HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2019 July 15.

Published in final edited form as:

Nat Methods. 2018 September ; 15(9): 681–684. doi:10.1038/s41592-018-0110-3.

METLIN–MRM: a cloud library and public resource for targeted analysis of small molecules

Xavier Domingo-Almenara[1,§], **J. Rafael Montenegro-Burke**[1,§], **Julijana Ivanisevic**[2], **Aurelien Thomas**[3,4], **Jonathan Sidibe**[3,4], **Tony Teav**[2], **Carlos Guijas**[1], **Aries E. Aisporna**[1], **Duane Rinehart**[1], **Linh Hoang**[1], **Anders Nordstrom**[5], **Maria Gomez-Romero**[6], **Luke Whiley**[6], **Mathew R. Lewis**[6], **Jeremy K. Nicholson**[6], **H. Paul Benton**[1,*], and **Gary Siuzdak**[1,7,*]

[1]Scripps Center for Metabolomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States. [2]Metabolomics Unit, Faculty of Biology and Medicine, University of Lausanne, Rue du Bugnon 19, 1005 Lausanne, Switzerland. [3]Unit of Toxicology, CURML, Lausanne University Hospital, Geneva University Hospitals, Chemin de la Vulliette 4, 1000 Lausanne, Switzerland. [4]Faculty of Biology and Medicine, University of Lausanne, Chemin de la Vulliette 4, 1000 Lausanne, Switzerland [5]Department of Molecular Biology, Umea University, Umea SE-90187, Sweden. [6]Biomolecular Medicine, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, United Kingdom. [7]Department of Molecular and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States.

Abstract

Small molecule quantitative tandem mass spectrometry analysis[1] is now widely used in life sciences and medicine. Quantification is usually accomplished by prior fragmentation of standard materials and the use of commercial software to quantitate the resulting peaks[2,3]. Despite its widespread application, there is a paucity of public libraries to expedite assay development. Here, we introduce a new library, METLIN-MRM, comprised of more than 15,500 optimized transitions for multiple reaction monitoring of a wide variety of low molecular weight compounds. METLIN-MRM includes (i) transitions optimized following the established protocol with standard materials and (ii) transitions computationally optimized for selectivity. This computational optimization was achieved by the analysis of a large collection of tandem mass spectrometry spectra, where an algorithm selected the most unique transitions for a given compound in comparison with other compounds with a mass within the error of the mass spectrometer. METLIN-MRM streamlines quantitative analyses with minimal resources and development time and also serves as a public

*To whom correspondence should be addressed.

§These authors contributed equally.

AUTHOR CONTRIBUTIONS. G.S. led the project. X.D-A. designed and programmed the algorithms for XCMS-MRM and the statistical ranking of METLIN-MRM transitions, with critical contributions from J.R.M-B., H.P.B. J.R.M-B. designed the analytical assays for the validation of XCMS-MRM and METLIN MRM. J.I., A.T. acquired the experimental transitions in METLIN-MRM and J.S., T.T conducted the analytical assays. J.R.M-B., C.G., L.H., J.S., A.N., M. G-R., L.W. analyzed the different samples in their respective laboratories. H.P.B. coordinated the online implementation of XCMS-MRM and METLIN-MRM, assisted by A.E.A., D.R., who programmed the online interface. X.D-A., J.R.M-B. wrote the manuscript with contributions from C.G., J.I., A.T, G. S. M. R. L., J. K. N. provided advice and revisions of the article. All authors approved the final version of the article.

DATA AVAILABILITY. XCMS-MRM and METLIN-MRM are publicly available at http://xcmsonline-mrm.scripps.edu and http://metlin.scripps.edu, respectively. Materials data are available from the corresponding authors upon request.

repository, allowing the community to upload, share and cite experimental transitions through accession numbers. Additionally, this library has been integrated with XCMS-MRM, a cloud-based data analysis platform that allows for data analysis and sharing across different platforms and laboratories. This platform is publicly accessible at http://metlin.scripps.edu/ and http://xcmsonline-mrm.scripps.edu

The quantitative analysis of small molecules using tandem mass spectrometry (MS) represents a broadly used scientific resource, routinely applied for purposes as diverse as food safety, sports medicine, clinical diagnostics, pharmacology, drug discovery, toxicology, forensics, environmental analyses and microbiology[4-15]. In tandem MS quantitative analysis, a chromatographic system typically coupled to a triple quadrupole (QqQ) mass spectrometer is configured to monitor a particular set of precursor–product ion transitions/reactions for molecules of interest[3]. The first quadrupole is set to select specific precursor ions, which are then fragmented in the second quadrupole acting as a collision cell. The resulting fragment ions are filtered by the third quadrupole prior to detection. This approach, known either as selected or multiple reaction monitoring (SRM or MRM)[16] is currently the gold standard for small molecule quantitative analysis due to its high sensitivity and specificity. However, both precursor and product mass-to-charge ratios have to be optimized for each target molecule with pure standard materials. In proteomics, for instance, public libraries such as SRMAtlas[3] or iMPAQT[17] have been developed to address this challenge and streamline targeted peptide quantification throughput. These libraries provide a compendium of selected reaction monitoring assays for targeting the human proteome. However, no public equivalent repository of such magnitude is currently available for the analysis of small molecules or metabolites. Individual optimization through pure materials by each laboratory is therefore necessary to determine the optimal transitions. Yet, these optimized transitions by each laboratory configure a rich source of information that is habitually not shared or it is relegated to spreadsheets as supplementary material of scientific papers.

Once the transitions have been optimized and the tandem MS experiments have been conducted, computational tools are used to convert raw MS signals into relative or absolute (in the case of spiked stable-isotope labeled standards being used) concentration values[16]. Computational processing is necessary to integrate each molecular fragment signal and retrieve accurate concentrations. This processing, however, heavily relies on commercial and vendor-specific software as relatively few tools have been developed for such purposes (Skyline[18], MRMAnalyzer[19], MRMPROBS[20]).

