps 6–8** and combine the supernatant with the previous extract.
12. Normalize the metabolomics sample to the tissue weight by diluting all of the samples to a final volume of 1.5 mL. Add as much of a 1:1 mixture of methanol and water prechilled to −20 °C as needed to achieve a final volume of 1.5 mL.
13. Split the sample into two 2 mL Eppendorf tubes. Aliquot 100 μL for MS analysis, and the remainder of the sample is used for NMR analysis.
14. Use a SpeedVac or a rotary evaporator to remove the methanol.
15. Flash-freeze the samples in liquid nitrogen.
16. Remove the water and bring to dryness using a lyophilizer.
17. Repeat **steps 3–16** for each replicate and for each group (*see* **Note 6**).
18. Store samples in a −80 °C freezer or proceed to preparing the NMR and/or MS samples (*see* Subheadings 3.7 and 3.8).

3.7 Preparation of NMR Samples

1. For one-dimensional (1D) NMR experiments, lyophilized cell-free lysates or tissue extracts are suspended in 600 μL of 100% 50 mM D_2O phosphate buffer (uncorrected pH 7.2) with 50 μM 3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt (TMSP-d_4).

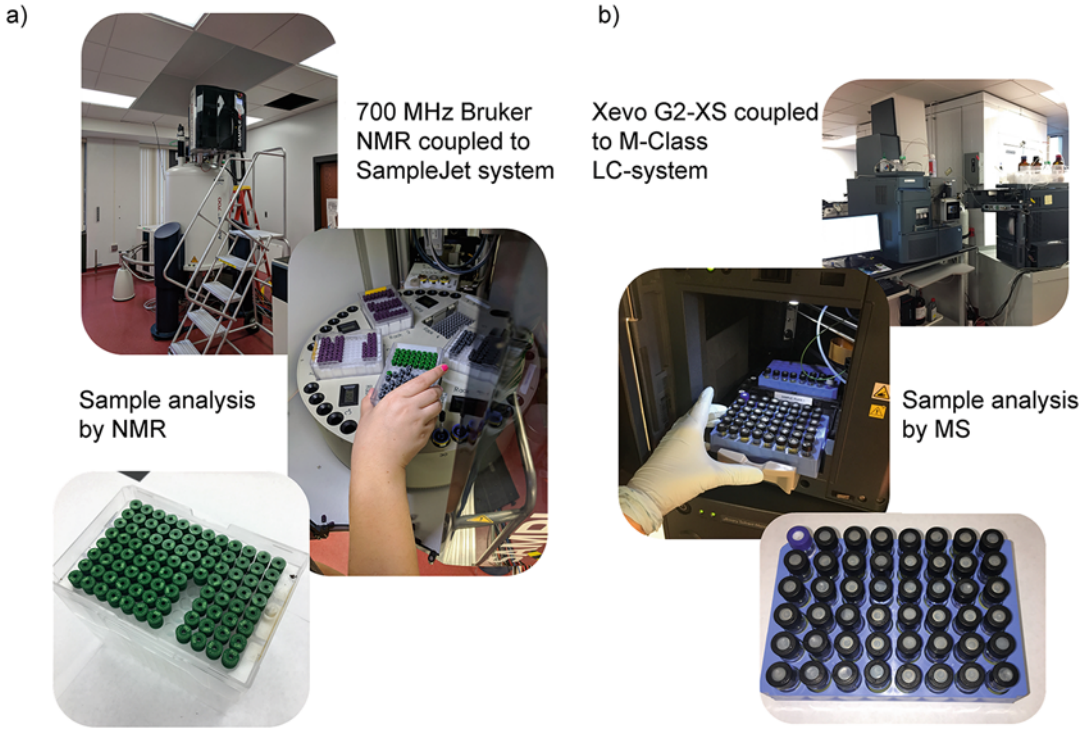

Fig. 4 High-throughput sample preparation. Images illustrating the loading of replicate metabolomics samples into the (**a**) 96-well plate SampleJet configuration and (**b**) the LC-MS autosampler

2. For two-dimensional (2D) NMR experiments, lyophilized cell-free lysates or tissue extracts are suspended in 600 μL of 100% 50 mM D_2O phosphate buffer (uncorrected pH 7.2) with 500 μM TMSP-d_4.
3. Centrifuge the sample at 14,000 × *g* for 10 min to remove any particulates.
4. The sample is transferred to a 4″ 5 mM SampleJet NMR tube with a pipette (*see* **Note 10**).
5. Repeat **steps 1–4** for each replicate and for each group (*see* **Note 6**).
6. Each sample is added to a 96-well plate SampleJet configuration equilibrated to 4 °C to prevent metabolite degradation (*see* Fig. 4).

3.8 Preparation of Mass Spectrometry Samples

1. Dissolve lyophilized cell-free lysates or tissue extracts in 20 μL of reconstitution solution and vortex for 30 s.
2. Centrifuge the solution at 14,000 × *g* for 10 min to remove any particulate matter.
3. Transfer the supernatant to LC vials and keep them in wet ice.

4. Repeat **steps 1–3** for each replicate and for each group (*see* **Note 6**).
5. Prepare quality control (QC) samples by pooling a 1 μL aliquot from each biological sample and transferring to a new LC vial labeled as QC.
6. Place all vials into the autosampler equilibrated to 4 °C to prevent metabolite degradation (*see* Fig. 4).

3.9 NMR Data Collection

All NMR experiments are conducted at 298 K using a Bruker AVANCE III HD 700 MHz spectrometer equipped with a 5 mm quadruple resonance QCI-P cryoprobe (^{1}H, ^{13}C, ^{15}N, and ^{31}P) with *z*-axis gradients. An automatic tune and match system (ATM) and a SampleJet automated sample changer system with Bruker ICONNMR software were used to automate the NMR data collection (*see* Fig. 5).

3.9.1 1D ^{1}H NMR

1. Load the NMR metabolomics samples into the SampleJet automated sample changer system (*see* Fig. 4). Check that the SampleJet is in the correct mode (i.e., 5 mm tubes).
2. Log into an account on the spectrometer workstation and start the Topspin software.
3. The first NMR sample is lowered into the magnet using the Bruker command, *sx 101*, where 101 corresponds to sample one in rack one.
4. The spectrometer is locked onto the D_2O solvent frequency using the Bruker command, *lock* D_2O.
5. The NMR sample is shimmed for optimal signal and suppression of the water signal by typing the Bruker command *topshim*. This will initiate an automated gradient shimming procedure, which may take a few minutes to complete (*see* **Note 11**).
6. The sample is automatically tuned and matched using the ATM system by typing the Bruker command *atma*.
7. The 90-degree pulse length (μs) is determined by measuring a null spectrum with an approximate 360-degree pulse using the Bruker *zg* pulse sequence (*see* **Note 12**).
8. A 1D ^{1}H NMR spectrum is obtained for each sample using a standard excitation sculpting water suppression pulse program (Bruker *zgesgp* pulse sequence) that provides optimal suppression of the residual water signal while maintaining a flat baseline (*see* **Note 13**).
9. Typical experimental parameters for a 1D ^{1}H NMR spectrum obtained on a Bruker 700 MHz spectrometer with a cryoprobe correspond to 128 scans, 16 dummy scans, 32,768 data points, a spectral width of 11,160.7 Hz, and a relaxation delay of 1.5 (*see* **Note 14**).

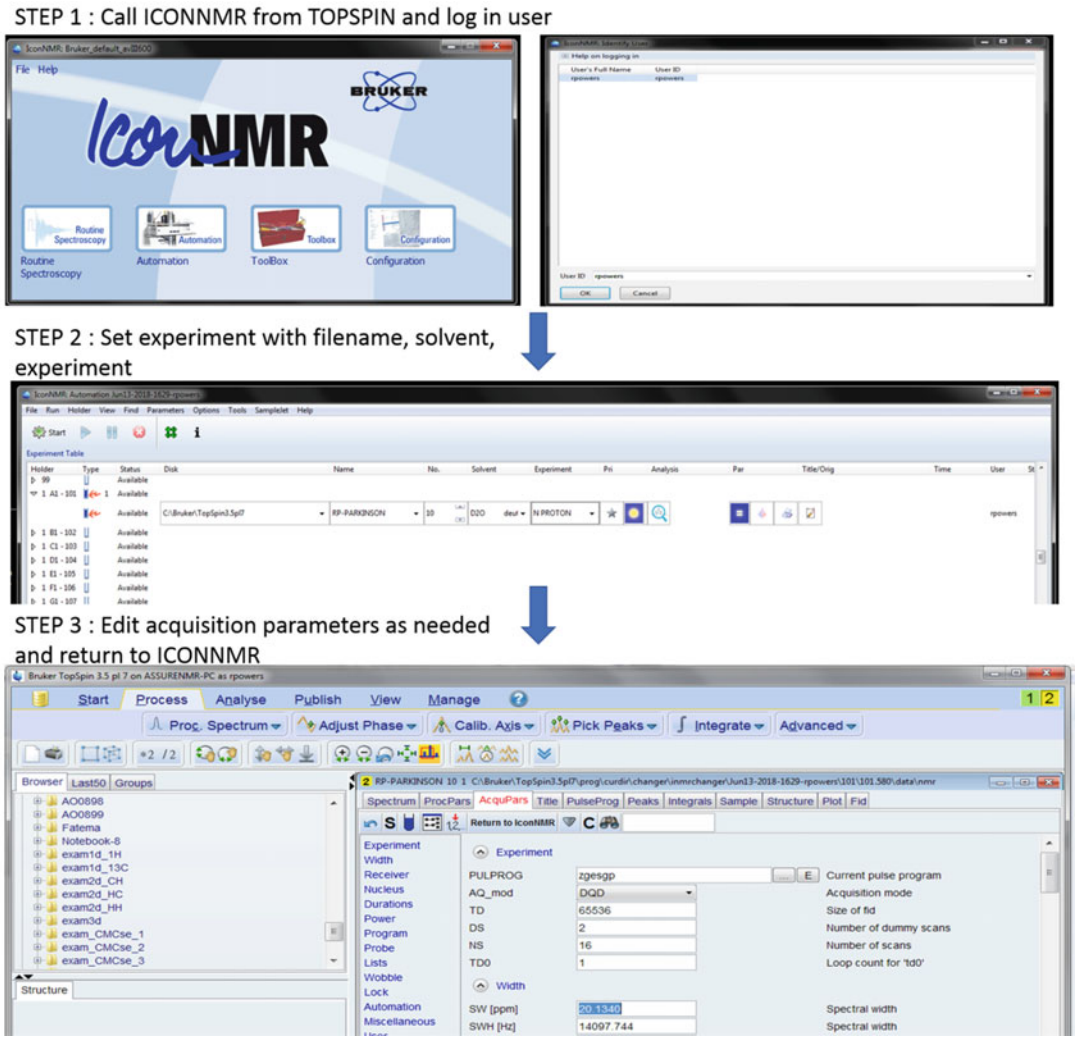

Fig. 5 High-throughput NMR data collection. ICONNMR screenshots illustrating the stepwise workflow for setting up a high-throughput 1D ^{1}H NMR metabolomics screen

10. Automated data collection of the entire set of metabolomics samples is accomplished using Bruker ICONNMR 5 (*see* Fig. 5).
11. The sample filename, solvent, pulse program, and temperature parameters are all defined in Bruker ICONNMR 5 (*see* **Notes 15–17**).
12. Collect the 1D ^{1}H NMR spectrum for each replicate and each group (*see* **Note 6**).
13. The data is processed initially with Topspin to verify spectral quality but exported for further analysis (*see* Fig. 6a).

