



Metabolite names and identifiers: how far are we from interoperability?

Ghina Hajjar¹ · Franck Giacomoni¹ · Mathieu Umeç¹ · Marie Lefebvre¹ · Mariana P. C. Pimentel^{2,3} · Sylvain Prigent^{2,3} · Cécile Cabasson^{2,3} · Monica Chagoyen⁴ · Pierre Pétriacq^{2,3} · Blandine Comte¹ · Estelle Pujos-Guillot¹

Received: 30 October 2025 / Accepted: 28 January 2026
© The Author(s) 2026

Abstract

Introduction Today, metabolomics literature suffers from ambiguity in metabolites' nomenclature, making difficult inter-comparison between publications and can lead to misinterpretations. Progress in the implementation of FAIR principles in metabolomics in various scientific communities is therefore imperative for successful comparisons across studies and for moving towards more large-scale metabolomics.

Objectives In this context, the aim of the present work was to explore the potential ambiguities that may be introduced during metabolite contextualization and reporting, and finally provide operational guidelines for metabolite name and identifier conversion to increase interoperability in metabolomics.

Methods From a list of 100 frequently annotated metabolites in human plasma, but also relevant for plant sciences, several workflows based on different existing identifier conversion tools were set up and evaluated, using two alternative approaches, one from an experimenter and the other from a data scientist's perspective.

Results Findings showed a high level of mismatches using metabolite names as input, whereas starting from identifiers showed heterogeneity in the conversion consistency, depending on the association between input identifiers, algorithm of the selected tool, their respective versions, as well as versions of databases used for mapping. Errors in cross-reference databases were also highlighted. Despite these facts, InChIKeys were found to provide the highest quality results using all identifier conversion tools.

Conclusion From these results, operational guidelines were proposed using a curation process based on computational iterations, testing the stability and consistency of this conversion process, thus guaranteeing future metabolite contextualisation (e.g. links with pathways or phenotypes) and the interoperability of result reports.

Keywords Interoperability · Metabolite · Chemical name · Identifier · Database · Curation · Reporting

1 Introduction

In metabolomics, the integration of standards and F.A.I.R. (Findable, Accessible, Interoperable, Reproducible) principles for processing and sharing complex data often remains limited today to recommendations or best practices (Wilkinson et al., 2016). Indeed, despite progress made in analytical methods and bioinformatics to automate data production, many current approaches still rely on manual curation by experts to validate annotations, select chemical names, or add identifiers for further data contextualization (e.g. links to pathways or phenotypes) (Hajjar et al., 2023). In this context, the Metabolomics Standards Initiative (MSI) has long-established recommendations for reporting metadata related

✉ Estelle Pujos-Guillot
estelle.pujos-guillot@inrae.fr

¹ Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, 63000 Clermont-Ferrand, France

² Univ. Bordeaux, INRAE, UMR1332 BFP, 33140 Villenave d'Ornon, France

³ Bordeaux Metabolome, MetaboHUB, PHENOME-EMPHASIS, 33140 Villenave d'Ornon, France

⁴ Systems biomedicine lab, Cajal Neuroscience Center, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

to metabolite identifications (Alygizakis et al., 2023; Creek et al., 2014; Sumner et al., 2007), but insufficient attention has been given to standardized metabolite nomenclature or identifiers. More recently, Koistinen et al. (Koistinen et al., 2023) published a comment to provide some recommendations for metabolite nomenclature in metabolomics reporting, but this initiative remains to be translated into operational guidelines.

Indeed, challenges in metabolite nomenclature arise from the fact that a single metabolite is referred to by multiple common names/synonyms across different scientific fields (Fahy & Subramaniam, 2020). In fact, dealing with chemical structure representation is complex, due to e.g. stereochemistry, tautomerism, protonation state, charges. This complexity is reinforced by the availability of a huge number of databases and resources (Pence & Williams, 2010) used at different steps of the metabolomics workflow, as they provide different levels of information and complementary data on chemical structures, physicochemical properties, biological functions, and pathway mapping of metabolites (Vinaixa et al., 2016). All these resources developed by several scientific communities included various semantic or chemoinformatic identifiers (Witting et al., 2024), resulting in multiple descriptions of the same features, without full interoperability. Chemoinformatic identifiers encode molecular structure for computational analysis, whereas semantic database identifiers are database-specific references without intrinsic structural information. Indeed, Akhondi et al. reported that considerable inconsistencies do exist in structural representation and systematic chemical identifiers within and between databases, in particular because of stereochemical considerations (Akhondi et al., 2012). In addition, even if the compact hash code of the IUPAC International Chemical Identifier “InChIKey” seems to be appropriate as a unique identifier since it provides information on the molecular backbone, isomer identity or isotopes of formally identified Level 1 metabolites (Sumner et al., 2007), it is not adapted to compounds with incomplete structural annotation, such as lipids, for which other alternatives are necessary (e.g., SMILES, <http://opensmiles.org/>, or ChEBI ontology) (Hajjar et al., 2023). Moreover, Pham et al. (Pham et al., 2019) identified that different name conventions between biochemical databases hinder reusability of genome-scale metabolic models (GEMs) by the community that covers different research areas ranging from systems medicine to biotechnology and plant or environment studies.