In this paper, we introduce METLIN-MRM and XCMS-MRM, as cloud-based tandem MS library and data analysis platform, respectively, which streamline experimental design, data processing, analysis and data sharing (Fig 1).

The METLIN-MRM library is a compendium of small molecule transitions for multiple reaction monitoring tandem MS, compiled into a freely and publicly web-available library, designed to allow users to perform cost-effective quantitative analyses with minimal method development. Currently, it contains transitions for more than 15,500 unique small molecules (Fig. 1). In this resource, three different types of transitions are available: (i) traditional experimentally optimized (EO) transitions, (ii) computationally optimized (CO)

experimental transitions and (iii) public repository (PR) transitions. EO transitions were acquired for cover more than 1000 molecules in both positive and negative mode by analysis of pure standards using two different liquid chromatography – triple quadrupole systems across our laboratories (Agilent iFunnel 6495 QqQ and Sciex QTRAP 5500, see Methods for details). These small molecules transitions allow targeting a broad metabolite spectrum and in their optimization, the collision energy was tuned to retrieve those fragments showing the highest intensity.

In addition to experimentally acquired data, transitions for over 14,000 and 4,700 molecules in positive and negative mode, respectively, were computationally optimized using the METLIN spectral library[21] (acquired at different collision energies on a qToF instrument) by ranking empirical MS/MS fragments according to their selectivity (uniqueness of a product-fragment for a given molecule). The developed ranking algorithm compared MS/MS spectra of compounds with precursors within a ±0.7 Da window and fragments showing the best selectivity, without compromising intensity, were retained as transitions (Fig. 2a). Up to three transitions were determined for each molecule since the use of a single transition can lead to potential misidentification of molecules[22]. This strategy allowed a high-throughput transition optimization where instead of optimizing transitions based mainly on compound-dependent properties such as the signal-to-noise ratio, transitions were optimized by prioritizing those transitions that are less likely to be masked by shared transitions from interfering molecules (Fig. 2a,b). In addition, METLIN-MRM allows users to evaluate these transitions via thump-up/down buttons, enabling the cooperative validation by the community. It also is worth highlighting our commitment to the continuous expansion of the library by the inclusion of more transitions, as well as maintaining its free availability to all the registered users.

Finally, METLIN-MRM serves as a public repository, where the community can upload their own transitions. Submitted lists of transitions are assigned with a unique accession number that can be used as a reference for publications. This ultimately facilitates the deposition of transitions used in experiments and scientific literature into a standardized and searchable database, increasing the traceability and reproducibility of experiments and the reuse/sharing of optimized transitions calculated by other laboratories. We launched this feature with a total of 3,300 transitions for over 1,500 small molecules from 8 peer-reviewed papers. These PR transitions will be displ}ayed along their corresponding original source (DOl) and will be downloadable.

The proposed cloud-based workflow integrating METLIN-MRM with XCMS-MRM is illustrated in Fig. 1. First, the transitions (precursor to fragment ions) to be monitored for the different molecules are searched against the METLIN-MRM library. Either, experimentally optimized, computationally optimized or transitions from the public repository can be selected (Extended Data Fig. 5). Upon transition selection, a chromatography system coupled to a mass spectrometer including: triple quadrupole (QqQ), quadrupole time of flight (q-ToF) using multiple reaction monitoring, or quadrupole-Orbitrap configured using parallel reaction monitoring[23], are used to monitor these transitions corresponding to each targeted molecule. Following data acquisition, raw data files – in any vendor format – are uploaded into the XCMS-MRM platform for quantitative analysis data processing. This

processing is accomplished through signal processing techniques to automatically detect, and integrate and align the peaks across samples, minimizing false peak integration and reducing manual labor (Extended Data Fig. 1). This yields relative molecule concentration levels, which can be converted to absolute values through the use of stable-isotope dilution assay, calibration curves by internal standard addition as well as external calibration. Additional peak integration refinement can be manually conducted for those transitions that might be affected by inherent tandem MS artifacts such as strong co-elution with interfering molecules or matrix effects. Finally, quantitative results are displayed including quality control indicators (capable of assessing accuracy and specificity, limits of detection and quantification and linear dynamic ranges) and biologically interpretable results and graphs (such as p-values or fold changes) to assess the statistical significance of the metabolite concentration changes among phenotypes. Moreover, in contrast to a vendor or offline software, the cloud-computing implementation allows data to be easily shared with internal and/or external collaborators. This enables all researchers across all fields to interact with data collectors (MS specialists) and play an active role in the experiment and its data analysis, facilitating inter-laboratory data sharing and collaboration, as for instance required in multi-centric clinical trials (Extended Data Fig. 2).

In order to validate the CO transitions (ranked) available in METLIN-MRM, we compared them against the EO (unranked). The unranked transitions were provided by two of our laboratories in addition to transitions from scientific literature reported by four independent laboratories. For a total of 641 molecules for which transitions were available in both groups, it is observed that for 90% of the molecules, at least one ranked and unranked transition matched, and in 62% of the cases, both the fragment and collision energy matched (Fig 2c). Moreover, we observed that for 60% of the transitions the collision energy differences between the ranked (CO) and unranked (EO) transitions was below 5 V (Fig 2d) despite the limited discrete values available for the ranked transitions (10, 20 and 40 V). Furthermore, in order to determine if the CE error for the ranked transitions compared to unranked is within the error range observed among unranked transitions, we determined, for each molecule, the maximum and minimum collision energy for the same experimentally optimized transitions among laboratories (those sharing the same fragment *m/z* value). Fig. 2e shows a two-dimensional density plot of the maximum and minimum collision energies found across laboratories. It can be appreciated that at low CE (10 V), a 5 V error range might be expected, whereas at higher CE (20 – 40 V), no significant error is observed. Interestingly, the general trend is that all the values determined experimentally following the standard protocol are centered among 10, 15, 20 and 40 V, agreeing with the discrete CE values provided by the ranked transitions from METLIN. To further evaluate the quality of the CO (ranked) transitions, we focused on evaluating their selectivity. For this, we calculated the number of putative interfering molecules (PIMs) that share the same transition based on a search in METLIN, and selected the number of PIMs of the most selective transition per molecule. We applied this procedure for both the ranked and unranked transitions separately. We then calculated the minimum PIM value per transition type (an indicator of how selective the most selective transition is) for and we plotted the ratio per molecule, calculated as the minimum PIM number of the unranked transitions divided by the minimum number of PIM of the ranked transitions, with values >1 and <1 indicating higher