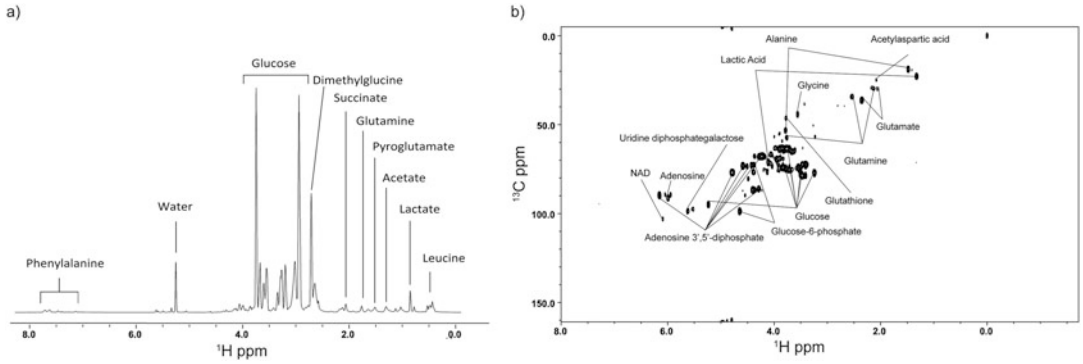

Fig. 6 NMR metabolomics spectral data. Examples of a typical (**a**) 1D ^{1}H NMR spectrum and a (**b**) 2D ^{1}H-^{13}C HSQC spectrum acquired from PD metabolomics samples

*3.9.2 2D ^{1}H-^{13}C-HSQC NMR (*See ***Note 18****)*

1. Follow **steps 1–7** from Subheading 3.9.1.
2. Using ICONNMR 5, the sample filename, solvent, pulse program, and temperature parameters are adjusted (*see* **Notes 15–17**).
3. The ICONNMR setup is similar to a 1D ^{1}H NMR data collection as shown in Fig. 5.
4. A standard 2D ^{1}H-^{13}C-HSQC experiment (Bruker *hsqcetgp-sisp2* pulse program) is used to determine the ^{1}H-^{13}C chemical shift correlations for all ^{13}C-labeled metabolites in the metabolomics sample (*see* **Note 19**).
5. Typical experimental parameters for a 2D ^{1}H-^{13}C-HSQC NMR spectrum obtained on a Bruker 700 MHz spectrometer with a cryoprobe correspond to 128 scans, 32 dummy scans, and a 1.0 s relaxation delay. The spectrum is collected with 2 K data points and a spectrum width of 4734 Hz in the direct dimension and 64 data points and a spectrum width of 18,864 Hz in the indirect dimension (*see* **Note 14**).
6. Implementation of fast NMR methods that includes nonuniform sampling significantly decreases data acquisition time and/or increases spectral resolution but may introduce artifacts (*see* **Note 20**).
7. Collect the 2D ^{1}H-^{13}C-HSQC NMR spectrum for each replicate and each group (*see* **Note 6**).
8. The data is processed initially with Topspin to verify spectral quality but exported for further analysis (*see* Fig. 6b).

3.10 Mass Spectrometry Data Collection

3.10.1 Direct-Injection (DI) Mass Spectrometry

1. Positive-ion direct infusion electrospray ionization mass spectrometry (DI-ESI-MS) data are collected on a Synapt G2 HDMS quadrupole time-of-flight (TOF) MS instrument equipped with an ESI source.
2. The mass spectrometry experiments are carried out at a flow rate of 10 μL/min for 1 min.
3. The mass spectra are acquired in positive ion and negative mode over a mass range of *m/z* 50–1200.
4. Mass spectra are acquired for 0.5 min using the following optimized source conditions: 2.5 kV for ESI capillary voltage, 60 V for sampling cone voltage, 4.0 V for extraction voltage, 80 °C for source temperature, 250 °C for desolvation temperature, 500 L/h for desolvation gas, and 15 μL/min flow rate of injection.
5. Collect the DI-ESI-MS mass spectral data for each replicate and each group (*see* **Note 6** and Fig. 7).

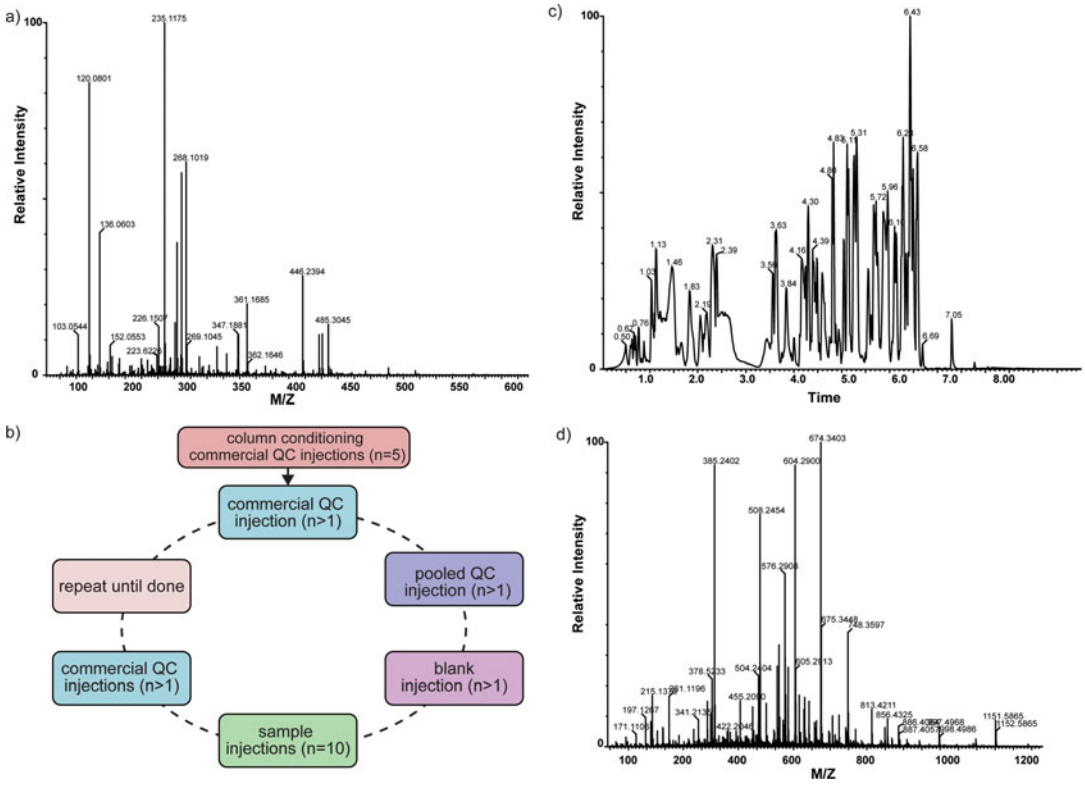

Fig. 7 MS metabolomics data. Examples of (**a**) typical DI-ESI-MS spectrum, (**b**) typical analysis sequence, (**c**) LC base peak chromatograms, and (**d**) MS spectrum acquired from metabolomics samples

3.10.2 Liquid Chromatography-Mass Spectrometry

1. Liquid chromatography-mass spectrometry (LC-MS) data are collected on a Waters ACQUITY M-class Xevo G2-XS QToF MS instrument equipped with an ESI source.
2. The LC-MS mass spectrum are acquired with the following system parameters:
   - LC system: Waters ACQUITY M-class.
   - Column: ACQUITY HSS T3 column 1.0 mm × 100 mm.
   - Mobile phase A: 0.1% formic acid in water.
   - Mobile phase B: 0.1% formic acid in acetonitrile.
   - Flow rate: 70 μL/min.
   - Run time: 10 min.
   - Injection volume: 2 μL.
   - MS system: Xevo G2-XS QToF.
   - Ionization mode: ESI + and –.
   - Capillary voltage: 2.8 kV.
   - Cone voltage: 30 V.
   - Source temp: 120 °C.
   - Desolvation temp: 500 °C.
   - Cone gas flow: 18 L/h.
   - Lock mass:
     - Positive mode: leucine-enkephalin, *m/z* 556.2771.
     - Negative mode: leucine-enkephalin, *m/z* 554.2615.
   - Acquisition mode: MSE.
   - Acquisition range: 50–1200 *m/z*.
   - Collision energy (LE): 6 eV.
   - Collision energy (HE): 20–40 eV.
3. The temperature for the LC column and autosampler is set to 40 °C and 4 °C, respectively.
4. Create a sample analysis sequence and inject the QC samples five times for column conditioning. After second QC injection, monitor peak area (<25% RSD), retention time (±0.05 min), and mass accuracy (±3 ppm) until the end of the fifth injection. If the QC samples pass the minimal system performance parameters, then acquire data. If not, do not collect data until the issue has been resolved and the QC samples pass the minimal system performance parameters.
5. Collect the LC-MS mass spectral data for each replicate and each group (*see* **Note 6** and Fig. 7).

3.11 NMR Data Processing (See Note 21)

All NMR data is processed and analyzed with our MVAPACK software [23], our PCA/PLS-DA utilities [24], NMRPipe [25], and NMRViewJ [26]. See example processing scripts at http://bionmr.unl.edu/wiki/Scripts.

3.11.1 1D ^{1}H NMR (See Fig. 8a)

1. A 1.0 Hz exponential apodization function is applied to the FID.
2. Fourier transform the FID.
3. The resulting NMR spectrum is automatically simultaneously phase corrected and normalized with the phase-scatter correction algorithm [45].
4. The NMR spectrum is referenced to the peak of TMSP-d_4 (0.0 ppm).
5. Noise and solvent regions are manually removed.

