Pau Orts Macip

LARGE-SCALE ANALYSIS OF MATABOLOMICS DATASETS

TREBALL DE FI DE GRAU

dirigit pel Dr. Xavier Domingo Almenara

Grau d'Enginyeria Biomèdica

UNIVERSITAT ROVIRA I VIRGILI

Tarragona

2024

This project is the result of the external internship carried in the Centre for Omics Sciences group at EURECAT, Reus, (EURECAT) under the supervision of Dr Xavier Domingo-Almenara.

It contains confidential information for the mentioned entity and for Dr Xavier Domingo-Almenara (xavier.domingoa@eurecat.org). In the complete version, from sections 2 to 5, including the objectives and hypothesis, materials and methods, results, and discussion, is confidential. Also, the annex (Section 7) is confidential.

Index

1 Introduction

1.1 Metabolomics, a growing omics science

Biological entities are systems, the collection of simple parts that work together as a single unity. Systems biology is an integrative discipline that connects molecular components within a single or multiple biological scales to physiological functions and phenotypes [1]. This biology-based field is subject to computational and mathematical analysis from experimental data which provide the understanding of complex interactions and dynamics at various levels, within cells, tissues, organs, and organisms [2]. Systems biology is sustained by the integration of the whole structural and functional information acquired from omics sciences, like genomics (including metagenomics and epigenomics), transcriptomics, proteomics, or metabolomics [4].

Particularly, metabolomics aims at investigating the activity and status of cellular and organismal metabolism, to delineate the end points of physiology and pathophysiology [5-6]. It involves the measurement of small molecule compounds, including endogenous and exogenous molecules, known as metabolites, that are the products and substrates of chemical reactions within biological systems. These metabolites are of huge importance since they can influence or even alter metabolic pathway regulation [7]. Amongst them, amino acids, lipids, nucleotides, carbohydrates, and organic acids can be found. Given this variety in chemical compounds and chemical properties, metabolomic research is a challenge in analytical chemistry, since there is no universal method for metabolome analysis. Moreover, there is a dynamic range of the metabolome [8] and compounds are frequently distributed over a broad extent of concentrations [9].

Metabolomics experiments directly reflect the activity of the metabolic network that leads to the production of metabolites and yields essential information about the underlying biological status of the system in question. In fact, some studies have given insights for the understanding of disease mechanisms and drug effects, as well as to improve the ability to predict personal disease progression or variation in drug response phenotypes.

To accomplish this, generally metabolomics' studies are not focused by any experiment but by the study of metabolism in a comprehensive and holistic approach. This is known as "untargeted" screen where thousands of unknown features are profiled and the relative differences in two conditions or across a population (semi-quantitation) are measured [9].

However, targeted experiments are also used in some cases since they often provide deeper insights by testing a specific hypothesis. For example, the absolute concentrations of molecules can be measured (absolute quantitation) or the rates or fluxes of the conversion of one molecule to another can be obtained [12]. Thus, a targeted metabolomics analysis requires substantial pre-existing knowledge, and its success depends on strength of the hypothesis being tested.

We can observe the main features of these three different types of analysis when talking about metabolomics in Figure 1.

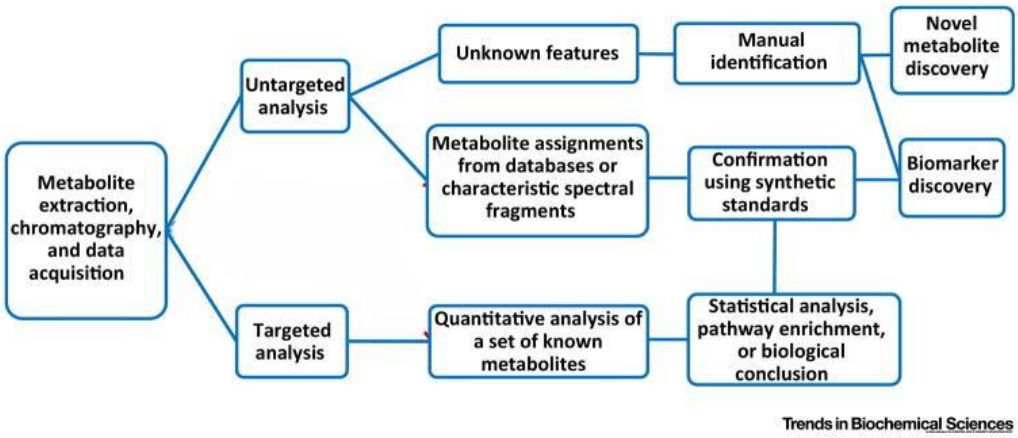


Figure 1. Targeted, semi-targeted and untargeted analysis. Picture from Xiaojing Liu et. Al. [9].

During this project, we are focusing on the first group: untargeted analysis. With this, we aim at finding patrons in the spectral features of 2 or more samples.