To this end, various strategies and tools were developed to achieve more interoperability. The introduction of external cross-referencing of compounds between different databases was one of the first steps (Chambers et al., 2013). A great effort has also been made for data curation including

harmonization of compound names and matching of identifiers between different databases. However, this remains specific to metabolite classes such as lipids in general (Kopczynski et al., 2024; Liebisch et al., 2020; Poupin et al., 2020; Witting et al., 2024), acylcarnitines as a specific case of lipid classes (Dambrova et al., 2022), tautomers (Sitzmann et al., 2010), and volatile organic compounds (Tevis et al., 2021). More globally, a Reference Set of Metabolite Names (RefMet) has been developed in the NIH Common Fund’s National Metabolomics Data Repository on the Metabolomics Workbench (Fahy & Subramaniam, 2020), as standardized names and links to a metabolite classification system, to be able to perform comparative analysis of metabolomics data across studies. In addition, several tools, such as ID conversion services, do exist, enabling the cross-mapping between databases. In a metabolomics workflow, the initial steps of identification and signal annotation are supported by chemical translation tools such as PubChem Identifier Exchange Service (Kim et al., 2015), CTS’s batch conversion utility (Wohlgemuth et al., 2010) or MetaboAnalyst’s Metabolite ID Conversion tool (Xia et al., 2012), taking into account chemical structures.

As the workflow progresses to biological interpretation, focusing on contextualizing metabolites within networks, a broader array of tools becomes relevant. These include: (i) RaMP (Braisted et al., 2023; Zhang et al., 2018) which facilitates the integration of metabolomic and transcriptomic data with curated pathway information; (ii) MetaCyc’s Metabolite Translation Service, a web service that translates metabolite names, identifiers, InChI strings, InChI keys, monoisotopic molecular weights, and molecular formulas, between metabolite databases; (iii) MetaNetX’s MNXref ID Mapper (Ganter et al., 2013; Moretti et al., 2016, 2021) an interactive resource for genome-scale metabolic networks that maps identifiers; and (iv) MetaFetchR (Yones et al., 2021) which links identifiers from different databases and resolves any ambiguities to help with standardization. Whether integrated into a complete mapping workflow or used independently, these tools enable the conversion of identifiers and mapping of metabolites, thereby enhancing biological insights.

Some authors proposed matching metabolites and reactions between metabolic networks by employing techniques that include approximate string matching, similarity score functions and multi-step filtering techniques, all enhanced by a set of rules based on the underlying metabolic biochemistry (Qi et al., 2014). These algorithms are implemented and available at the RECON Models site (<https://vmlife/#home>) as metabolite and reaction identification tools.

Despite the various initiatives in place, metabolomics literature continues to grapple with significant ambiguity in metabolite nomenclature. This lack of clarity hinders the

intercomparison of findings across different publications and can potentially result in misinterpretations (Koistinen et al., 2023; Witting et al., 2024).

In this context, the aim of the present work was to explore, from an operational point of view, the potential ambiguities that may be introduced during metabolite contextualization and reporting. To fulfil this objective, several workflows based on different existing identifier conversion tools were set up and evaluated, using two alternative approaches, one from an experimenter's perspective, and the other from a data scientist's perspective. From these results, the objective was to produce operational guidelines for metabolite name and identifier conversion and reporting to go towards more interoperability in metabolomics.

2 Materials and methods

2.1 Material

A list of 100 metabolites was built for this study from the intersection lists of published studies (Supplementary Table 1). These metabolites were identified as being the most frequently annotated in the human plasma (González-Domínguez et al., 2020; Liu et al., 2020) and in large-scale selected studies in medicine and epidemiology (Pietzner et al., 2021; Yu et al., 2019). Additionally, many are found in plant tissue (Lin et al., 2021; Thomason et al., 2018; Vlahakis et al., 2019; Wu et al., 2018), as well as in public databases of plant metabolites and collections of authentic standards, such as those provided by IROA Technologies (Bolton, MA, USA). This list was limited to metabolites formally identified (reported as level 1, according to MSI guidelines (Sumner et al., 2007)). The molecular weight of these compounds ranged between 75 and 777 g/mol with a polarity ranging between -4 and 8 (ALOGPS 2.1), thus leading to most frequently found variety of chemical classes (22 subclasses according to the ClassyFire computer program (Djoumbou Feunang et al., 2016), based on ChemOnt chemical taxonomy, i.e., amino acids, purines, carboxylic acids, fatty acids, steroids, hydroxy acids, keto acids, indoles etc...). Lipids identified in metabolomics are often not fully identified, explaining their relative absence in the final intersection list.