selectivity for the unranked and ranked transitions, respectively (Fig 2f). A total of 61% of molecules had the same minimum PIM value, from the remaining, 35% had a selectivity fold-change greater than one, *i.e.,* ranked transitions are more selective than unranked transitions. Interestingly, previous studies reported a rate of 30% of misidentification when a single transition per compound with retention time was used in combination[22]. Only in 4% of the remaining cases, the unranked transitions presented better selectivity. After inspection of these fragments in METLIN, these were low-intensity fragments or they fell outside the collision energy range of METLIN (40 eV or above). Overall, these results confirmed the statistical ranking as a valid complementary optimization strategy. As an example, molecules such as leucine and isoleucine showed different transitions after the statistical ranking compared to unranked transitions. This allowed the necessary fragment selectivity for the correct identification and quantification of the aforementioned molecules in a reverse phase multiple reaction monitoring configuration (see Extended Figure 2), which otherwise would have required additional assay development. Despite the increased capability of selective transitions to distinguish between similar molecules, in-source fragments from the same structure might be inaccurately attributed to the target molecule[24]. This is an important limitation that emphasizes the continuing need for chromatography and retention time libraries to completely circumvent the need for standard materials. Finally, it is worth noting that there is a compromise between quantitative response and selectivity, *e.g.,* selective transitions do not always have an optimal intensity response, and vice versa. This is why the use of a qualitative in addition to a quantitative transition is encouraged to increase both quantitative and qualitative method performance.

The experimental validation of the computationally optimized transitions was performed by analyzing 98 metabolites (74 had both CO and EO transitions and 24 had only CO transitions) at 6 different concentrations (unspiked and 5 spiked samples) in a murine cell line (RAW264.7) using HILIC chromatography (HPLC and UPLC) in both positive-and negative-ion modes using Waters TQ-XS and Agilent 6495 QqQ instrumentation, respectively. R^2 values between real concentration and peak area were determined. From the 74 compounds, 62 of them showed R^2 values above 0.8 for both CO and EO transitions, demonstrating a linear response and validating both types of transitions (Extended Data Table 3 and 4). Those compounds showing R^2 values below 0.8 showed however a high linear relationship between CO and EO transition peak areas, which indicates that CO transitions performance is comparable to EO transitions. Furthermore, 24 other metabolites where analyzed with the CO transitions (EO transitions were not available) with 20 of them showing R^2 values above 0.8, increasing the number to a total of 82 validated metabolite transitions.

To demonstrate the capabilities of XCMS-MRM as a cooperative resource, six samples from the same pool of human serum sample spiked with standards (standard addition calibration method) were analyzed and processed by XCMS-MRM in four different laboratories worldwide. The independent quantitative analysis of those molecules in all four laboratories with different vendor equipment lead to similar quantitative results under expected instrumental and analytical error (a median coefficient of variation between laboratories of 15%, see Extended Data Table 1). All the laboratories were able to easily share their data and results through the XCMS-MRM platform without using third-party transmission

systems or particular reports, making their results universally accessible to collaborators and, optionally, to the community. Furthermore, XCMS-MRM performance for high-throughput data processing was assayed in a drug quantification study involving 354 human blood samples (see Methods for details), where a supervised quantification of two drugs and two of their metabolites (6 compounds in total) in an accredited toxicology laboratory was compared against the quantification by XCMS-MRM. All drugs and metabolites were quantified in the majority of cases with relative errors below 10% (see Extended Figure 4), demonstrating the capability of XCMS-MRM to perform an accurate quantification of small molecules.

Together, METLIN-MRM and XCMS-MRM facilitated the data acquisition, results sharing and reproducibility of the results in four independent laboratories with different vendor instruments. The platform also decreased the cost and resources involving additional experimental transition optimization and subsequent data processing. Additionally, METLIN-MRM constitutes the largest publicly available ensemble of tandem MS transitions for small molecules, allowing the identification and quantification of a large number of endogenous as well as exogenous molecules. We anticipate that these resources will enable one of the most widely used technologies, MRM, to meet the demands of this ever-growing scientific community.

ONLINE METHODS

Transition optimization via triple quadrupole tandem mass spectrometry.

Most of the material standards used for transition optimization were purchased from Sigma-Aldrich (St. Louis, MO). Transitions for metabolites of interest were acquired in two laboratories using two triple quadrupole (QqQ) mass spectrometers. In the first laboratory, flow injection analysis (FIA) using an UH-PLC system (1290 series, Agilent Technologies) coupled to Triple quadrupole 6495 (QqQ, Agilent Technologies) mass spectrometer was used. The MS was operated in Single reaction monitoring mode (SIM) using the Optimizer in order to automatically optimize the collision energies and product ions for each metabolite of interest. Cycle time was 500 ms. ESI source conditions were set as following: dry gas temperature 230 °C, drying gas 14 L/min, nebulizer 40 psi, sheath gas temperature 400 °C, sheath gas flow 12 L/min, nozzle voltage at 500 V, capillary voltage 3500 V in ESI positive and 2000 V in negative mode. The analyses were performed in ESI positive mode using an isocratic gradient, 50% 20 mM ammonium acetate, 0.1% formic acid in water and 50% 0.1% formic acid in acetonitrile for 1 min. In ESI negative mode the isocratic gradient, 50% 10 mM ammonium acetate in water and 50% acetonitrile was applied, also during 1 minute. The flow rate was 400 μL/min and the injection volume was 2 μL, for all analyzed standards. Standard compound mixtures were reconstituted in methanol: water (5:95, v/v), transferred to HPLC vials and stored at −80 °C prior to flow injection analysis.