In all metabolomics experiments, to achieve the end goal of knowing the behaviour and relations of the metabolites, there are some common steps that every study follows (Figure 2), with the end goal of acquiring the intensities of certain features characterised by their mass to charge ratio (m/z). This will then be used to acquire knowledge about certain metabolomics pathways and patrons in some illnesses or behaviours.

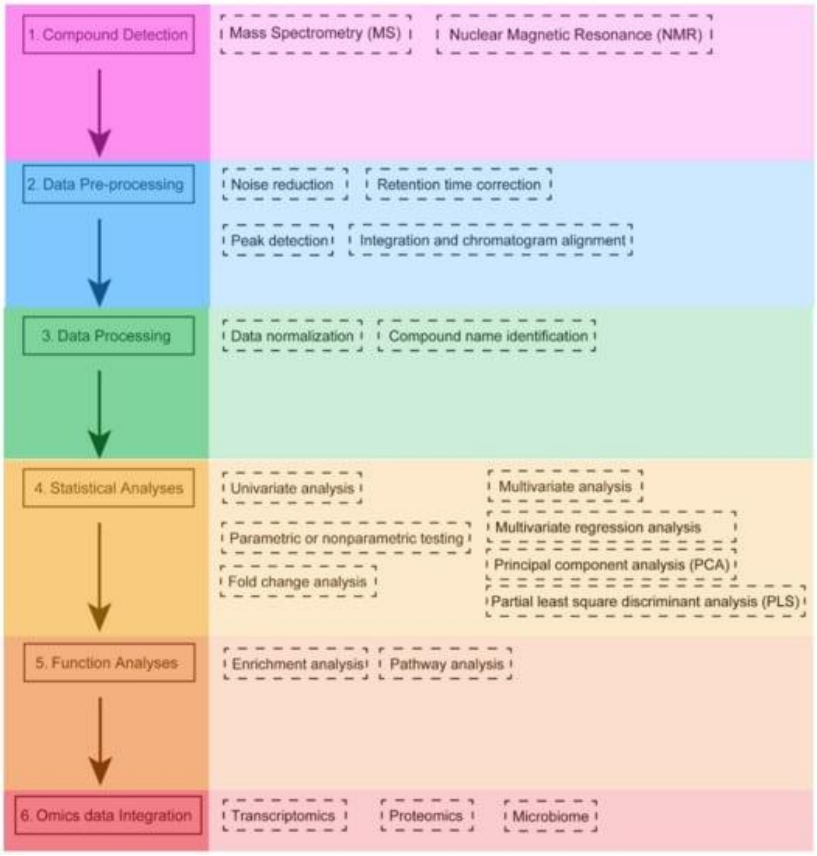


Figure 2. Typical workflow of metabolomics analysis. Metabolites are detected by using specific detection techniques (compound detection). Raw signals are then pre-processed to produce data in a suitable format for subsequent statistical analysis (data pre-processing). Then, data normalization is used to reduce the system and technical bias. For untargeted studies, metabolites are identified from spectral information in some given database (data processing). Univariate and multivariate statistical analyses are used to identify significantly expressed

metabolites (statistical analyses). Next, the significantly expressed metabolites are subsequently linked to the biological context by using enrichment and pathway analysis (function analyses). Finally, metabolomics data may be integrated with other omics data (transcriptomics, proteomics, or the microbiome) to gain a comprehensive understanding of the molecular mechanisms of pathophysiological processes (Omics data Integration). Picture from Yang Chen et al. [13].

According to what was being explained, an essential step is to acquire the information related to the molecular composition of the sample. We have two main methods for this: nuclear magnetic resonance (NMR) and mass spectrometry (MS). A quick overview of both options is presented in Table 1.

Table 1. Metabolomics platforms [9].

	Basis	**Pros**	**Cons**
MS	Acquires spectral data in the form of a mass-to-charge ratio (m/z) and a relative intensity of the ions.	High sensitivity Broad metabolite coverage Open-source software analysis	Not quantitative Destructive
NMR	Detects molecular features by measuring an intrinsic magnetic property of atomic nuclei.	Real time measuring Deeper structural information	Low sensitivity Less metabolite coverage

For this project mass spectrometry will be used, following the choice of untargeted analysis that was mentioned before. This decision will allow to analyse much more metabolites but not focusing that much on specific ones, thus giving rise to new and unexpected information.

1.2 Basis and workflow of mass spectrometry

Mass spectrometry (MS) is a technique used to analyse molecular masses of individual compounds and atoms by converting them into charged ions. This technique also provides quantitative information of an analyte at levels of structure specificity and sensitivity. Apart from allowing the study of these, also the reaction dynamics and chemistry of ions can be studied by analysing the ionization energy, enthalpy, and so on. Given all this, MS is the most versatile and comprehensive analytical technique currently in use [14].