3.11.2 2D ^{1}H-^{13}C-HSQC NMR (See Fig. 8b)

1. A sine-bell apodization function is applied to the t_2 dimension.
2. The t_2 dimension is zero filled three times.
3. The t_2 dimension is Fourier transformed and manually phase corrected and the imaginary data deleted.
4. The matrix is transposed.
5. A sine-bell apodization function is applied to the t_1 dimension.
6. The t_1 dimension is zero filled three times.

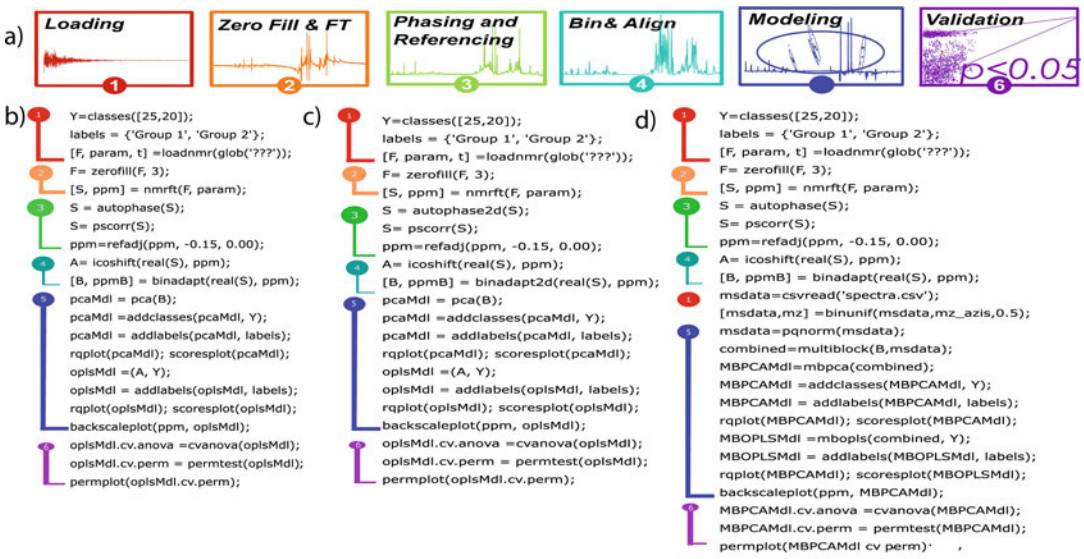

Fig. 8 MVAPACK processing scripts. (**a**) Schematic illustration of the major processing steps. Examples of MVAPACK processing script for (**b**) 1D ^{1}H NMR dataset, (**c**) 2D ^{1}H-^{13}C HSQC dataset, and (**d**) combined NMR and MS datasets. The numbered steps in the flow diagram correspond to the numbered lines in the processing scripts

7. The t_1 dimension is Fourier transformed and manually phase corrected.
8. The NMR spectrum is referenced in both dimensions to the peak of TMSP-d_4 (0.0 ppm).

3.12 Mass Spectrometry Data Processing: DI-ESI-MS (See Fig. 8c)

1. Mass spectral data processing is first performed using MassLynx V4.1.
2. A background subtraction is performed on all spectra using appropriate reference spectra, such as a free drug or toxin used to treat a cell culture. The background subtraction of each spectrum is performed in a class-dependent manner (i.e., only the MS reference spectrum of the drug/toxin used to treat the cell culture is used for background subtraction). Accordingly, mass spectral signals from the drug/toxin treatments are guaranteed to not influence subsequent analyses. An example of a typical MS spectrum from a metabolomics sample is shown in Fig. 7.
3. The background-subtracted mass spectra are then loaded into MVAPACK as a text file for binning and normalization.

3.13 Mass Spectrometry Data Processing: LC-MS (See Fig. 9)

All LC-MS data is processed and analyzed with Progenesis QI (version 2.0.). Please see the Progenesis QI user guide (http://storage.nonlinear.com/webfiles/progenesis/qi/v2.2/user-guide/Progenesis_QI_User_Guide_2_2.pdf) for detailed instructions.

3.13.1 Data Upload

1. Go to *File* and create a new experiment. Select a location to store the experiment file. Click *Next*.
2. Select the machine type (i.e., high-resolution mass spectrometer) and the polarity used to collect the mass spectrum (i.e., positive or negative). Click *Next*.
3. Select the expected adducts [e.g., $M+Na^+$ (+), $M+H+CH_3OH^+$ (+, −), $M+H+CH_3N^+$(+, −), $M+H_3O^+$ (+)] and click *Create experiment*.
4. Go to *Select your run data*, choose the MS *Format*, and click *Import*. Click *Next*.
5. Apply Lock mass calibration. Click *Next*.
6. Select *Import*.

3.13.2 Perform Automatic Processing

1. Click on *Start automatic processing*.
2. Select an alignment reference by choosing *Use the most suitable run from candidates that I select*. Click *Next*.
3. Select all QC runs. Click *Next*.
4. Select *Yes, automatically align my runs*. Click *Next*. Click *Next* again.
5. After processing is complete, click *Section Complete* to move forward to the *Review Alignment* stage.

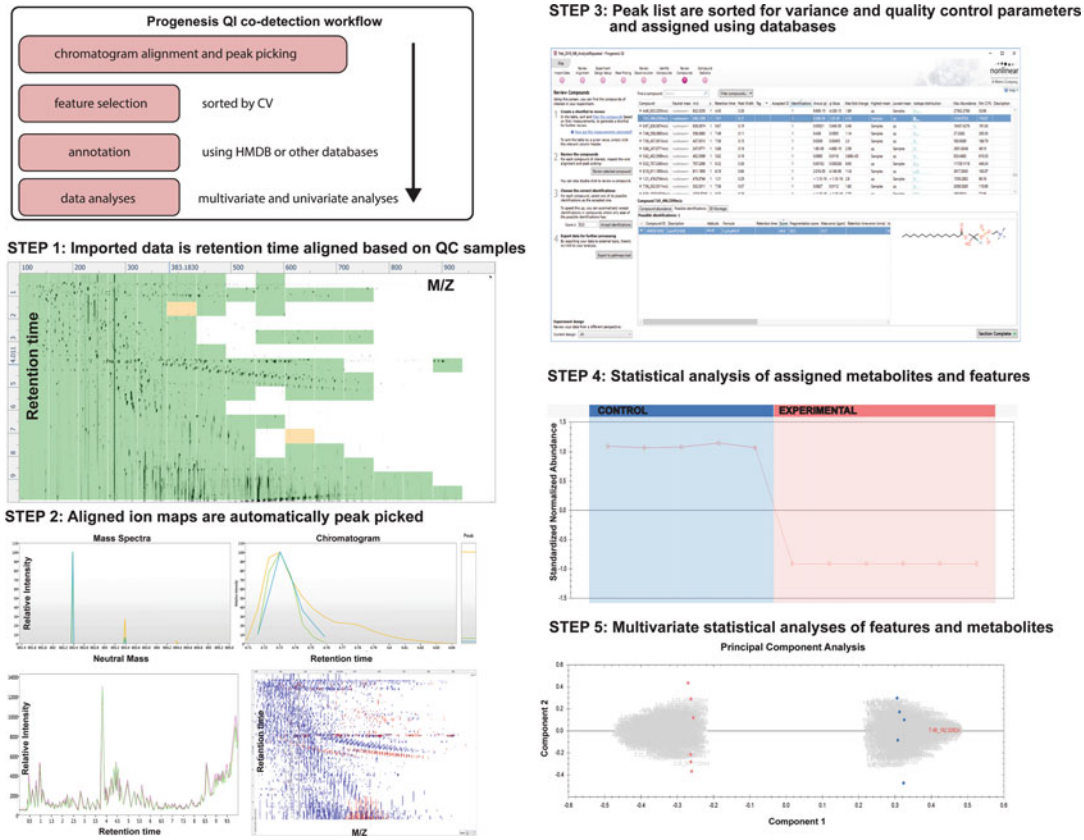

Fig. 9 LC-MS processing protocol. The small molecule discovery workflow using the Progenesis QI software is diagrammed. (*top left*) Summary of the major steps in the LC-MS processing protocol, which also describes each figure block in order starting from *middle-left* to *bottom-right*. Images are screenshots from the Progenesis QI software

3.13.3 Review Alignment

1. Interrogate the number of vectors and alignment scores.
2. Examine the distribution of green (good alignment), yellow (acceptable alignment), and red (needs review) alignments present in the ion intensity map.
3. As necessary, manually edit the alignments. Make sure that each ion is properly aligned across all replicates and to the reference mass spectrum. This is accomplished by interactively adjusting the alignment vector positions.
4. After processing is complete, click *Section Complete*.

3.13.4 Create Experiment Design

1. Choose the type of experiment and click *Create* (*see* **Note 22**).
2. Click *Add condition* and rename it according to the groups in the study (e.g., control, treated, etc.).
3. Drag and drop each replicate mass spectrum into each of the defined groups from **2**.

4. After processing is complete, click *Section Complete* to move forward to the *Peak Picking* stage.

3.13.5 Peak Picking

1. Click *Change parameters*.
2. Go to the *Peak picking limits* grid and define a minimum peak width to reject noise spikes. A typical minimum peak width is 0.05 min.
3. Click *Start peak picking*.
4. After the process is completed, go to *Review normalization*, and choose the normalization method. Normalize to all metabolites is the default. A preferred choice is to normalize to an internal standard (e.g., reserpine).
5. After processing is complete, click *Section Complete* to move forward to the *Deconvolution Review* stage.

3.13.6 Review Deconvolution (See **Note 23**)

1. Go to the *Deconvolution Review* grid.
2. On the left panel, choose *organize the compound features by adducts*.
3. Click over an ion metabolite to review its adducts (*see* **Note 24**).
4. To remove an adduct assigned to a metabolite, *right-click* on the peak in the adduct panel and click *Remove from compound*.
5. After the processing is complete, click *Section Complete* to move forward to the *Compound Statistics* stage.

3.14 NMR Data Preprocessing for Multivariate Modeling

In order to obtain an accurate and reliable multivariate statistical model, it is essential that the dataset is properly preprocessed to remove normal systematic variations resulting from biological variability, instrument instability, and inconsistency in sample handling and preparation. Key preprocessing steps include (1) alignment, (2) normalization, (3) binning, and (4) scaling, which are illustrated in Fig. 8. Examples of results from a variety of statistical models are shown in Fig. 10. All NMR datasets are processed with our MVAPACK software [23] and our PCA/PLS-DA utilities [24]. See example MVAPACK scripts at http://bionmr.unl.edu/wiki/Scripts.