The second laboratory used a QTRAP 5500 (Sciex) coupled to a LC Dionex Ultimate 3000 (Dionex, Thermo Scientific). Analyses were performed in positive and negative electrospray ionization using a TurboV ion source. The chromatographic separation was performed on a column Kinetex C18 (100 × 2.1 mm, 2.6 μm). The mobile phases were composed A: 0.1% formic acid in H_2O and B: 0.1% formic acid in Acetonitrile (ACN) for the positive mode. In

the negative mode, the mobile phases were constituted by A: 0.5 mM ammonium fluoride in water and B: ammonium fluoride 0.5 mM in ACN. The run started at 2% B for 1.5 min, it was ramped to 98% B during 5 minutes, held at 98% B for 2 additional minutes, ramped down to 2% B in 0.5 min and re-equilibrated for the rest of the run.

Transition optimization via computational ranking.

The METLIN-MRM statistical optimization algorithm was designed for high-through put transition optimization. The algorithm aims at finding the most selective transitions for each molecule by comparing experimental MS/MS spectra from the METLIN library among molecules with a precursor within a ±0.7 Da window. For each given target molecule in the METLIN library, a two-step strategy to optimize their transitions was applied: (i) determining the maximum collision energy to be considered and (ii), determining up to three transitions per molecule. These three transitions are determined using three slightly different methods and aim at mimicking the routine MRM transition setup composed of a quantitative and two qualitative transitions.

In the first step (i), we aim to determine what is the maximum collision energy (CE) that we can apply to each compound without compromising sensitivity. In general, higher collision energies applied to a compound result in lower fragment intensities, especially when the molecule is in low concentration (a common situation in real samples). It is then of special importance to limit the CE to be applied to each molecule. In that sense, to determine this maximum CE, the algorithm compares the intensity of the precursor ion among different collision energies (10, 20 and 40 V), and it retains those collision energies were the the precursor signal is still detected. This is an indirect (non-experimental) way to measure the rate of dissociation of the precursor, and thus an optimal maximum CE. Once the maximum CE for each compound is determined, the algorithm proceeds to step two (ii), which consists in calculating the selective fragments. In this step, for a given target molecule, METLIN is searched for all molecules with a precursor within a ±0.7 Da window (typical quadrupole mass resolution). We will refer to these molecules as putative interfering molecules (PIM) throughout the text. Next, for each fragment and CE of the target molecule MS/MS spectrum, an intensity density value was determined, calculated as the sum of the intensities of the same fragment in the PIM's MS/MS spectra (Extended Data Figure 4a). This value acts as an indicator of the selectivity (uniqueness) of a certain fragment for a given molecule. A low-density value indicates that this fragment has a low or null intensity in the rest of interfering molecules, and thus has a high degree of selectivity. The fragment with the highest degree of selectivity is retained as a qualitative fragment. Only fragments above 20% of relative intensity (relative to the highest peak) are considered. To determine the second and third transitions, the algorithm determines the maximum interfering spectra, composed by all fragments across all PIM spectra, retaining the most intense fragment when this is present in two or more spectra. Then, the algorithm selects the two remaining transitions by retaining those showing the largest absolute (Eq. 1) and relative (Eq. 2) intensity distance between the target and the maximum interfering spectra.

$$T_t = \arg\max(x_i - y_i) \quad (1)$$

$$T_l = \arg\max(\frac{x_i}{y_i}) \quad (2)$$

Where ***Tt*** and ***Tl*** are the fragment indexes for the quantitative and qualitative transitions; index ***i*** denotes each of the shared fragments in the target spectra ***x*** and the maximum interfering spectra ***y***. The absolute distance will prioritize the fragments showing the best intensity response, while taking into account the selectivity of each fragment (quantitative transition). The relative distance will prioritize highly selective fragments (secondary qualitative transition). This strategy maximizes the Pareto efficiency[25] of the process, that is, it determines a set of fragments showing an optimal selectivity without compromising intensity. Of note, common neutral losses including loss of water and carbon dioxide for positive and negative mode and loss of ammonia for positive mode are not considered for qualitative transitions but are considered for quantitative transitions.

To determine the second and third transitions, the algorithm considers only fragments appearing in interfering spectra that share the qualitative fragment (first transition). This is to prevent other putative interfering molecules from sharing all the same three transitions (Extended Data Figure 4c). Thus, the combined use of these transitions translates into a more confident quantification of this target molecule when their peak area/concentration show a high linear relation. This means that the three transitions should lead to the same absolute concentrations, and only when another compound is interfering, different concentrations could be obtained Overall, the MRM configuration can not provide absolute specificity, however, METLIN-MRMs computational optimization reduces the number of interfering molecules at the same time that facilitates the identification of possible interferences.

XCMS-MRM automated data processing algorithm:

To process transition signals, first, an automatic baseline removal preprocessing is applied to subtract the baseline drift that LC/MS chromatograms are usually affected with, and provide with a more reliable peak-area quantification. Similarly as in previous work[26], this baseline is approximated by a moving-minimum filter, and according to a minimum compound peak width (a user-defined value in seconds). After the preprocessing, a fast Fourier transform-based peak detection algorithm is applied to detect the different peak signals based on the same user-defined minimum peak width parameter. If multiple transitions for the same metabolite are used, XCMS-MRM focuses on the information from all the transitions for a given compound and uses this information to detect the peak of interest more efficiently. This is of special importance when more than one peak is found in a transition, and only by inspecting the transitions, the peak related to the compound of interest can be identified (Extended Data Figure 1). Unimodality through an isotonic regression is applied to the final

chromatographic peak profile to improve the quality of the extracted peak, to assure that only one peak is integrated and thus minimize the interference of noise and low-concentrated and co-eluted peaks[27].