As explained in the Table 1, amongst other features, the most desired from this technique are its unsurpassed molecular specificity because of its unique ability to measure accurate molecular mass and to provide information on structurally diagnostic ions of an analyte. Its ultrahigh detection sensitivity is another advantage: it has the ability to detect a single molecule. Furthermore, it has unparalleled versatility to determine structures of most classes of compounds, it is applicable to all types of samples, and it is combinable with high-resolution separation devices [14].

For reviewing the basic principles that mass spectrometry relies on, we need to fully understand that it works with ions, since unlike neutral species, the motions and direction of these are easy to manipulate and thus detect them. A quick overview of how a mass spectrometer works is shown in Figure 3.

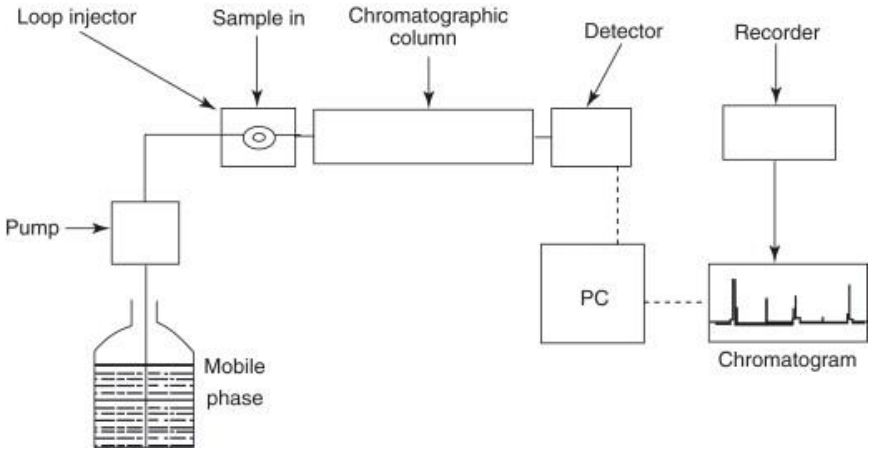


Figure 3 Workflow of mass spectrometer linked with liquid chromatography. Picture from science direct [15].

The first step of MS is ionization that converts analyte molecules or atoms into gas or liquid-phase ionic species. This step requires the removal or addition of an electron or proton. The excess energy transferred during an ionization event may break the molecules into characteristic fragments [14]. But this is a problem that will be addressed later. For this first step, many techniques are used, which can be classified as soft or hard, depending on the fragmentation degree [16]. Amongst others, we can find liquid chromatography (LC) that is the one chosen for the studies of this project.

As can be seen in Figure 3, in this process the ionization involves the four first blocks: the mobile phase, the pump, the injector, and the column. Thus, when coupled with a peak and features detection algorithm we can obtain the m/z and intensities of the metabolites.

A quick scope of how the LC works is the following: first, in the pump a stable flow rate is provided. Then, a loop injector is responsible of introducing the sample into a flowing liquid stream by using a conventional syringe. The mobile phase, which is a mix of the sample with a solvent, then runs through the column where the stationary phase is located, with certain chemical properties. While the mobile phase is running through the column the components interact is some different manner: more polar components interact more strongly and elute later [17].

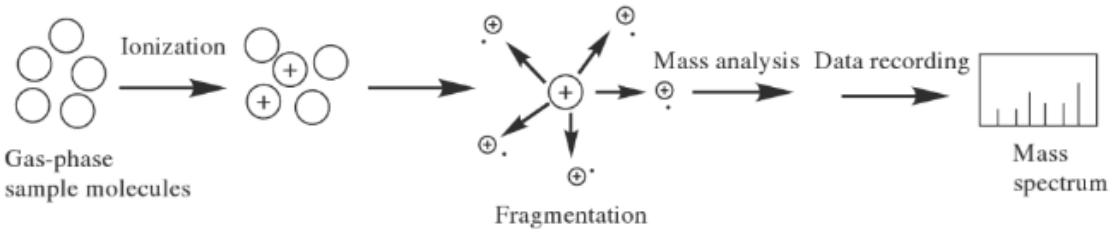


Figure 4 Basic concept of mass spectrometry analysis. Picture from Chabbil Dass et al. [14].

After the ionization process, the compound that entered now has produced one or multiple ion species (Figure 4). These species enter the mass analyser, where the ions are accelerated giving information about the m/z and intensities of them [18,19]. After being accelerated, the chromatography separation takes place. Every mass analyser uses a different technique, but most of them rely on magnetic or electric fields to control the motion of ions (Figure 5) [14]. The most used ones are Quadrupole Mass, Time of Flight (ToF), Ion Trap. The first one has a continuous mode of operation, whereas ToF

use a pulsed-based operation mode. Finally, ion trap, as its name refers, uses a ion trapping mode [20].