3.14.1 1D ^{1}H NMR

1. Spectra may first be normalized based on either the total cell count or the total protein concentration using the BCA (bicinchoninic acid) protein estimation assay using parallel dishes treated similarly on the same day.
2. Spectra are normalized with the PSC algorithm [47].
3. Spectra are aligned and/or binned. For principal component analysis (PCA), use the following parameters:

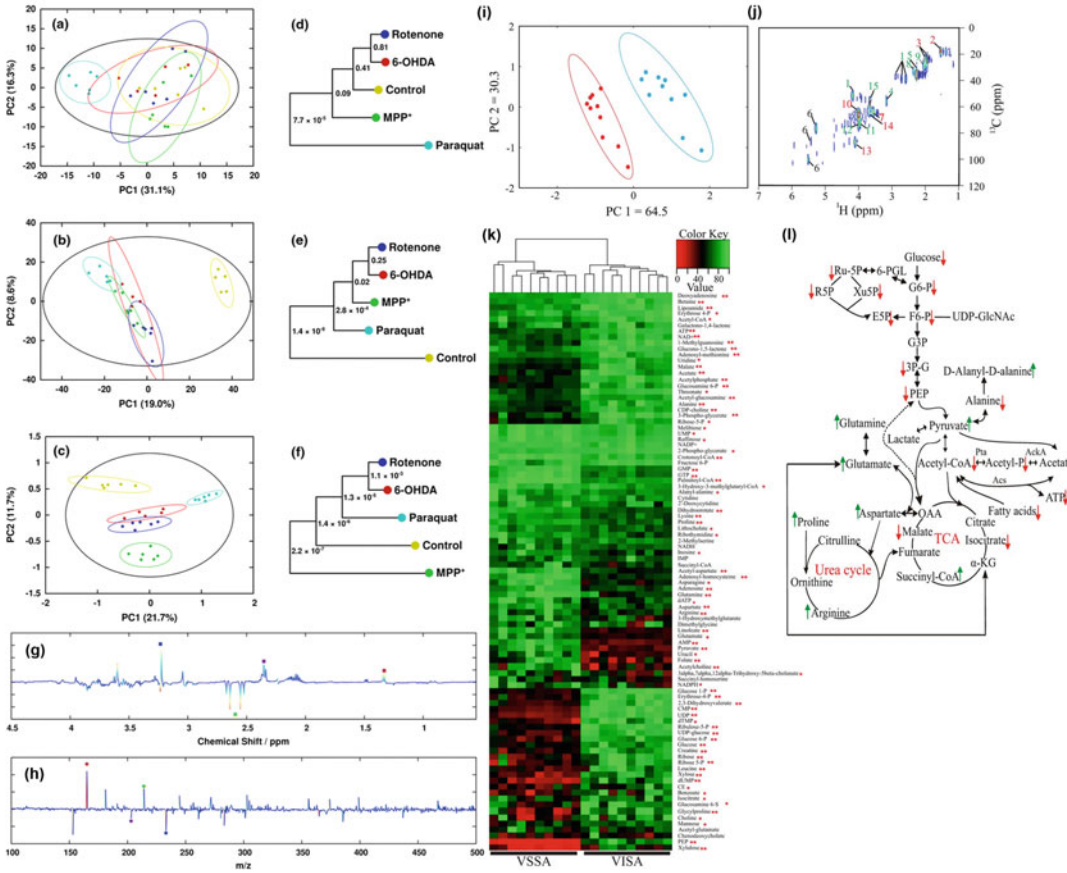

Fig. 10 Univariate and multivariate statistical models. Example of PCA scores plot and the associated metabolomics tree diagram for (**a**, **d**) 1D ^{1}H NMR dataset, (**b**, **e**) DI-ESI-MS dataset, and (**c**, **f**) combined 1D ^{1}H NMR DI-ESI-MS dataset. (**g**) NMR and (**h**) MS back-scaled loadings from an OPLS model generated from combined 1D ^{1}H NMR DI-ESI-MS dataset. Reproduced with permission from [46]. (**i**, **j**) PCA scores plot and OPLS back-scaled loadings generated from 2D ^{1}H-^{13}C HSQC NMR dataset. Reproduced with permission from [1]. (**k**) Example heat-map with hierarchal clustering summarizing specific metabolite changes per replicate and the relative clustering of each individual replicate. Reproduced with permission from [1]. (**l**) Example metabolic pathway summarizing the major metabolite changes between the two groups. Reproduced with permission from [1]

- The spectral data are globally aligned to the peak of TMSP-d_4 at 0.0 ppm.
- The spectral data are regionally aligned using the icoshift algorithm [48].
- The spectral data are binned using the adaptive intelligent binning algorithm [49].

For orthogonal projection to latent structures (OPLS), use the following parameters:

- The spectral data are globally aligned to the peak of TMSP-d_4 at 0.0 ppm.
- The spectral data are regionally aligned using the icoshift algorithm [48].
- The spectral data is not binned. Instead, the full-resolution spectral data is used to build the model.

4. Solvent peaks and noise regions are manually removed.
5. The dataset is scaled using Pareto scaling.
6. A PCA or OPLS model is generated from the data matrix.

3.14.2 2D ^{1}H-^{13}C-HSQC NMR

1. Spectra may first be normalized based on either the total cell count or the total protein concentration as explained above.
2. The spectral data is normalized using standard normal variate normalization.
3. The spectral data is binned using a generalized adaptive intelligent binning algorithm [49].
4. The data are Pareto scaled.
5. A PCA or OPLS model is generated from the data matrix.

3.15 Mass Spectrometry Data Preprocessing for Multivariate Modeling

LC-MS datasets need to be preprocessed in a similar manner to NMR spectra. The LC-MS datasets are processed with Progenesis QI (version 2.0.).

3.15.1 DI-ESI-MS

1. All mass spectra are linearly re-interpolated onto a common axis that spanned from *m/z* 50 to 1200 in 0.003 *m/z* steps, resulting in 383,334 variables prior to processing.
2. The mass range *m/z* 1100–1200 is removed prior to binning because of the low probability of observing a metabolite in this region.
3. The mass spectra are uniformly binned using a bin width of 0.5 *m/z*, resulting in a data matrix of 2095 variables.
4. The MS data matrix is normalized using probabilistic quotient (PQ) normalization.
5. The MS data matrix is then scaled to unit variance prior to modeling.
6. A PCA or OPLS model is generated from the data matrix.

3.15.2 LC-MS (See **Note 25** and Fig. 9)

1. The LC-MS datasets are processed with Progenesis QI (version 2.0.). Please see the Progenesis QI user guide (http://storage.nonlinear.com/webfiles/progenesis/qi/v2.2/user-guide/Progenesis_QI_User_Guide_2_2.pdf) for detailed instructions.
2. Right-click on the *Compounds table* and select *Quick Tags*.

3. Set the ANOVA cutoff value to 0.05.
4. Click *Create tag*.
5. All metabolites with an ANOVA p-value ≤ 0.05 will be marked with a red tag.
6. Repeat the process to create a tag for fold change (*see* **Note 29**). Right-click on the *Compounds table* and select *Quick Tags*.
7. Set the fold change cutoff value to 2.
8. All metabolites with a fold change greater than 2 will be marked with a green tag.
9. Create a filter to show only tagged metabolites. Click *Create* on *Filter* grid to open the filter dialog box.
10. Select the tags and then drag to the box *Show compounds that have all these tags*. Click *OK*.
11. Only the metabolites that match the criteria are showed and will be used for metabolite identification.
12. Go to the *Compound statistics* grid. The statistical analysis is available as a PCA scores plot. A statistically relevant dataset is indicated by replicate samples clustering together in the scores plot. Furthermore, the set of control and treated replicates form distinct clusters from each other.
13. Go to *File* and select *export all measurements*. A comma-separated value (csv) file will be created with a list of several values per metabolite: (1) metabolite identification, (2) *m/z* value, (3) charge, (4) retention time, (5) relative abundance, and (6) ANOVA value and other parameters.

3.16 Statistical Analysis (See Fig. 10)

Datasets are analyzed with our MVAPACK software [23], our PCA/PLS-DA utilities [24], and R [50]. See example MVAPACK and R scripts at http://bionmr.unl.edu/wiki/Scripts.

A major challenge in the analysis of metabolomics datasets, and a common source of error, is the incorrect application of statistics. This results from a number of prevailing misconceptions within the metabolomics community. For example, a multivariate model, especially supervised methods such as PLS or OPLS, needs to be properly validated. Validation can be accomplished with CV-ANOVA [51] and/or response permutation testing [52]. Conversely, the resulting R^2 and Q^2 values only provide a measure of the model fit to the original data and an internal measure of consistency between the original and cross-validation predicted data, respectively. R^2 and Q^2 values do not provide for model validation without a proper standard of comparison.

PCA, PLS, and OPLS are routinely used to model metabolomics data. Nevertheless, there are misconceptions regarding the proper application and interpretation of the resulting models, especially in regard to comparing PCA, PLS, and OPLS models. For

example, PCA finds the largest source of variance in the dataset irrespective of the intent of the study. So, an observed separation between treated and untreated groups in a PCA scores plot may have nothing to do with the treatment if some other larger variant is present in the dataset. Supervised methods, like PLS and OPLS, address this issue by aggressively forcing group separation based on the defined group membership. Hence, PLS and OPLS models almost always yield separated groups, as by design! As a result, PLS and OPLS models are easily over-fitted, especially for metabolomics datasets since the number of variables (e.g., metabolites) is typically larger than the number of replicates. Again, model validation is essential for PLS and OPLS.

Another serious misconception is the false belief that PLS/OPLS is a *better* method than PCA and may find group differences when PCA fails to expose group separation. Instead, PCA, PLS, and OPLS are simply different models that extract different information and achieve different goals. Thus, if PCA fails to identify group separation, it is unlikely that PLS/OPLS will yield a valid model [52]. Remember, PCA finds the largest source of variance. If PCA doesn't find any major variance, then there cannot be any smaller group-specific variance.

PLS and OPLS appear to provide similar models. In fact, a comparison of PLS and OPLS scores plots generated from the same dataset may suggest the only difference is a relative rotation of the group-defined ellipses. Nevertheless, this apparently subtle change highlights a critical difference. Simply, OPLS places group-independent variance (e.g., confounding factors such as differences in diet, age, race, etc.) orthogonal to group-dependent variance. Conversely, PLS entangles both group-independent and group-dependent variance. In this regard, a metabolite identified as a major contributor to an OPLS model is strictly the result of the defined group difference. For PLS, metabolite changes may be a result of the group difference, a confounding factor or a combination of both. Accordingly, a PLS-identified metabolite may be of little interest to the intent of the study. In this regard, we strongly recommend always using OPLS instead of PLS.