XCMS-MRM workflow:

After tandem MS measurements, raw data files are uploaded into the XCMS-MRM cloud platform. Raw data files can be uploaded in the commonly accepted chromatography interchange open standard formats, mzML, or also in different vendor files such as Agilent or Bruker .d files, Waters raw files or Sciex scan files. Many other vendor files can be easily translated into mzML by the broadly used ProteoWizard software[28]. Next, data is processed by XCMS-MRM to automatically detect and integrate fragment peaks. XCMS-MRM allows multiple transitions for the same molecule, which are analyzed as a compound (e.g., detecting the retention time when multiple peaks are found in a transition). To allow this automatic processing, users must provide with a list of the target molecule names along with their respective precursoi ions, and quantitative/qualitative fragments *m/z* values. Details on the selection of those parameters are described in the XCMS-MRM documentation available in the cloud platform. After processing, XCMS-MRM provides with a list of the target molecules, along with the mean retention time, and the coefficient of determination (R^2) between integrated peak areas across samples. This R^2, which value ranges between 0 and 1, allows for a rapid assessment of the automated processing results of XCMS-MRM, even when standards were not spiked. For example, values close to 0 mean that the quantitative and qualitative fragment peak areas do not follow a linear relation among samples, which suggest the presence ol co-eluting molecules and/or suppression effects, and a manual review of the peak integration is necessary. R^2 values close to 1 suggest successful peak integration. Users can manually inspect the XCMS-MRM automated integration of each peak in each sample and adjust integration boundaries in case of strong co-elution or chromatographic shifts.

Cooperative application of XCMS-MRM for the analysis of tandem MS samples.

To demonstrate the capabilities of XCMS-MRM as a cooperative resource, six samples from the same pool of human serum sample spiked with standards (standard addition calibration method) were analyzed in four different laboratories worldwide. The sample was analyzed with four different QqQ instruments including an Agilent iFunnel 6495 and 6490, a Sciex QTRAP 5500 and a Waters Xevo TQ-S.

A pool serum sample spiked with ten standards (Extended Data Table 2) at different concentrations (a total of 1 serum sample and 5 spiked samples, see Methods for details) was analyzed in four independent laboratories (The Scripps Research Institute, University of Umea, University of Lausanne and Imperial College). Samples were analyzed with different instruments, with reverse phase chromatography with slightly chromatographic method variations. XCMS-MRM was used to analyze the data by each laboratory. Results (Extended Data Table 1) show the concentration of each metabolite in the sample, obtained using a calibration curve by standard addition. The independent quantitative analysis of those molecules in all four laboratories with different vendor equipment lead to similar quantitative results under expected instrumental and analytical error (a median coefficient of

variation between laboratories of 15%). The quantitative analysis of cholesterol was discarded because its concentration was under the limits of quantification. Also, due to its small concentration (median concentration of 30 nM), palmitoylcarnitine showed a high coefficient of variation.

Application of XCMS-MRM for drug and metabolite quantification.

To demonstrate the performance of XCMS-MRM for high-throughput quantification of molecules we compared the quantitative results by XCMS-MRM over the results after a supervised quantification of drugs and metabolites in a certified toxicology laboratory. The concentrations of cocaine, benzoylecgonine and cocaethylene were analyzed by triple quadrupole MS in 57 whole blood samples in duplicate (114 samples) and the concentrations of tetrahydrocannabinol (THC), 11-Hydroxy-THC and 11-Nor-9-carboxy-THC were analyzed in 120 serum samples in duplicate (240 samples). We calculated the relative concentration error between XCMS-MRM and the reference values by the certified laboratory. We observed that most of these errors were below 10%. Of note, larger errors for 11-Hydroxy-THC, THC and 11-Nor-9-carboxy-THC were observed compared to cocaine, benzoylecgonine, cocaethylene. This can be explained due to the low ionization efficiency of the THC and its metabolites, together with their relatively low concentration in blood due to their pharmacokinetics. This explains why a small difference in the peak integration will induce a larger difference in the calculated concentration, independently of the peak integration algorithm.

Small molecule extraction and tandem QqQ/MRM analysis.

Human plasma samples (100 μL), mice plasma (30 μL) metabolites were extracted using 400 μL and 120 μL, respectively of cold MeOH:ACN (1:1, v/v) to maintain MeOH:ACN:H_2O (2:2:1, v/v) ratio, respectively. RAW264.7 cells (1×10^6) metabolites were extracted with 500 μL MeOH:ACN:H_2O (2:2:1, v/v). The samples were then vortexed for 30 s. To precipitate proteins, the samples were then sonicated for 15 min and incubated for 1 h at −20 °C, followed by 15 min centrifugation at 13000 rpm and 4 °C. The resulting supernatant was removed and evaporated to dryness in a vacuum concentrator. Dry extracts were then reconstituted in 100 μL of ACN:H_2O (1:1, v/v), sonicated for 15 min and centrifuged 15 min at 13000 rpm and 4 °C to remove insoluble debris. Supernatants were transferred to HPLC vials and stored at −80 °C prior to LC-MS/MS analysis.

For the interlaboratory quantitative analysis, the human plasma was extracted with MeOH:ACN (1:1, v/v) spiked with different concentrations of the metabolites listed in Extended Data Table 2. Different instrumentation was used as in the different laboratories. A Sciex QTRAP 5500, an Agilent 6495 and a Waters Xevo TQ-S were used for this analysis. For the separation, reversed phase chromatography was selected as it is commonly used in most laboratories. The chromatography settings for all laboratories are based on the following method with some slight variations depending on availability at the different sites. An ACQUITY UPLC BEH C18 (2.1 × 100 mm, 1.7 μm, Waters Corp. Milford, MA) column held at 50 °C was used. The mobile phases used were H_2O + 0.1% formic acid and ACN + 0.1% formic acid fro A and b, respectively. 2μL were injected and a flow rate of 400

μL/min was used. The gradient used consisted in 99% A for 1 minute to 1% A over 9 minutes and held at 1% A for 1 minute.