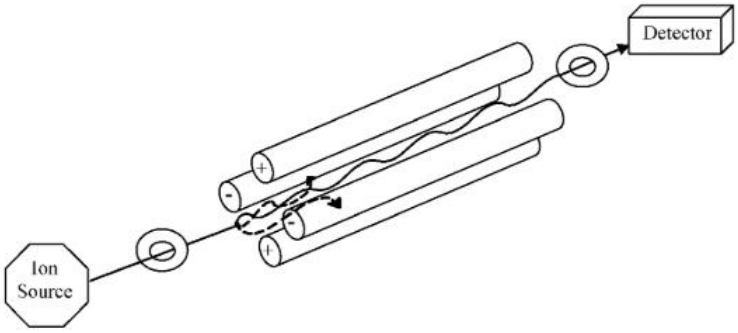


Figure 5. Representation of a quadrupole mass analyzer. Picture from Anas el Aneed et al. [21].

Once the particles have interacted in the mass analyser in different intensities and velocity, they are detected. A mass detector mainly detects the current signal generated from the passes or incident ions which are absolute or relative concentration of each analyte [22]. This is done by the phenomenon known as secondary electron emission. The number of these secondary electrons depends on the basic properties of the incident primary particle. The two basic forms of electron multipliers are discrete-dynode electron multiplier and the continuous-dynode electron multiplier [23]. Some important elements of the detectors are the electron multiplier, the faraday cups, and the photographic plates amongst others [23].

By this point, the data has already been generated. After being detected in the detector, all the ions' species under study have been classified and assigned a m/z value, a retention time value, and a intensity value. But the number of data is huge, so in order to process it and find relevant information (features), there exist different feature and peak detection algorithms.

1.3 Peak detection and feature detection: XCMS, ASARI

The feature detection algorithm will provide useful information by generating output data, including features (peaks with specific retention time and mass to charge ratio m/z); peak area, which is usually the preferred parameter to represent relative abundance of each metabolite in different samples [9]. For metabolomics datasets, m/z is used as the only criteria for feature identification, and therefore, many features often return multiple metabolite identities, which is caused by ion source fragmentation or isomers [24].

There exist different algorithms that can perform this key step, some of the most used are presented in Figure 6:

Software	Data Format	Statistics	Pathways[b]	Data Visualization	Isotope tracing	MS/MS[c]	Semi-targeted analysis[d]	Multi-omics integration	Source
XCMS	All[a]	√	√	√				√	xcmsonline.scripps.edu
13C XCMS	All	√			√				pattilab.wustl.edu/x13cms
MAVEN	All	√		√	√		√		maven.princeton.edu
MsXelerator	All	√		√	√		√		msmetrix.com
MetaboAnalyst	All	√	√	√					metaboanalyst.ca
MetAlign	All	√		√				√	metalign.nl
MZmine	All	√		√			√		mzmine.sourceforge.net
SIEVE	.raw	√		√	√		√		Thermo
Compound discover	.raw	√	√	√	√	√	√		Thermo
Mass Profiler	All	√	√	√				√	Agilent
MarkerLynx	.raw	√		√					Waters
MarkerView	.d	√		√					AB Sciex
MS-DIAL	All	√		√		√			prime.psc.riken.jp/Metabolomics_Software/MS-DIAL

Figure 6. Metabolomics data processing software. Picture from Xiaojing Liu et al. [9].

As we can see, each one has its own features, but in the end they all perform similarly (Figure 7). In general, this step includes noise reduction, retention time correction, peak detection and integration, and chromatographic alignment [13].

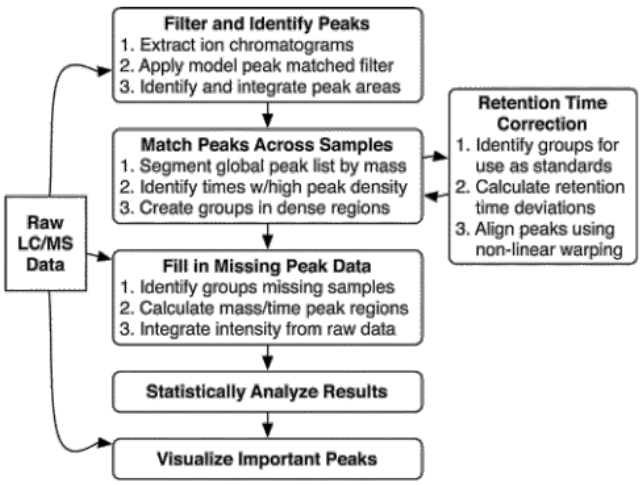


Figure 7 Flowchart of the strategy for feature detection. Picture from Smith CA et al. [25].