3.16.1 Univariate Analysis

1. Relative metabolite abundances are inferred from NMR and/or mass spectral peak heights and/or peak volumes.
2. Relative metabolite abundances are then normalized on a per spectrum basis. One common approach is to convert the absolute peak intensities (arbitrary units) to a Z-score:

$$Z = \frac{I_i - \bar{I}}{\sigma} \tag{1}$$

where $\bar{I}$ is the average peak intensity for the spectrum, I_i is the intensity of peak i, and σ is the standard deviation of peak

intensities. Peak intensities may also be normalized to the total number of cells, to the total protein concentration (*see* Subheading 3.14.1), to the average spectral noise, or to an internal standard (*see* **Note 26**). Relative metabolite abundances may also be converted to fold changes:

$$F = \frac{I_i}{I_o} \tag{2}$$

where I_i is the normalized peak intensity of metabolite i from a treated spectrum and I_o is the normalized peak intensity of metabolite i from the control or untreated spectrum.

3. A standard Student's t-test is commonly used to determine statistical significance only for a pairwise comparison of metabolite changes based on either fold changes or normalized peak intensities (*see* **Note 27**). A statistically significant difference is typically identified by a p-value <0.05.
4. A Student's t-test is insufficient for the multiple comparisons that are common to a metabolomics study [52, 53]. In order to identify the set of metabolites that exhibit a statistically significant change, a multiple hypothesis test correction method such as a Benjamini-Hochberg [54] or a Bonferroni [55] correction must be applied (*see* **Note 28**).
5. A heat-map with hierarchical clustering (*see* Fig. 10k) is commonly generated from the fold changes or normalized peak intensities using R (*see* example R script at bionmr.unl.edu/wiki/scripts). The heat-map may contain relative metabolite abundances for each individual replicate in the study or simply the replicate averages for each group (*see* **Note 29**).

3.16.2 Multivariate Analysis

1. Generate a PCA and or OPLS model from the data matrix.
2. Fractions of explained variation (R^2_X and R^2_Y) are computed during PCA or OPLS model training.
3. The PCA or OPLS model is internally cross-validated using sevenfold Monte Carlo cross-validation [56] to compute Q^2 values (*see* **Note 30**).
4. For an OPLS model, the Q^2 value is compared against a distribution of null model Q^2 values in 1000 rounds of response permutation testing [52]. Group membership is randomly reassigned to generate the set of null models. A p-value is calculated from a comparison of the true Q^2 value to the set of null model Q^2 values (*see* **Note 31**).
5. The model is further validated using CV-ANOVA significance testing, which is used to calculate another model p-value [51] (*see* **Note 31**).

6. Scores plots (*see* Fig. 10a, b, c, i), back-scaled loadings plots (*see* Fig. 10g, h, j), *S*-plots, and/or SUS-plots are often generated from OPLS models.
7. PCA/PLS-DA utilities [24] are used to define group membership by drawing an ellipse per group onto the scores plots (*see* Fig. 10a, b, c, i). Each ellipse corresponds to 95% confidence interval for a normal distribution. The PCA/PLS-DA utilities also generate a metabolomics tree diagram that identifies the statistical significance (*p*-value and/or bootstrap value) and the relative similarity of each group in the scores plot (*see* Fig. 10d, e, f). The *p*-value or bootstrap number from the pairwise comparison is labeled at each node in the tree.

3.17 Data Analysis: Metabolite Assignment from 1D ^{1}H NMR Data

All NMR data is analyzed with NMRPipe [25], NMRViewJ [26], and Chenomx. See example scripts at http://bionmr.unl.edu/wiki/Scripts.

1. The identification of metabolites in a 1D ^{1}H NMR spectrum is performed with software programs such as Chenomx. Chenomx matches the experimental 1D ^{1}H NMR spectrum to a database of 1D ^{1}H NMR spectra of known metabolites. Chenomx attempts to explain or describe the experimental NMR spectrum by combining or summing as many of the individual reference metabolite NMR spectra as needed. In addition to metabolite identification, Chenomx also provides an estimate of the metabolite concentration (*see* **Note 32**).
2. Upload the 1D ^{1}H NMR spectrum for processing. The NMR spectra can be batch processed or processed one at a time.
3. The 1D ^{1}H NMR spectrum is phased.
4. The 1D ^{1}H NMR spectrum is calibrated and referenced to TMSP-d_4, using the known concentration of TMSP-d_4.
5. The properly phased and calibrated 1D ^{1}H NMR spectrum is then sent to the Chenomx profiler where the spectrum is compared against the metabolite library.
6. Chenomx will overlay a 1D ^{1}H NMR reference spectrum for each metabolite identified in the experimental 1D ^{1}H NMR spectrum. The spectral overlay needs to be manually adjusted to optimize the alignment of the experimental 1D ^{1}H NMR spectrum with the reference spectrum. Figure 6 shows an example of a labeled 1D ^{1}H NMR spectrum.

3.18 Data Analysis: Metabolite Assignment from 2D ^{1}H-^{13}C-HSQC NMR Data

All NMR data is analyzed with NMRPipe [25], NMRViewJ [26], and Chenomx. See example scripts at http://bionmr.unl.edu/wiki/Scripts.

3.18.1 NMRPipe Processing to Obtain .ft2 and .nv Files

1. The data files from ICONNMR can be used directly by NMRPipe to process the 2D ^{1}H-^{13}C-HSQC spectra.
2. On a Linux workstation, open a terminal and go to the directory that contains the NMR data. Type *bruker* to start the NMRPipe software.
3. Read in the experimental parameters file by clicking *Read Parameters* and verify that all of the parameters have been correctly updated. Confirm that the mode of data collection has been set to *echo-antiecho* if the NMR spectrum was collected with the *hsqcetgpsisp2* pulse program.
4. Click *Update Script* to save an NMRPipe processing script *fid.com* file in the working directory.
5. Type. */fid.com* to start the NMRPipe processing script.
6. When the NMRPipe processing has finished, type *nmrDraw* to view the processed NMR spectrum. Please see the NMRPipe and nmrDraw tutorial (https://spin.niddk.nih.gov/NMRPipe/doc1/) for detailed instructions.
7. Phase the NMR spectrum in NMRPipe, and note the p0 and p1 values for both the ^{1}H and ^{13}C dimensions.
8. Edit the NMRPipe processing script *hsqcproc.com*, and replace the parameters associated with the NMRPipe phase correction command, *ps*, with the p0 and p1 values obtained from **step 7**.
9. Type. */hsqcproc.com* to start the NMRPipe processing script.
10. Repeat **steps 3–9** for each 2D ^{1}H-^{13}C-HSQC NMR spectrum in the dataset. This produces a set of *.ft2* files. One *.ft2* file is created for each 2D ^{1}H-^{13}C-HSQC NMR spectrum collected for each replicate from each group.
11. Copy all of the *.ft2* files into a new folder and use the NMRPipe script *addnmr.com* to generate NMRviewJ files from the *.ft2* files. A *.nv* file will be generated for each individual spectrum (*.ft2* file) with an numerically incremented root name of "*Final_*". In addition, the script will combine all of the NMR spectra together into a single file called *results.nv*. The script will also generate the text file, *rate.txt*, that lists all of the individual *.nv* files (*Final_*).

3.18.2 Peak Picking and Peak Integration of 2D ^{1}H-^{13}C-HSQC Spectra in NMRviewJ

1. Type *nmrviewj* to start NMRviewJ. Please refer to NMRViewJ documentation (http://docs.nmrfx.org/) for more details.
2. From the *Dataset* toolbar in the *main* window, use the *Open and Draw Datasets* function to select the *result.nv* file.
3. Right-click and select *attributes* to open the *attributes* window.
4. In the *attributes* window, select the *PeakPick* tab.
5. In the blank *Lists* field in the *attribute* window, type a filename (i.e., *lists*) for the new peak pick list. Click the *Pick* button. The

software will automatically peak pick the displayed spectrum and populate *lists* with the peak ID number, chemical shifts, and intensity.

6. Choose *Show Peak Table* from the *Peak* toolbar on the *main* window. A *peak table* window will open that lists the peak ID, peak intensity, and peak chemical shifts.
7. Manually edit the peak list and remove solvent peaks, noise peaks, or other spectral artifacts. Peaks are deleted from the peak table by using the delete function in the *PeakPick* tab in the *attributes* window along with the *spectrum display* window. In the *spectrum display* window, use the mouse to position the two cursors around any peak or spectral region to form a box. Then, click the *Delete* button under the *PeakPick* tab in the *attributes* window to remove the peak(s).
8. After the peak table has been completely edited, on the *peak table* window choose the *Edit* tab and select *Compress & Degap*. Answer *yes* to the pop-up question. This will finalize changes to the peak list and prevent any further edits.
9. On the *peak table* window choose the *Edit* tab and then select *Save Table*. A file browser window will open in order to choose a name and location to save the new peak list file. The saved peak pick file can be viewed and edited by Excel.
10. In order to obtain peak intensities across the entire set of NMR spectra in the dataset, click on the *Analysis* tab on the main window and select *Rate Analysis*. A setup window for the *Rate Analysis* will open.
11. In the *Rate Analysis* setup window:
    - Set the *Prefix for matrix numbers* field to *Final_*.
    - Set the *Peaklist* field to *lists* (defined in Subheading 3.18.2, **step 5**).
    - Make sure *Auto fit* is checked.
    - Use all other default settings.
    - Click *Load time file*.
    - In the file browser window, select *rate.txt* (created in Subheading 3.18.1, **step 11**).
    - Click *Measure All*. The software automatically populates the table in the *Rate Analysis* setup window with all of the peak intensities across the entire NMR dataset.
    - Click *Save Table*. In the file browser window, save the peak intensities table to a new filename (i.e., *intensities*).
12. The peak list (i.e., *list*) and the peak intensities (i.e., *intensities*) files are merged in Microsoft Excel using the common peak ID column. The *ppm1* (^{1}H ppm) and *ppm2* (^{13}C ppm) columns

are added to the peak intensities columns to generate a complete matrix of NMR peaks and intensities across the entire dataset.