For the drug quantitation analysis, blood samples were from roadside drug testing were utilized. 10 μL (dried blood spot) and 1 mL of whole-blood were used for the quantitation of cocaine and THC, respectively. 200 μL of MeOH were used for the extraction of cocaine. For the THC analysis, 200 μL of acetic acid 10% in H_2O (v/v), and 5 mL of hexane: ethyl acetate (9:1, v/v) were used as extraction solvents. The analysis was conducted in a QqQ 5000 (Sciex, Framingham, MA) coupled to LC Ultimate 3000 (Dionex, Suunnyvale, CA). The analysis of the cocaine and its metabolites was performed with a Kinetex HILIC column (50 × 2.1 mm, 2.6 μm, Phenomenex, Torrance, CA). H_2O (10 mM ammonium formate, pH = 3.3) and ACN were used as mobile phases A and B, respectively. Concerning the tetrahydrocannabinol (THC) and its metabolites, the analysis was performed with a Kinetex C18 column (50 × 2.1, 2.6 μm, Phenomenex, Torrance, CA). H_2O (5 mM ammonium formate, pH = 7.7) and ACN were used as mobile phases A and B, respectively. The gradient for cocaine analysis consisted in 3% A for 0.3 minutes, increased to 30% A at 2.8 minutes, then to 50% A at 2.9 minutes and held for 1 minute at a flow rate of 700 μL/min. The gradient for THC analysis consisted in 70% A to 10% A in 2 minutes, held at 10% A for 4 minutes at a flow rate of 600 μL/min. The MRM transitions selected were: m/z 304.0 -> m/z 182.0; m/z 290.0 -> m/z 168.0 and m/z 200.0 -> m/z 182.0 for cocaine, benzoylecgonine, cocaethylene, respectively. Calibration curves were made of four points (10, 20, 50 and 200 ng/mL) and analyzed in duplicate. The transitions m/z 315.2 -> m/z 193.0; m/z 331.300-> m/z 313.0 and m/z 343.2 -> m/z 191.0 were used for THC, 11-Hydroxy-THC and 11-Nor-9-carboxy-THC respectively. Calibration curves were made of four points (1, 5, 10 and 20 ng/mL for THC and 11-Hydroxy-THC) and for (5, 10, 50 and 100 ng/mL for 11-Nor-9-carboxy-THC) and analyzed in duplicate.

The analysis of N1,N12-Diacetylspermine in mice plasma with calibration curves was conducted as described elsewhere[29]. Briefly, 8 μL of plasma extracts were injected into a Scherzo SM-C18 column (150 × 0.5 mm, 3 μm, Imtakt, Philadelphia, PA) using an Agilent Technologies series 1290 UPLC with a gradient mobile phase of 5 mM ammonium acetate (mobile phase A) and 50 mM acetate and acetonitrile (1:1, v/v) (mobile phase B) at a flow rate of 20 μL/min: 2% B for 5 min, to 17% B at 11 min, to 98% B at 13.5 min and held for 5 min. The MRM transitions selected were: *m/z* 287.2 -> *m/z* 100.1 and *m/z* 287.2 -> 171.1 and an Agilent 6495 triple quadrupole mass spectrometer was used.

RAW264.7 macrophage-like cells were obtained from the American Type Culture Collection (Rockville, MD). The cells were grown in Dulbecco's modified Eagles medium supplemented with 10% (v/v) fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin, and 2 mM 1-glutamine at 37 °C in a humidified atmosphere of C02/air (1:19). The cells were washed with PBS, scraped in ice-cold water and stored at −80 °C prior to metabolite extraction and LC-MS/MS analysis.

The validation study of 97 metabolites in RAW264.7 cells was performed by spiking 5 samples with different concentrations of the 97 metabolites (6 samples total, 1 unspiked and 5 spiked) and analyzed using both the ranked (computationally optimized) and unranked

(experimentally optimized and public repository, when available) transitions during the same run. For this analysis, 2 different HILIC methods were used; one using a UPLC column ACQUITY UPLC BEH Amide (2.1 × 100 mm, 1.7 μm, Waters Corp. Milford, MA) in positive mode and the other used an HPLC column Luna HILIC (150 × 2 mm, 3 μm, Phenomenex, Torrance, CA) in negative mode. The gradient for UPLC analysis consisted on (400 μL/min flow rate) 99% A for 1 minute, 35% A over 13 minutes, 60% A over 3 minutes and held at 60% A for 1 minute. The composition for the mobile phases A and B consisted on H_2O + 0.1% formic acid and ACN + 0.1% formic acid, respectively. The gradient for the HPLC analysis consisted on (300 μL/min flow rate) 5% A for 2 minutes, 90% A over 13 minutes, 100% A over 3 minutes and held 100% A for 3 minutes. The composition for the mobile phases A and B consisted on 5% ACN (40 mM ammonium acetate and 20 mM ammonium hydroxide) and 95% ACN, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS.

This research was partially funded by the Medical Research Council (UK) grant number MC-PC-12025 (M.R.L., J.K.N.); by Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory for the US Department of Energy, Office of Science, Office of Biological and Environmental Research under contract number DE-AC02-05CH11231 (G.S.); and National Institutes of Health grants R01 GM114368-03, P30 MH062261-10, and P01 DA026146-02 (G.S.).

REFERENCES.