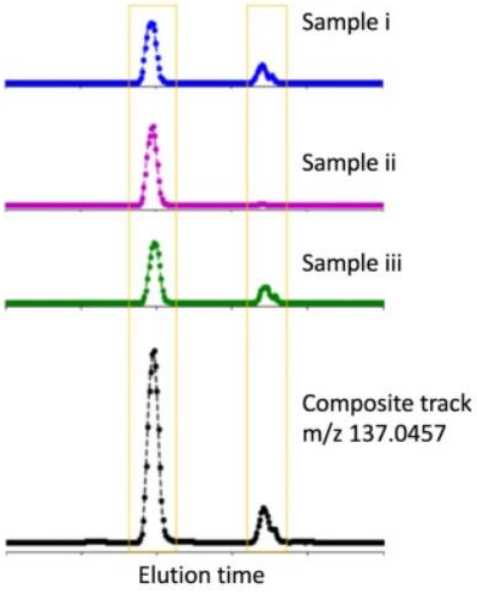
For instance, XCMS relies on four significant steps in raw data processing. Firstly, there is a peak detection where the software identifies whether a peak is due to a metabolite or due to noise. Then, there is peak matching for grouping peaks with similar retention times. This eliminates insignificant groups and resolves cases where a sample has multiple peaks in a group. After, retention time correction is performed by calculating the median retention time and deviation of each sample in the group. Finally, once the peak groups are established, absent samples in each group are identified; and when all retention times across samples is aligned the missing chromatographic peaks of samples are filled.

ASARI is also an interesting arising algorithm that works in a different basis. Here, all samples are treated at the same time, relying on the principle that "Mass alignment should not be conditioned on elution peak detection" [26].

1. In order to accomplish this, the first step here is mass track construction: the m/z values that are within a certain range are grouped, and then nearest

neighbour clustering (NNC) is performed, where each m/z value is assigned to its nearest peak, which is defined across the m/z distribution within the condition that it needs to be separated from another peak. Parallel processing is used for all samples [26].

2. After, there is the alignment of mass tracks, where the sample with the highest anchor mass tracks is designated as the reference sample [26].
3. Now retention time alignment is performed, first in the reference sample, and then locally weighted scatterplot smoothing (LOWESS) regression is computed for obtaining the relationship of the RT between the peaks [26].
4. Finally, the composite map is built given all the aligned mass tracks (Figure 8) [26].



5.

Figure 8 The "composite mass track" is a representation of data from all samples, by adding up the signals in corresponding mass tracks after retention time (RT) alignment. Picture from Shuzhao Li et. Al. [26].

Overall, in untargeted analysis the feature detection step can be challenging due to frequent ion fragmentation in samples, and thus, the results can be hard to interpret, since some features can be result of these ion source fragments that cannot be recognised [9].

1.4 Metabolite annotation, a key step

In large-scale studies searching for metabolic biomarkers with hundreds of samples, researchers often split the samples into smaller groups for analysis. This avoids lengthy analytical processes or the burden of preparing a massive number of samples at once. The identified ions from each analysis are then compared to find those that appear consistently across the groups. A key challenge, however, lies in recognizing a substantial number of derivative ions, such as isotopes, adducts, and fragments, which are typically overlooked. This oversight can lead to inaccurate metabolite identification during mass-based searches, as databases typically assume each derivative represents a unique molecular ion. To enhance the accuracy of metabolite identification, it's crucial to account for ions originating from the same metabolite [27].

After all the mass spectrometry process, and the feature identification, there must be performed peak or metabolite annotation. The term "annotation" can have two different meanings in metabolomics studies: (i) the tentative identification of a metabolite and (ii) the assignation of different metabolic features (adducts, charges, and losses) into a single value [28]. Here, the second one is wanted in other to relate these ion features to a bigger molecule that can achieve the first one.

Along this annotation process we need to focus on three different ion elements: adducts, isotopes, and in-source fragments (Figure 9). Adducts is an ion formed by interaction of two species, usually an ion and a molecule, and often within the ion source, to form an ion containing all the constituent atoms of one species as well as an additional atom or atoms [29]. Isotopes are variants of atoms of the same chemical elements, which have the same number of protons but different number of neutrons. As a result, the atoms of the same element may have different masses depending on the number of neutrons they have [27]. Although these two ion elements are of great importance, we will only focus on the third type. In-source fragments is the only one that is not wanted, and ionization techniques try to be as soft as possible to avoid it, but it still happens.

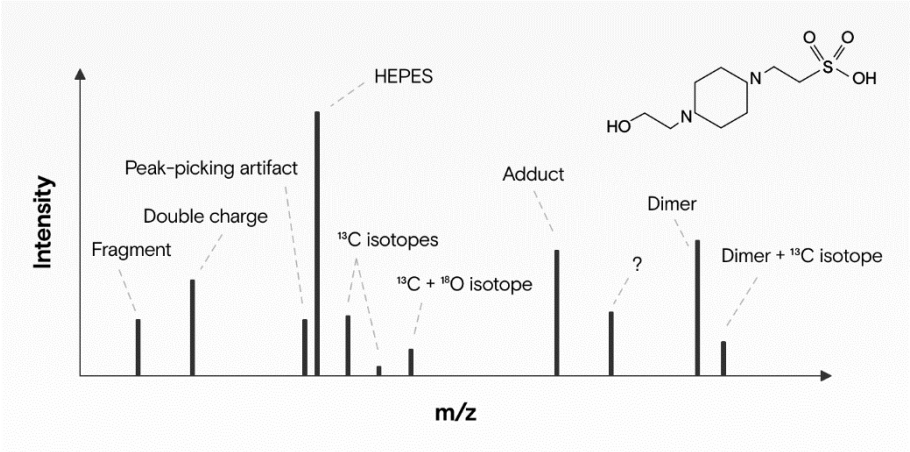


Figure 9. Example of the m/z values of the three different ion elements. Picture from Markus Schmitt [30].