13. The merged Excel file is saved to a new filename.

3.18.3 Metabolite Assignments from 2D ^{1}H-^{13}C-HSQC Peak Lists

1. The complete list of peaks obtained from the NMRviewJ analyses is searched using NMR metabolomics databases such as HMDB [31], BMRB [32], or other databases (*see* **Note 33**).
2. On the HMDB home page, choose the *Search* tab and select *2D NMR Search*.
3. From the *Spectra Library* pull-down menu, choose *13C HSQC*.
4. Cut and paste the 2D ^{1}H-^{13}C-HSQC peak lists into the *Peak List* field. One set of ^{1}H and ^{13}C chemical shifts, respectively, per line. Chemical shift values should only be separated by white space.
5. Set the ^{1}H chemical shift error tolerance to 0.05 ppm (*X-axis Peak Tolerance* ± field) and the ^{13}C chemical shift error tolerance to 0.10 ppm (*Y-axis Peak Tolerance* ± field).
6. Click the *Search* button. Depending on the size of the peak list, the software will return a ranked-order list of possible metabolites based on the number of chemical shift matches to reference spectrum.
7. Manually curate the list of potential metabolite assignments based on the number of chemical shift assignments, the quality of the spectral overlap (i.e., chemical shift match), number of other metabolites in the same metabolic pathway, and the biological system (i.e., is it a reasonable or possible metabolite for the organism),
8. Obtain additional NMR (e.g., HMBC, HSQC-TOCSY) and/or MS spectral data to confirm or refute the assignment.
9. An assigned 2D ^{1}H-^{13}C-HSQC spectrum is shown in Fig. 6b.

3.19 Data Analysis: Metabolite Assignments from LC-MS Data

The identification process is accomplished using the Progenesis QI (version 2.0.) software (*see* **Note 34** and Fig. 9). Please see the Progenesis QI user guide for detailed instructions (http://storage.nonlinear.com/webfiles/progenesis/qi/v2.2/user-guide/Progenesis_QI_User_Guide_2_2.pdf).

3.19.1 Identification of Compounds (See **Note 36**)

1. Make sure the filter created in Subheading 3.15.2, **step 9** is applied, and then proceed to *Identity Compounds* grid.
2. At the left panel, define the method to be used. In this case, select *Progenesis MetaScope*.
3. Choose the search parameter, in this case choose HMDB (*see* **Note 35**).

4. Click *Search for identifications*.
5. After a few minutes, a dialog box will open identifying the number of metabolites identified. Click *ok* to close.
6. All ions with possible identifications will be presented as a solid gray icon on the left side.

*3.19.2 Incorporation of Theoretical Fragmentation (See **Note 36**)*

1. On the left panel, select *ChemSpider* [30] as the identification method (*see* **Note 37**).
2. In the *Choose search parameter* field, choose *default* and then click *edit*.
3. Set the following parameters:
   - Select name as *theoretical fragmentation*.
   - Set *precursor tolerance* to 5 ppm.
   - Tick *Perform theoretical fragmentation* box.
   - Set the *Fragment tolerance* to 5 ppm.
4. Click *Save*.
5. Click *Search for identifications*.
6. After a few minutes, a dialog box will open displaying the number of metabolites identified. Click *ok* to close.

3.19.3 Accepting Compounds Assignment

1. Proceed to *Review Compounds* grid.
2. Go to the option *Choose the correct identification* and set a threshold of 45. The choice of a threshold is empirical and may need refinement based on the specific properties of the dataset. The higher the threshold setting, the more confident are the assignments, but the more restrictive analysis may result in a lower number of assignments.
3. Click *Accept identifications*. All identifications with a score of 45 or above will be accepted automatically.

*3.19.4 Review and Accept the Identifications Manually (See **Note 38**)*

1. Select a metabolite from the list.
2. Go to the *Possible identifications* grid.
3. In the bottom panel, select the desired identification threshold for the metabolite.

4 Notes

1. Isotopically labeled reagents commonly used for NMR are not radioactive and do not require special handling or safety precautions. However, gloves and eye protection are standard safety protocol for preparing all types of metabolomics samples.

2. Deuterated solvents, such as D_2O or DMSO, are very hygroscopic and require storage in a dry box and need to remain sealed until used.
3. The pH of a 100% D_2O sample using a standard pH probe may not report the correct pH. A standardly applied correction is pD = pH + 0.4. Conversely, a recent study by K. A. Rubinson [57] suggests the variance is not as significant, especially for a phosphate buffer, and a correction may not be required.
4. Complete cell survivability in each group is essential to a successful metabolomics study. This may be particularly challenging in a study that involves treating cells with a drug, toxin, or some other condition (including nutrient depletion or supplementation) that is expected to alter cell viability. Thus, the goal is to identify a dosage and time for the experimental paradigm that will stress the cells, prior to the induction of cell death. In this regard, the observed metabolomic changes will be a result of the cell's immediate response to the mechanism of action of the experimental condition, or the adaptation of the cell to the stress, and not a general cell death response. We typically identify the dosage by collecting a series of growth curves over a range of drug/toxin concentrations and compare them to a growth curve from untreated cell culture.
5. The resulting composition of the metabolome is easily perturbed by any difference in the protocol. Thus, it is essential that every sample is handled in exactly the same manner as reasonably as possible. Bias can be induced if cells are cultured in different incubators or shakers; if cells are handled by different personnel; if cells are treated with a different wash, buffer, or media (even if it is the exact same recipe as prepared by the same individual); if the time to process the cells differ, etc. In essence, any source of variance (regardless of how slight) may lead to a significant biologically irrelevant change in the metabolome. As a result, an important aspect of the protocol is to randomize the processing of each sample to minimize any bias induced by sample order. The order of sample processing should change at each step of the protocol. It is especially critical to randomly interleave replicate samples from each group.
6. Randomization of samples throughout the protocol is essential to avoid the introduction of bias. For example, if all of the control samples are processed together first and all of the treated samples are processed second, a difference between the controls and treated samples may be due to the processing order instead of the expected response to treatment. Consider another example consisting of a set of 20 samples numbered 1–20. If the samples are always processed in the order of the

sample number, then a time bias will be induced across the entire dataset. Sample 20 will always be processed after a maximal wait-time, and sample 1 will always be processed the quickest. Accordingly, biologically relevant differences in the metabolomes will accumulate between the samples due to the difference in processing time. Instead, if the order is constantly changed at each step, the processing time and any impact on the metabolome will be randomized, which in turn will minimize or eliminate any bias.

7. The number of replicates per group will have a significant impact on the quality of the study and the statistical validity of the outcomes. In general, it is best to maximize the number of replicates per group, within reason, with a typical target of ten replicates per group. A variety of experimental considerations may impact the number of replicates that are practical for a given study. For example, a large number of groups may require a reduction in the number of replicates per group. Another consideration is the impact of the number of replicates on the quality of the metabolomics samples. Sacrificing quality for a greater number of replicates will not likely lead to a successful outcome. Conversely, a limited number of replicates <4 per group will likely provide meaningless results.
8. Other studies have used a combination of isotopically labeled and non-labeled carbon sources. The conditions of optimal labeling should be standardized for every cell line/type used while also considering the composition/recipe of the culture media and the required carbon sources (glucose, pyruvate, or glutamine) for cell growth. A time course between 1 and 48 h should be performed to assess the rate of carbon consumption. Examples of media used for ^{13}C-carbon-labeled metabolomics are Dulbecco's Modified Eagle Medium (DMEM) (11966-025, 10938-025, 11960-044, and A14430-01) and RPMI (11879020) from GIBCO/Life Technologies.
9. Removal of proteins and other biomolecules by methanol or ethanol precipitation is preferred over mechanical filtration methods or the application of Carr-Purcell-Meiboom-Gill (CPMG) NMR T_2 filtering techniques. Filtering techniques may remove metabolites that bind to biomolecules leading to biologically irrelevant group differences [52].
10. Smaller diameter NMR tubes of 3 mm (160 μL) or 1.7 mm (35 μL) may be needed if the available metabolomics sample is limited. Filling of these smaller diameter NMR tubes may require a liquid handling robot, such as a Gilson 215 Liquid Handler. In addition, the NMR acquisition parameters will likely need to be adjusted to account for the lower sensitivity due to the lower number of nuclei in the samples.

11. Topshim requires that the sample contain either a D_2O or H_2O solvent. It is advisable to create a shim file with a parameter set that produces an optimal set of shims for your sample type. Read in a shim file using the Bruker command *rsh* and select the appropriate Topshim shim file. If you are doing this for the first time, complete the command *topshim*; if you are not satisfied with the shim performance, use command *topshim tuneb tunea* to obtain an improved set of shims. Write the shim set parameters with the Bruker command *wsh*, and save it to a new filename for future reference.
12. The 90-degree pulse length is commonly measured by incrementing the P1 pulse in the *zg* pulse program by 1 μs or smaller increments and by plotting the relative peak heights or intensities. A maximum peak height should be observed at the pulse length corresponding to the 90-degree pulse. Conversely, a minimum or null spectrum should be observed at the pulse length corresponding to the 360-degree pulse length. In practice, a more accurate measure of the 90-degree pulse is obtained by measuring the 360-degree pulse length and dividing by four to obtain the 90-degree pulse length. A typical 90-degree pulse length for a metabolomics sample ranges from approximately 8–13 μs or longer. Among other factors, the relative salt concentration of the metabolomics sample affects the 90-degree pulse, in which a higher salt concentration results in a longer 90-degree pulse. Other factors also contributed to the observed 90-degree pulse, so it is always necessary to experimentally determine the 90-degree pulse for each sample or set of samples.
13. Excitation sculpting parameters (*zgesgp*)—32768 data points (TD), SW = 12.02 ppm, O1P (transmitter offset) = 4.70 ppm, D1 = 1 s, NS (number of scans) = 128, DS (dummy/steady state scans) = 16, P1 = 9.5–13.5 μs, SPNAM (shaped pulse for water suppression) = SINC1.1000 at 26.39 dB or 0.00228 W.
14. The NMR data acquisition parameters need to be adjusted to compensate for differences in the field strength and sensitivity of the NMR spectrometer actually used for the data collection. Specifically, the number of scans, the number of data points, the sweep width (13.79 ppm, ^1H frequency range), and the frequency offset (centered on water peak at 4.70 ppm) need to be adjusted according to the type and configuration of the NMR spectrometer used for the study.
15. For high-throughput NMR data collection, please refer to the Bruker ICONNMR manual to explore various configuration options. For example, composite experiments allow for the collection of multiple 1D and 2D experiments for the same metabolomics sample. An experimental set consisting of a 1D

^{1}H, a 2D ^{1}H-^{13}C HSQC, and a 2D ^{1}H-^{13}C HMBC experiment may be subsequently collected for the same sample before moving to the next sample in queue.