1. Glish GL, Vachet RW The basics of mass spectrometry in the twenty-first century. Nat. Rev. Drug Discov. 2, 140–150 (2003). [PubMed: 12563305]
2. Patti GJ, Yanes O, Siuzdak G Innovation: Metabolomics: the apogee of the omics trilogy. Nat. Rev. Mol. Cell Biol. 13, 263–269 (2012). [PubMed: 22436749]
3. Kusebauch U et al. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. Cell 28, 766–778 (2016).
4. Nordströmm A, Lewensohn R Metabolomics: Moving to the Clinic. J. Neuroimmune Pharmacol. 5, 4–17 (2010). [PubMed: 19399626]
5. Shevchenko A, Simons K Lipidomics: coming to grips with lipid diversity. Nat. Rev. Mol. Cell Biol. 11, 593–598 (2010). [PubMed: 20606693]
6. Wang TJ et al. Metabolite profiles and the risk of developing diabetes. Nat. Med. 17, 448–453 (2011) [PubMed: 21423183]
7. Strathmann FG, Hoofnagle AN Current and future applications of mass spectrometry to the clinical laboratory. Am. J. Clin. Pathol. 136, 609–616 (2011). [PubMed: 21917684]
8. Vogeser M, Seger C Mass spectrometry methods in clinical diagnostics – state of the art and perspectives. Trends Anal. Chem. 84, 1–4 (2016).
9. Beck A Goetsch L Dumontet C Corvaia N Strategies and challenges for the next generation of antibody-drug conjugates. Nat. Rev. Drug Discov. 16, 315–337 (2017). [PubMed: 28303026]
10. Fan TW-M. et al. Stable isotope-resolved metabolomics and applications for drug development. Pharmacol. Ther. 133, 366–391 (2012). [PubMed: 22212615]
11. Triolo A, Altamura M, Cardinali F, Sisto A, Maggi CA Mass spectrometry and combinatorial chemistry: a short outline. J. Mass. Spectrom. 36, 1249–1259 (2001). [PubMed: 11754116]
12. Hoffmann WD, Jackson GP Forensic mass spectrometry. Annu. Rev. Anal. Chem. 8, 419–440 (2015).

13. Sauer S, Kliem M Mass spectrometry tools for the classification and identification of bacteria. Nat. Rev. Microbiol. 8, 74–82 (2010). [PubMed: 20010952]
14. Waters MD, Fostel JM Toxicogenomics and systems toxicology: aims and prospects. Nat. Rev. Genet. 5, 936–948 (2004). [PubMed: 15573125]
15. Nicholson JK, Connelly J, Lindon JC, Holmes E Metabonomics: a platform for studying drug toxicity and gene function. Nat. Rev. Drug Discov. 1, 153–161 (2002). [PubMed: 12120097]
16. Lange V, Picotti P, Domon B, Aebersold R Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. 4: 222 (2008). [PubMed: 18854821]
17. Matsumoto M et al. A large-scale targeted proteomics assay resource based on an in vitro human proteome. Nature Methods. 14, 251–258 (2017). [PubMed: 28267743]
18. MacLean B et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968 (2010). [PubMed: 20147306]
19. Cai Y, Weng K, Guo Yan., Peng J, Zhu Z-J An integrated targeted metabolomic platform for high-throughput metabolite profiling and automated data processing. Metabolomics 11, 1575–1586 (2015).
20. Tsugawa H, Kanazawa M, Ogiwara A, Arita M MRMPROBS suite for metabolomics using large-scale MRM assays. Bioinformatics 30, 2379–2380 (2014). [PubMed: 24753485]
21. Guijas C et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal. Chem. 90, Just Accepted (2018).
22. Melnik AV et al. Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. Anal. Chem. 89, 7549–7559 (2017). [PubMed: 28628333]
23. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol. Cell. Proteomics 11, 1475–1488 (2012). [PubMed: 22865924]
24. Domingo-Almenara X et al. Annotation: A Computational Solution for Streamlining Metabolomics Analysis. Anal. Chem. 90, 480–489 (2018). [PubMed: 29039932]
25. Tomoiagă B, Chindriş M, Sumper A, Sudria-Andreu A, Villafafila-Robles R Pareto Optimal Reconfiguration of Power Distribution Systems Using a Genetic Algorithm Based on NSGA-II. Energies6, 1439–55 (2013).
26. Domingo-Almenara X et al. Compound identification in gas chromatography/mass spectrometry-based metabolomics by blind source separation. J. Chromatogr. A. 1409, 226–233 (2015). [PubMed: 26210114]
27. De Juan A, Vander Heyden Y, Tauler R, Massart DL Assessment of new constraints applied to the alternating least squares method. Anal. Chim. Acta 346, 307–318 (1997).
28. Kessner D, Chambers M, Burke R, Agus D, Mallick P ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. Bioinformatics. 24, 2534–2536 (2008). [PubMed: 18606607]
29. Johnson CH et al. Metabolism links bacterial biofilms and colon carcinogenesis. Cell Metabolism 21, 891–897 (2015). [PubMed: 25959674]