In the past few years, many new algorithms and computational tools for improving this annotating step in MS-based metabolomics have been introduced [31].

For instance, the R-package CAMERA (Collection of Algorithms for MEtabolite pRofile Annotation) is quite known. This algorithm performs ion-annotation in two steps [32]. In the first step, the detected peaks with similar retention times are roughly grouped together using a sliding retention time window. Within each group, the EICs of the peaks are extracted and the peaks are clustered into smaller groups based on the Pearson correlation between their EICs. The m/z difference between each peak pair within a group is calculated and compared to known m/z relationships between different ion formations. The two ions are considered to come from the same compound if their m/z difference can be explained by one of the known m/z relationships [27].

Other algorithms also have been developed, most of them consisting of the comparison of identified features or peaks with ion fragments databases, which have almost every molecule and its possible fragments, isotopes and adducts.

Approaches based on detecting metabolite in-source fragments have been shown as an efficient alternative for peak annotation [33, 46] These algorithms rely on the fact that more than 90% of molecules in a typical metabolomics analysis undergo molecular

fragmentation yielding in-source fragments [35, 47]. These in-source fragments in MS1 data are then matched to low-collision energy spectral libraries to provide both feature and metabolite annotation (Figure 10).

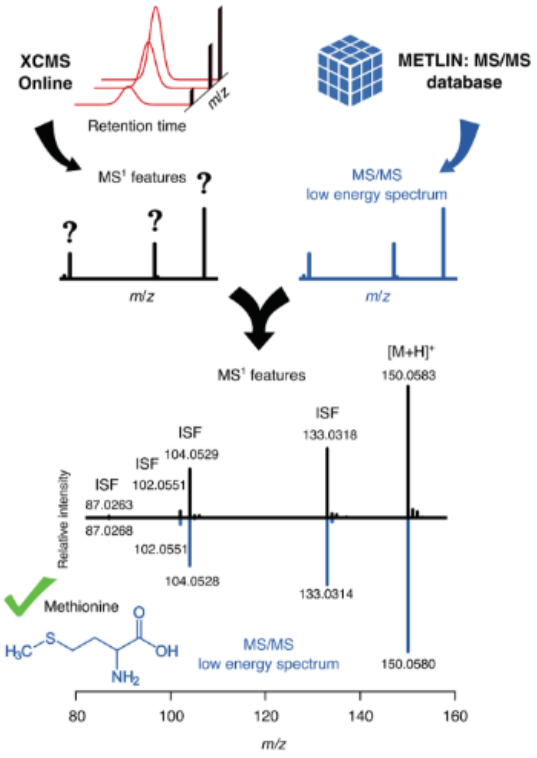


Figure 10. An example of how the procedure of metabolite annotation would be performed. Here METLIN database is used. Picture from Xavier Almenara et. Al. [33].

However, a proper metabolite identification is still a big struggle in untargeted metabolomics since only a small fraction of the thousands of metabolites in samples can be annotated and identified at a satisfactory confidence level [34]. This can be associated to different facts: during the sample process there is a high redundancy of features which are linked to the same metabolite owing to the existence of many in-source fragments, isotopes and adducts, also without a previous knowledge of monoisotopic masses the search in libraries of significant features may lead to miss annotations [35,36].

From this point and to the conclusion, the project is confidential. What is explained here are materials and algorithms used and implemented in order to annotate metabolites and molecules result of metabolic reactions from different public studies. After annotating them, high correlated metabolite relations will be extracted.

2 Hypothesis and objectives

3 Materials and methods

3.1 Fragments library: NIST

3.2 Experimental data: metabolomics workbench

3.3 Annotation algorithm implementation

4 Results

4.1 Metabolite correlation

4.2 Metabolite count

4.3 Highly correlated metabolites

4.4 Variation in metabolite correlations

4.4.1 L-Lysine – L-Arginine

4.4.2 Benzylimidazole and tryptophan

4.4.3 Xanthine and xanthose

5 Discussion

6 Conclusion

It was concluded that the annotation functions using the NIST database have given results of much interest, even though using free studies that vary between them. Many metabolites have been annotated across the studies and many of them hold relations that keep consistent in different studies, what suggests that the efficacy and accuracy of the process is good.