16. It is imperative that NMR data be collected at the same temperature for a queue of metabolomics samples. ICONNMR assists this by allowing for a temperature delay when a large number of samples are in the SampleJet queue. For example, a 15–60 s delay may be inserted prior to data acquisition to allow each sample to equilibrate to the probe temperature. We recommend a 60-second delay for both pre- and post-sample insertion to prevent any temperature variation.
17. Parameters to check before you queue experiments in ICONNMR for 1D ^{1}H NMR are number of scans *ns*, number of dummy scans *ds*, 90-degree pulse *p1*, delay *d1*, sweep width *sw*, receiver gain *rg*, experiment temperature *te*, and automation setup *aunm*. We recommend using *au_zgonly* as the automation setup. This will collect all samples at the same receiver gain, which will avoid peak intensity variation across the dataset.
18. In addition to 2D ^{1}H-^{13}C-HSQC NMR experiments, NMR metabolomics studies may make use of HMBC, TOCSY, HSQC-TOCSY, 2D J-resolved spectra, or other experiments. Similarly, ^{15}N-, ^{31}P-, and other isotope-labeled metabolites may be detected in addition to ^{1}H- and ^{13}C-labeled metabolites. Accordingly, experimental parameters, data processing and preprocessing methods, and data analysis techniques all need to be adjusted to accommodate the specifics of each NMR experiment. Nevertheless, there is enough similarity that the detail discussion of the application of 2D ^{1}H-^{13}C-HSQC NMR experiments may provide a useful initial guide to the application of other NMR experiments to metabolomics.
19. 2D ^{1}H-^{13}C-HSQC parameters (*hsqcetgpsisp2*)—1024 data point in F2 and F1, nonuniform sampling at 25%, O1P = 4.7 ppm, O2P (offset for ^{13}C) = 75 ppm, NS = 64, DS = 16, d1 = 2, P1 = 10–13 μs depending on salinity, CPDPRG2 = garp (decoupling program), and PCPD2 = 55 μs at PLW12 = 4.09 W
20. Nonuniform sampling of 2D ^{1}H ^{13}C HSQC data can be performed on metabolomics samples. We have successfully acquired data at 20% sparsity using a burst augmented scheduler available from http://bionmr.unl.edu/dgs-gensched.php [33]. Download the sampling schedules as a text file for Topspin.
21. A minimalistic approach to the processing of NMR and mass spectrometry data is optimal for a metabolomics analysis utilizing multivariate statistics such as PCA and OPLS. The resulting

multivariate statistical model is dependent on the choice of processing and preprocessing protocols. In effect, a different statistical model is likely to be obtain based on the presence (or absence) of baseline correction and the type of baseline correction method chosen. Similarly, the type of weighting (apodization) function, the type of spectral alignment or referencing, the resulting phase correction or phase correction algorithm, the number of zero-fills or the application of linear prediction, or any other data manipulation method will affect the outcome of the statistical model. Accordingly, it is best to avoid any unnecessary data processing steps since it is difficult to ascertain if the data processing induced a biologically irrelevant bias to the data or actually improved the model.

22. Before proceeding to statistical analysis, it is necessary to create an experiment design. Progenesis QI supports *Between-subject design* and *Within-subject design*. *Between-subject design* separates samples according to the experimental condition (control vs treated) for the statistical comparison. *Within-subject design* is a repeated measures study design where the same subject (i.e., cell, animal, or human) is compared across the full range of experimental conditions (before treatment and after treatment, different time points, etc.).
23. The ions and adducts for a compound are automatically recombined by Progenesis QI, but it is advisable to review the deconvolution results. It is important to make sure the same pattern of adducts are assigned equally across all replicates and between all groups. Progenesis compares each detected ion with each of its co-eluting ions. If by chance their mass difference matches the difference between two adduct masses (i.e., from the previously chosen list), then it is probably an adducted form of the same compound. Progenesis groups the two ions as the same compound and automatically assigns the ions to the respective adduct. However, if an interesting compound is identified in the sample, it is important to review the deconvolution process to make sure all of the ions grouped together are actually adducts of the same compound.
24. Adducts assigned to a compound should have the same retention time as the compound. Thus, compare the chromatograms from the potential adduct with the compound to determine how well the chromatograms overlay. If a poor match is observed, then remove the adduct.
25. A primary goal of the LC-MS data analysis is to identify metabolites that exhibit significant concentration differences between groups. This is accomplished in the Progenesis software by creating tags to identify metabolites that exhibit a statistically significant (ANOVA [51] p-value <0.05) difference

in relative abundance between the groups. Progenesis relies on PCA for this analysis.

26. For NMR, relative peak intensities are averaged across all replicates per group and also for each NMR peak assigned to the metabolite. Most metabolites will have more than one peak in an NMR spectrum, and all NMR peaks should be incorporated into an average relative peak intensity. Please note that NMR peaks may need to be scaled by the number of attached hydrogens, since peak intensity is proportional to the number of nuclei.
27. Of course, there are a variety of options beyond the standard Student's t-test such as Mann-Whitney *U* test [58], Welch's *t*-test [59], Hotelling's t-squared statistic [60], and one-way analysis of variance [52], among others. The proper choice of a statistical test depends on a number of factors, which is well beyond the scope of this protocol review. For an introduction to the topic, please see *A Biologist's Guide To Statistical Thinking And Analysis* [61].
28. In effect, the uncertainty in each pairwise comparison (as determined by the Student's *t*-test) is compounded with the addition of each metabolite to a set. The actual *p*-value for a set of metabolites is defined as:

$$p = 1 - (1 - \alpha)^m \tag{3}$$

where m is the number of hypotheses (metabolites) and α is typically defined as 0.05. Accordingly, a set of ten metabolites becomes an insignificant $p = 0.401$ even though each individual metabolite is statistically significant based on a pairwise Student's *t*-test with a $p < 0.05$.

29. A heat-map displaying all of the replicates from each group is preferred to only a group average plot. Specifically, the hierarchical clustering of each replicate is indicative of the relative group separation and provides further confirmation of an observed group separation from a PCA, PLS, or OPLS scores plot.
30. A valid PCA, PLS, or OPLS model typically has R^2 values $>Q^2$ values and Q^2 values >0.4.
31. While *p*-values <0.05 are typically acceptable, more often than not, high-quality PLS/OPLS models from metabolomics datasets yield *p*-values $<<0.001$.
32. Chenomx maintains a series of 1D ^{1}H NMR databases for a variety of NMR field strengths and sample pH. Use the database that matches the experimental conditions of the dataset being analyzed.

33. Most NMR metabolomics databases function in a similar manner to HMDB [31]. Simply upload a peak list with a set of chemical shift tolerances and obtain a list of potential matches.
34. The identification process is also available in open-source software such as MZmine [28] or web-based tools such as MetaboAnalyst [29].
35. It is also possible to create or select your own search parameter. Click on *Edit* and select *Create New*. Select a database file in Structure Data Format (SDF) as input.
36. The possible compound assignments are based on an overall *score* determined by the *mass error*, *retention time error*, *isotope similarity*, *fragmentation score*, and, if available, *collision cross section*. The confidence of the identification may be increased by including *theoretical fragmentation* (*see* **Note 37**).
37. ChemSpider is a database comprised of 67 million compounds and, accordingly, is not restricted to known metabolites [30]. But, Progenesis can use the ChemSpider database for in silico prediction of fragmentation patterns. Progenesis cannot do this with HMDB [31].
38. Subheading 3.19.3 sets a global threshold setting for all metabolites. Sometimes this may be too restrictive for specific metabolites, where a lower global threshold setting may cause a large number of erroneous assignments. Subheading 3.19.4 describes a manual approach to adjust the threshold settings for individual metabolites to recover incorrectly missed assignments while avoiding a high false assignment rate.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 1660921. This work was supported in part by funding from the Redox Biology Center (P30 GM103335, NIGMS) and the Nebraska Center for Integrated Biomolecular Communication (P20 GM113126, NIGMS). The research was performed in facilities renovated with support from the National Institutes of Health (RR015468-01). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Gardner SG, Marshall DD, Daum RS, Powers R, Somerville GA (2018) Metabolic mitigation of *staphylococcus aureus* vancomycin intermediate-level susceptibility. Antimicrob Agents Chemother 62(1). https://doi.org/10.1128/AAC.01608-17
2. Anandhan A, Jacome MS, Lei S, Hernandez-Franco P, Pappa A, Panayiotidis MI, Powers R,

Franco R (2017) Metabolic dysfunction in Parkinson's disease: bioenergetics, redox homeostasis and central carbon metabolism. Brain Res Bull 133:12–30. https://doi.org/10.1016/j.brainresbull.2017.03.009