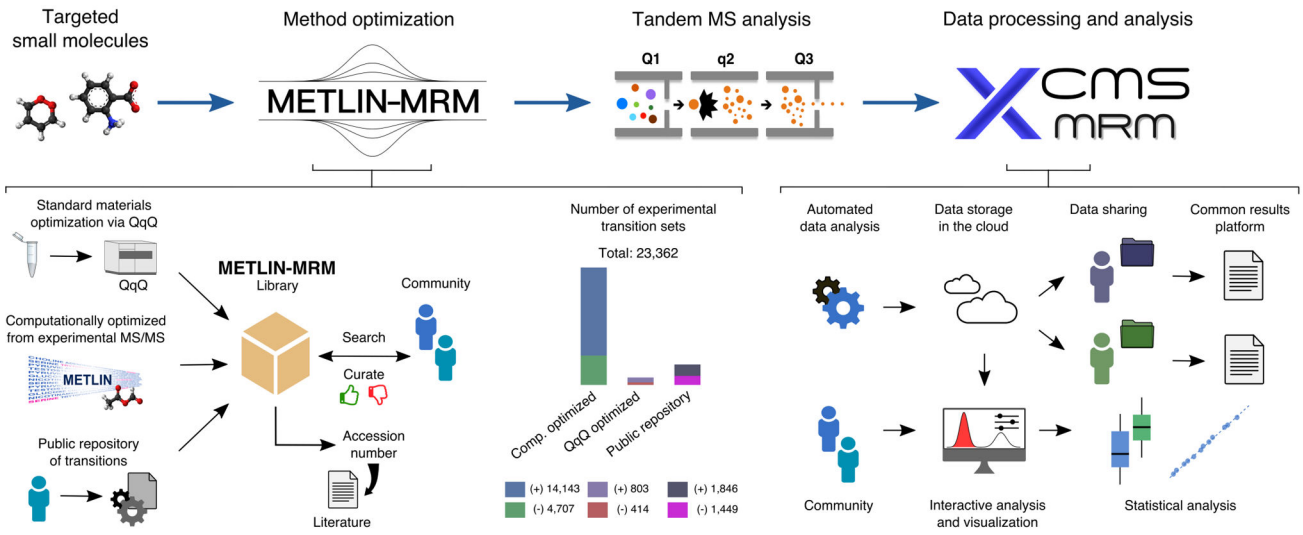

Fig. 1. Cloud-based tandem MS workflow.

Transitions for target molecules can be selected by means of METLIN-MRM library. Next, the resulting tandem MS data is uploaded into the cloud and automatically processed with XCMS-MRM. METLIN-MRM is composed of three types of transitions: (i) experimentally optimized by standard materials with QqQ via the established protocol, (ii) computationally optimized transitions using METLIN's MS/MS spectra and (iii) and public repository transitions extracted from literature. Experimentally optimized and computationally optimized transitions were optimized for sensitivity and selectivity, respectively. Additionally, METLIN-MRM serves as a public repository, where the community can upload and populate the library and use the accession number provided as citation in literature. The cooperative aspect of this workflow is highlighted by the fact that METLIN-MRM can be accessed by the community as opposed to optimized transitions remaining as in-house libraries or spreadsheets in supplementary materials of scientific literature. Furthermore, the community can also play an active role in database curation through a transition quality rating system. In XCMS-MRM, the cloud-based workflow allows users to universally process, visualize and share data from different vendor formats, facilitating results reproducibility and comparative analysis within and across laboratories. The community can interactively visualize experiments made public by users and use the accession numbers provided as citation in literature.

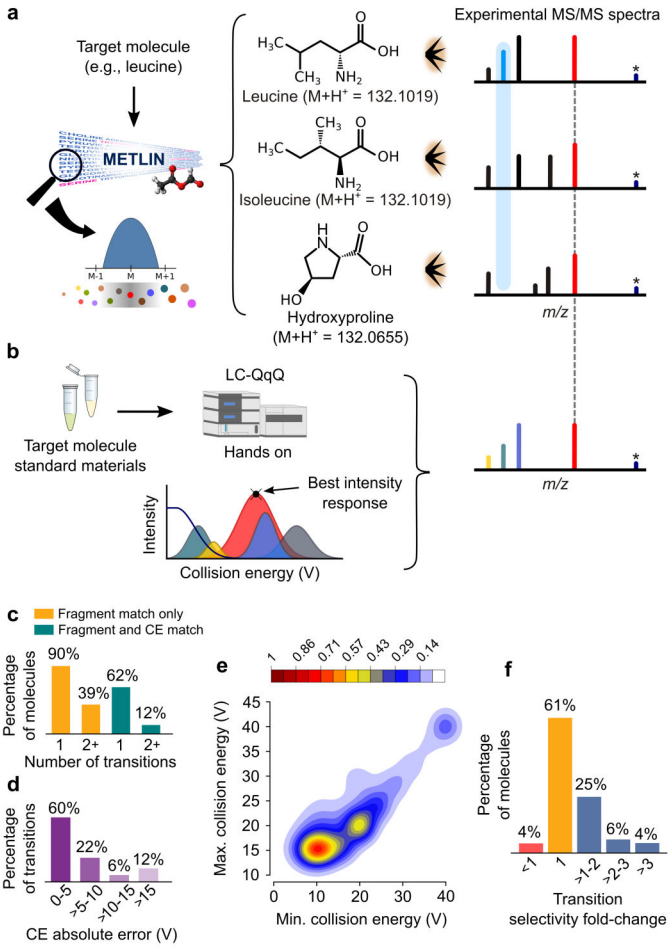

Fig. 2. METLIN-MRM.

a, in METLIN-MRM statistical ranking system, all molecules in the METLIN library with a precursor within ±0.7 Da window of the target molecule are compared (e.g., leucine is compared to isoleucine and hydroxyproline), and candidate transitions are selected based on their fragment selectivity. **b,** Experimental triple quadrupole transitions are optimized using standard materials prioritizing high intensity fragments (sensitivity) and can lead to misidentifications. **c**, Percentage of a total of 641 molecules sharing at least one (1) or two and more (2+) transitions (taking *m/z* and CE into account). For this plot, CE in EO and PR transitions (unranked) were rounded to the nearest discrete value (10, 20 or 40 V). **d**, CE error distribution between CO (ranked) and unranked transitions (for those transitions with the same product-fragment value). **e**, Two-dimensional density plot of the maximum and minimum collision energies for the same transition found across laboratories (EO and PR). **f**, Selectivity ratio: for each molecule, the ratio of minimum number of putative interfering molecules (PIM) of the unranked transitions divided by the minimum number of PIM of the ranked transitions. The plot shows the number of molecules where unranked (red) and ranked (blue) transitions are more selective (ratios <1, and >1 indicate better selectivity for unranked and ranked transitions, respectively).