NIST database is an interesting database that, in first place, is free of use. Also, it includes many metabolites and its fragments, providing a good matching for many unknown features. Thus, if complemented with a good algorithm, good results can be obtained as it can be observed in our study. The algorithm was sometimes heavy to run due to the large data that was processing but has proved to work for many different studies in humans as well in mice. Also, the annotation functions are very precise and produce lots of information for each feature annotated.

In the results, it is observed that across studies of the same species, same metabolites are annotated in many studies, and many of these hold strong relations between them. Between both species analysed, also many similarities can be observed as for the metabolites annotated and the relations that these hold. All in all, it can be concluded that even though the necessity of some possible optimization of the code for a better time of computation and better annotation percentage and metabolite identification, this code could already be of use in large-scale analysis of metabolomics datasets.

In this project, the lack of deep knowledge in the field of bioinformatics has resulted in a slow process, where understanding many parts of the project took more time than desired. Also, a result of this, is an algorithm that can still be more optimised giving better results, and a better analysis of the results. But still, a good treatment and processing of the data was accomplished all along, understanding in all moments what was being done and the reason of it. Many R-based functions and instructions were used, and many were learnt.

The need to continue advancing and innovating in the field of metabolomics is clear. As reviewed, the organism is a very complex system that we still do not fully understand. As a result, when performing omics sciences, specially metabolomics, much of the information obtained is dismissed, what prevents from acquiring new knowledge about different conditions of the organism under study. In metabolomics, the bottleneck is found in the annotation process, rather than in the sample analysis or feature detection. Thus, further research should be centred in this topic. Based on our experience on this project, a more complete database should be developed (if that is still possible) in order to match even the smallest fragments and features, and also an algorithm that was the most precise possible and could distinguish perfectly between different metabolites.

In conclusion, current tools and methods for metabolite annotation have greatly advanced the field of metabolomics. However, fully annotating and interpreting metabolites from complex mass spectrometry data still requires the creation of new computational tools, algorithms, and instrument improvements. Ongoing efforts in these areas are essential to further progress and fully realize the potential of metabolomics for comprehensive metabolite annotation.

7 Annex

7.1 Annotation functions

7.2 Code to adapt NIST library.

7.3 Code to download data

7.4 Code to process data.

7.5 Code to make correlations.

7.6 Code to count the frequence in which metabolites are annotated across studies.

7.7 Code to find highly correlated pairs of metabolites.

Bibliography

1. Tavassoly I, Goldfarb J, Iyengar R. Systems biology primer: the basic methods and approaches. Essays Biochem. 2018 Oct 26;62(4):487–500.
2. Voit EO. A First Course in Systems Biology. Second edition. | New York : Garland Science, 2017.: Garland Science; 2017.
3. Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multiomics data for machine learning analysis. Comput Struct Biotechnol J. 2021;19:3735–46.
4. Cajka T, Fiehn O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. Anal Chem. 2016;88(1):524–45.
5. Guma M, et al. Metabolomics in rheumatic diseases: desperately seeking biomarkers. Nat Rev Rheumatol. 2016;12(5):269–81.
6. Grüning NM, Rinnerthaler M, Bluemlein K, Mülleder M, Wamelink MMC, Lehrach H, et al. Pyruvate Kinase Triggers a Metabolic Feedback Loop that Controls Redox Metabolism in Respiring Cells. Cell Metab. 2011 Sep;14(3):415–27.
7. Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'auria J, Ewald J, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Available from: https://doi.org/10.1038/s41592-021-01197-1
8. Castelli FA, Rosati G, Moguet C, Fuentes C, Marrugo-Ramírez J, Lefebvre T, et al. Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. Available from: https://doi.org/10.1007/s00216-021-03586-z
9. Xiaojing Liu and Jason W.Locasale, Metabolomics - a primer.
10. Breitling R, et al. Ab initio prediction of metabolic networks using Fourier transform mass spectrometry data. Metabolomics. 2006;2(3):155–164.
11. Shin SY, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46(6):543–50.
12. Park JO, et al. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat Chem Biol. 2016;12(7):482–9.
13. Yang Chen, En-Min Li and Li-Yan Xu, Guide to Metabolomics Analysis: A Bioinformatics Workflow. Metabolites 2022. 12(4), 357; https://doi.org/10.3390/metabo12040357
14. Chhabil Dass. Fundamentals of Contemporary Mass Spectrometry. 2006.
15. https://www.sciencedirect.com/topics/chemical-engineering/liquid-chromatography
16. Brunnée C. The ideal mass analyzer: Fact or fiction? Int J Mass Spectrom Ion Process. 1987 Jun;76(2):125–237.
17. Robert E. Ardrey, Liquid chromatography – mass spectrometry: An introduction. University of Huddersfield, Huddersfield, UK.
18. Ho CS, Lam CWK, Chan MHM, Cheung RCK, Law LK, Lit LCW, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. Clin Biochem Rev. 2003;24(1):3–12.
19. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: An integrated strategy for compound spectra extraction and annotation of LC/MS data sets. Available from: http://pubs.acs.org/
20. Faull KF, Dooley AN, Halgand F, Shoemaker LD, Norris AJ, Ryan CM, et al. Chapter 1 An Introduction to the Basic Principles and Concepts of Mass Spectrometry. In 2008. p. 1– 46.
21. Anas El-Aneed, Aljandro Cohen, Joseph Banoub. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. In 2009
22. Stanislav SR, Jonathan VS (2010) A mass spectrometry primer for mass spectrometry imaging. Methods Mol Biol. 656: 21– 49.
23. Sharad Medhe. Mass Spectrometry: Detectors Review. Chemical and Biomolecular Engineering. Vol. 3, No. 4, 2018, pp. 51-58. doi: 10.11648/j.cbe.20180304.11
24. Xu YF, et al. Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography-mass spectrometry-based metabolomics. Anal Chem. 2015;87(4):2273–81.
25. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal Chem. 2006 Feb 1;78(3):779–87.
26. Shuzhao Li, S., Siddiqa, A., Thapa, M. et al. Trackable and scalable LC-MS metabolomics data processing using asari. Nat Commun 14, 4113 (2023). https://doi.org/10.1038/s41467-023-39889-1