3. Camandola S, Mattson MP (2017) Brain metabolism in health, aging, and neurodegeneration. EMBO J 36(11):1474–1492. https://doi.org/10.15252/embj.201695810
4. Powers R, Lei S, Anandhan A, Marshall DD, Worley B, Cerny RL, Dodds ED, Huang Y, Panayiotidis MI, Pappa A, Franco R (2017) Metabolic investigations of the molecular mechanisms associated with parkinson's disease. Meta 7(2). https://doi.org/10.3390/metabo7020022
5. Gebregiworgis T, Powers R (2012) Application of NMR metabolomics to search for human disease biomarkers. Comb Chem High Throughput Screen 15(8):595–610
6. Halouska S, Zhang B, Gaupp R, Lei S, Snell E, Fenton RJ, Barletta RG, Somerville GA, Powers R (2013) Revisiting protocols for the NMR analysis of bacterial metabolomes. J Integr OMICS 3(2):120–137. https://doi.org/10.5584/jiomics.v3i2.139
7. Powers R (2014) The current state of drug discovery and a potential role for NMR metabolomics. J Med Chem 57(14):5860–5870. https://doi.org/10.1021/jm401803b
8. Botas A, Campbell HM, Han X, Maletic-Savatic M (2015) Metabolomics of neurodegenerative diseases. Int Rev Neurobiol 122:53–80. https://doi.org/10.1016/bs.irn.2015.05.006
9. Han W, Sapkota S, Camicioli R, Dixon RA, Li L (2017) Profiling novel metabolic biomarkers for Parkinson's disease using in-depth metabolomic analysis. Mov Disord 32 (12):1720–1728. https://doi.org/10.1002/mds.27173
10. Luan H, Liu LF, Tang Z, Zhang M, Chua KK, Song JX, Mok VC, Li M, Cai Z (2015) Comprehensive urinary metabolomic profiling and identification of potential noninvasive marker for idiopathic Parkinson's disease. Sci Rep 5:13888. https://doi.org/10.1038/srep13888
11. Roede JR, Uppal K, Park Y, Lee K, Tran V, Walker D, Strobel FH, Rhodes SL, Ritz B, Jones DP (2013) Serum metabolomics of slow vs. rapid motor progression Parkinson's disease: a pilot study. PLoS One 8(10):e77629. https://doi.org/10.1371/journal.pone.0077629
12. Poliquin PO, Chen J, Cloutier M, Trudeau LE, Jolicoeur M (2013) Metabolomics and in-silico analysis reveal critical energy deregulations in animal models of Parkinson's disease. PLoS One 8(7):e69146. https://doi.org/10.1371/journal.pone.0069146
13. Chen X, Xie C, Sun L, Ding J, Cai H (2015) Longitudinal metabolomics profiling of parkinson's disease-related alpha-synuclein A53T transgenic mice. PLoS One 10(8):e0136612. https://doi.org/10.1371/journal.pone.0136612
14. Lewitt PA, Li J, Lu M, Beach TG, Adler CH, Guo L, Arizona Parkinson's Disease C (2013) 3-hydroxykynurenine and other Parkinson's disease biomarkers discovered by metabolomic analysis. Mov Disord 28(12):1653–1660. https://doi.org/10.1002/mds.25555
15. Lei S, Zavala-Flores L, Garcia-Garcia A, Nandakumar R, Huang Y, Madayiputhiya N, Stanton RC, Dodds ED, Powers R, Franco R (2014) Alterations in energy/redox metabolism induced by mitochondrial and environmental toxins: a specific role for glucose-6-phosphate-dehydrogenase and the pentose phosphate pathway in paraquat toxicity. ACS Chem Biol 9(9):2032–2048. https://doi.org/10.1021/cb400894a
16. Havelund JF, Andersen AD, Binzer M, Blaabjerg M, Heegaard NHH, Stenager E, Faergeman NJ, Gramsbergen JB (2017) Changes in kynurenine pathway metabolism in Parkinson patients with L-DOPA-induced dyskinesia. J Neurochem 142(5):756–766. https://doi.org/10.1111/jnc.14104
17. Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, Banning F, Seissler J, Grallert H, Adamski J, Lechner A (2014) Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. PLoS One 9(2):e89728. https://doi.org/10.1371/journal.pone.0089728
18. Worley B, Powers R (2013) Multivariate analysis in metabolomics. Curr Metabolomics 1 (1):92–107. https://doi.org/10.2174/2213235X11301010092
19. Dettmer K, Aronov Pavel A, Hammock Bruce D (2006) Mass spectrometry-based metabolomics. Mass Spectrom Rev 26(1):51–78. https://doi.org/10.1002/mas.20108
20. Nicholson JK, Lindon JC, Holmes E (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29(11):1181–1189. https://doi.org/10.1080/004982599238047
21. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, Schrag AE, Lang AE (2017) Parkinson disease. Nat Rev Dis Primers

3:17013. https://doi.org/10.1038/nrdp.2017.13

22. Bras J, Guerreiro R, Hardy J (2015) SnapShot: genetics of parkinson's disease. Cell 160 (3):570–570 e571. https://doi.org/10.1016/j.cell.2015.01.019
23. Worley B, Powers R (2014) MVAPACK: a complete data handling package for NMR metabolomics. ACS Chem Biol 9 (5):1138–1144. https://doi.org/10.1021/cb4008937
24. Worley B, Halouska S, Powers R (2013) Utilities for quantifying separation in PCA/PLS-DA scores plots. Anal Biochem 433 (2):102–104. https://doi.org/10.1016/j.ab.2012.10.011
25. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–293
26. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. In: Downing AK (ed) Protein NMR techniques. Humana Press, Totowa, NJ, pp 313–352. https://doi.org/10.1385/1-59259-809-9:313
27. Blesa J, Przedborski S (2014) Parkinson's disease: animal models and dopaminergic cell vulnerability. Front Neuroanat 8:155. https://doi.org/10.3389/fnana.2014.00155
28. Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11(1):395. https://doi.org/10.1186/1471-2105-11-395
29. Xia J, Sinelnikov IV, Han B, Wishart DS (2015) MetaboAnalyst 3.0-making metabolomics more meaningful. Nucleic Acids Res 43(W1): W251–W257. https://doi.org/10.1093/nar/gkv380
30. Pence HE, Williams A (2010) ChemSpider: an online chemical information resource. J Chem Educ 87(11):1123–1124. https://doi.org/10.1021/ed100697w
31. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A (2013) HMDB 3.0-the human metabolome database in 2013. Nucleic Acids Res 41(D1):D801–D807. https://doi.org/10.1093/nar/gks1065
32. Markley JL, Anderson ME, Cui Q, Eghbalnia HR, Lewis IA, Hegeman AD, Li J, Schulte CF, Sussman MR, Westler WM, Ulrich EL, Zolnai Z (2007) New bioinformatics resources for metabolomics. World Scientific Publishing Co. Pte. Ltd, Singapore, pp 157–168. https://doi.org/10.1142/9789812772435_0016
33. Worley B, Powers R (2015) Deterministic multidimensional nonuniform gap sampling. J Magn Reson 261:19–26. https://doi.org/10.1016/j.jmr.2015.09.016
34. Klein C, Westenberger A (2012) Genetics of Parkinson's disease. Cold Spring Harb Perspect Med 2(1):a008888. https://doi.org/10.1101/cshperspect.a008888
35. Goldman SM (2014) Environmental toxins and Parkinson's disease. Annu Rev Pharmacol Toxicol 54:141–164. https://doi.org/10.1146/annurev-pharmtox-011613-135937
36. Cannon JR, Greenamyre JT (2013) Gene-environment interactions in Parkinson's disease: specific evidence in humans and mammalian models. Neurobiol Dis 57:38–46. https://doi.org/10.1016/j.nbd.2012.06.025
37. Franco R, Li S, Rodriguez-Rocha H, Burns M, Panayiotidis MI (2010) Molecular mechanisms of pesticide-induced neurotoxicity: relevance to Parkinson's disease. Chem Biol Interact 188(2):289–300. https://doi.org/10.1016/j.cbi.2010.06.003
38. Falkenburger BH, Saridaki T, Dinter E (2016) Cellular models for Parkinson's disease. J Neurochem 139(Suppl 1):121–130. https://doi.org/10.1111/jnc.13618
39. Creed RB, Goldberg MS (2018) New developments in genetic rat models of parkinson's disease. Mov Disord 33(5):717–729. https://doi.org/10.1002/mds.27296
40. Mosley RL, Hutter-Saunders JA, Stone DK, Gendelman HE (2012) Inflammation and adaptive immunity in Parkinson's disease. Cold Spring Harb Perspect Med 2(1): a009381. https://doi.org/10.1101/cshperspect.a009381
41. Ascherio A, Schwarzschild MA (2016) The epidemiology of Parkinson's disease: risk factors and prevention. Lancet Neurol 15 (12):1257–1272. https://doi.org/10.1016/S1474-4422(16)30230-7
42. Anandhan A, Lei S, Levytskyy R, Pappa A, Panayiotidis MI, Cerny RL, Khalimonchuk O, Powers R, Franco R (2017) Glucose metabolism and AMPK signaling regulate dopaminergic cell death induced by gene (alpha-Synuclein)-environment (paraquat) interactions. Mol Neurobiol 54(5):3825–3842. https://doi.org/10.1007/s12035-016-9906-2

43. Geraghty RJ, Capes-Davis A, Davis JM, Downward J, Freshney RI, Knezevic I, Lovell-Badge R, Masters JR, Meredith J, Stacey GN, Thraves P, Vias M, Cancer Research UK (2014) Guidelines for the use of cell lines in biomedical research. Br J Cancer 111 (6):1021–1046. https://doi.org/10.1038/bjc.2014.166
44. Westerhoff HV, Chen YD (1984) How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. Eur J Biochem 142 (2):425–430
45. Siegel MM (1981) The use of the modified simplex method for automatic phase correction in fourier-transform nuclear magnetic resonance spectroscopy. Anal Chim Acta 133 (1):103–108. https://doi.org/10.1016/S0003-2670(00)00089-1
46. Marshall DD, Lei S, Worley B, Huang Y, Garcia-Garcia A, Franco R, Dodds ED, Powers R (2015) Combining DI-ESI–MS and NMR datasets for metabolic profiling. Metabolomics 11(2):391–402. https://doi.org/10.1007/s11306-014-0704-4
47. Worley B, Powers R (2014) Simultaneous phase and scatter correction for NMR datasets. Chemom Intell Lab Syst 131:1–6. https://doi.org/10.1016/j.chemolab.2013.11.005
48. Savorani F, Tomasi G, Engelsen SB (2010) icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson 202 (2):190–202. https://doi.org/10.1016/j.jmr.2009.11.012
49. De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, Gillebert TC, Bekaert S, Martins JC, Van Criekinge W (2008) NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. Anal Chem 80(10):3783–3790. https://doi.org/10.1021/ac7025964
50. Development Core Team R (2011) R: A language and environment for statistical computing, vol 1. R Foundation for Statistical Computing, Vienna, Austria
51. Eriksson L, Trygg J, Wold S (2008) CV-ANOVA for significance testing of PLS and OPLS® models. J Chemom 22 (11–12):594–600. https://doi.org/10.1002/cem.1187
52. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savarin P (2015) PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. Mol BioSyst 11(1):13–19. https://doi.org/10.1039/c4mb00414k
53. Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, Bessant C, Connor S, Capuani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai R, Sjöström M, Trygg J, Wulfert F (2007) Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 3(3):231–241. https://doi.org/10.1007/s11306-007-0081-3
54. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol 57(1):289–300
55. Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. BMJ 310 (6973):170
56. Xu Q-S, Liang Y-Z (2001) Monte Carlo cross validation. Chemom Intell Lab Syst 56 (1):1–11. https://doi.org/10.1016/S0169-7439(00)00122-2
57. Rubinson KA (2017) Practical corrections for p (H,D) measurements in mixed H2O/D2O biological buffers. Anal Methods 9 (18):2744–2750. https://doi.org/10.1039/C7AY00669A
58. Ruxton GD (2006) The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. Behav Ecol 17 (4):688–690. https://doi.org/10.1093/beheco/ark016
59. Fay MP, Proschan MA (2010) Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat Sur 4:1–39. https://doi.org/10.1214/09-SS051
60. Hotelling H (1931) The economics of exhaustible resources. J Polit Econ 39(2):137–175
61. Fay DS, Gerow K (2013) A biologist's guide to statistical thinking and analysis. WormBook:1–54. https://doi.org/10.1895/wormbook.1.159.1