27. Varghese RS, Zhou B, Nezami Ranjbar MR, Zhao Y, Ressom HW. Ion annotation-assisted analysis of LC-MS based metabolomic experiment [Internet]. 2012. Available from: http://www.proteomesci.com/content/10/S1/S8
28. Joanna Godzien, Alberto Gil de la Fuente, Abraham Otero, and Coral Barbas. Chapter Fifteen – Metabolite Annotation and Identification. Comprehensive Analytical Chemistry, Volume 82, 2018.
29. McNaught AD, Wilkinson A. IUPAC:Compendium of Chemical Terminology. 2. Oxford:Blackwell Science; 1997.
30. Markus Schmitt. Reducing Noise in Untargeted Metabolomics with Binner.
31. Bauermeister A, Mannochio-Russo H, Costa-Lotufo L V., Jarmusch AK, Dorrestein PC. Mass spectrometry-based metabolomics in microbiome investigations. Nat Rev Microbiol. 2022 Mar 22;20(3):143–60
32. Tautenhahn R, Böttcher C, Neumann S. In: Annotation of LC/ESI-MS Mass Signals in Bioinformatics Research and Development. Hochreiter S, Wagner R, editor. Vol. 4414. Springer Berlin/Heidelberg; 2007. pp. 371–380.
33. Xavier Domingo-Almenara, J. Rafael Montenegro-Burke, Carlos Guijas, Erica L.-W. Majumder, H. Paul Benton, and Gary Siuzdak. Autonomous METLIN-Guided In-source Fragment Annotation for Untargeted Metabolomics. Anal. Chem. 2019, 91, 3246-3253.
34. Beniddir MA, Kang K Bin, Genta-Jouve G, Huber F, Rogers S, van der Hooft JJJ. Advances in decomposing complex metabolite mixtures using substructure- and network-based computational metabolomics approaches. Nat Prod Rep. 2021;38(11):1967–93.
35. Domingo-Almenara X, Montenegro-Burke JR, Benton HP, Siuzdak G. Annotation: A Computational Solution for Streamlining Metabolomics Analysis. Vol. 90, Analytical Chemistry. American Chemical Society; 2018. p. 480–9.
36. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics. 2006 Dec 28;7(1):234.
37. National Institute of Standards and Technology (NIST) Mass Spectral Database: https://www.nist.gov/srd/nist-standard-reference-database-1a
38. https://chemdata.nist.gov/dokuwiki/lib/exe/fetch.php?media=chemdata:nist17:nistms_ver23man.pdf
39. Metabolomics Workbench website: https://www.metabolomicsworkbench.org/
40. https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
41. The Human Metabolome Database (hmdb): https://hmdb.ca/
42. Guengerich, F. P. (2008). Cytochrome P450 and chemical toxicology. Chemical Research in Toxicology, 21(1), 70-83.
43. Guengerich, F. P. (1991). Reactions and significance of cytochrome P-450 enzymes. Journal of Biological Chemistry, 266(16), 10019-10022.
44. Ohta, H., & Nishikimi, M. (1999). Differences in the activity of xanthine oxidase between mice and humans. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 123(1), 81-85.
45. Caskey, C. T., & Patterson, D. (1977). Biochemical Genetics of Purine Metabolism. Annual Review of Biochemistry, 46, 343-368.
46. Broeckling, C. D., et al. Enabling Efficient and Confident Annotation of LCMS Metabolomics Data
47. Giera, M., Aisporna, A., Uritboonthai, W. et al. The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation. Nat Metab (2024).