We manually curated this list in order to provide for each metabolite a consensus common name (the ChEBI name), an external database identifier (the ChEBI identifier), and a chemical identifier, the InChIKEY found on the EMBL-EBI Chemical Entities of Biological Interest (ChEBI) database. This information has been selected not only to establish a link for each metabolite to a public reference database, but also to extend the interoperability of this information so that new links can be built for each metabolite to other public

or private data repositories and databases. Chemoinformatic identifiers (e.g., InChIKey, SMILES, InChI) directly encode structural information through an algorithmic representation of atomic connectivity, stereochemistry, and tautomerism, enabling computational comparison of structures and substructure searching. Semantic database identifiers (e.g., HMDB_ID, ChEBI_ID, PubChem_CID, KEGG_ID), on the other hand, are database-specific access numbers that serve as opaque pointers to selected entries without inherent structural encoding.

2.2 Tools used for ID conversion

As previously mentioned, a wide range of tools are available for converting identifiers, but not all of them have been developed for the same purpose, which can introduce different biases mainly related inconsistency of referring to an exact match. Consequently, various criteria were applied to select the tools included in the setup workflows. Priority was given to tools that are commonly used in the metabolomics field, adhere to the F.A.I.R. principles (Wilkinson et al., 2016), with a particular focus on the Accessibility (A) of data sources, the capability for programmatic interaction (I), the availability of clear documentation (R), and the capacity to be easily used by either experimenters or data scientists. Beyond these shared attributes, additional factors were considered: (i) the ability to convert a batch of identifiers (IDs), (ii) the ease of use with no particular knowledge of IT. Supplemental Fig. 1 lists all the resulting tools considered in this study, as well as their global purposes and target users.

After the first use of all the tools, several of them were discarded to build our workflows. MetaFetchR was not selected despite its possibilities, as the installation processes were not updated for at least 3 years. Despite its online service and the possibility of batch conversion, the MetaCyc database was not selected because of a need to update the format of the cross-referenced HMDB identifiers, an observation that was raised and discussed with their support. Therefore, a total of 6 tools were included to build the alternative conversion workflows: Metabolomics Workbench RefMet tool, CTS Proxy Batch Conversion, MetaboAnalyst Metabolites ID conversion, RaMP, PubChem Identifier Exchange Service, and MetaNetX MNXref ID Mapper. The last two tools were not selected for the experimental user approach, as the output results were limited in terms of nomenclature (InChIKeys and SMILES) and not easy to handle without developing processing code. Supplemental Fig. 2 summarises inclusion and exclusion flowchart.

2.3 Description of the tested workflows

From the considered list of 100 metabolites and their metadata, two different approaches were used: one from an experimenter (wet lab), and the other from a data scientist (dry lab) perspective. These workflows are detailed in Fig. 1. The starting IDs were manually curated from cited publications, and therefore restricted to two identifiers among the most relevant ones: ChEBI IDs that are a semantic identifiers, proposed by the manually curated ChEBI database and organized within a comprehensive ontology, and InChIKeys that are chemoinformatics identifiers that links all resources. In a perspective to optimize the accuracy of the conversion step, all the conversion work was performed to require exact matches.

2.3.1 Experimenter approach

From an experimental user perspective, we focused only on the intercomparison of available metadata following data curation (see Sect. 2.1), namely ChEBI names, ChEBI IDs, and InChIKeys. We first considered ChEBI names for retrieval of IDs using the CTS ‘Batch Conversion’ and the MetaboAnalyst ‘Metabolite ID conversion’ tools (Fig. 1A). To check the conformity of output results, we compared retrieved ChEBI IDs and InChIKeys (when given, see

Fig. 1A) with corresponding ChEBI IDs and InChIKeys from our curated original list of 100 metabolites. The same comparison was repeated after standardization of nomenclatures using the ‘RefMet’ tool available on the Metabolomics Workbench repository (Fig. 1A). Second, we considered ChEBI IDs as inputs using CTS, MetaboAnalyst, and RaMP query tools, and we compared retrieved InChIKeys and nomenclatures to those of the curated original metabolite list (Fig. 1B). Finally, we considered InChIKeys as inputs using the CTS ‘Batch Conversion’ tool and we compared retrieved ChEBI IDs to those of the curated original metabolite list (Fig. 1C). The queries were run three times at separate times, on 23th April 2024, on 17th July 2024, and lastly on 8th August 2025, to check long term reproducibility of results.

2.3.2 Data scientist approach

In this approach, the intercomparison was enlarged to the conversion to external IDs that were not present within the initial metadata (see Sect. 2.1), with a perspective of using them within a metabolomics workflow that progresses to biological interpretation.

Then, the workflows described in Fig. 1B and C were also studied from a bioinformatics point of view, starting from lists of ChEBI IDs and InChIKeys only, without

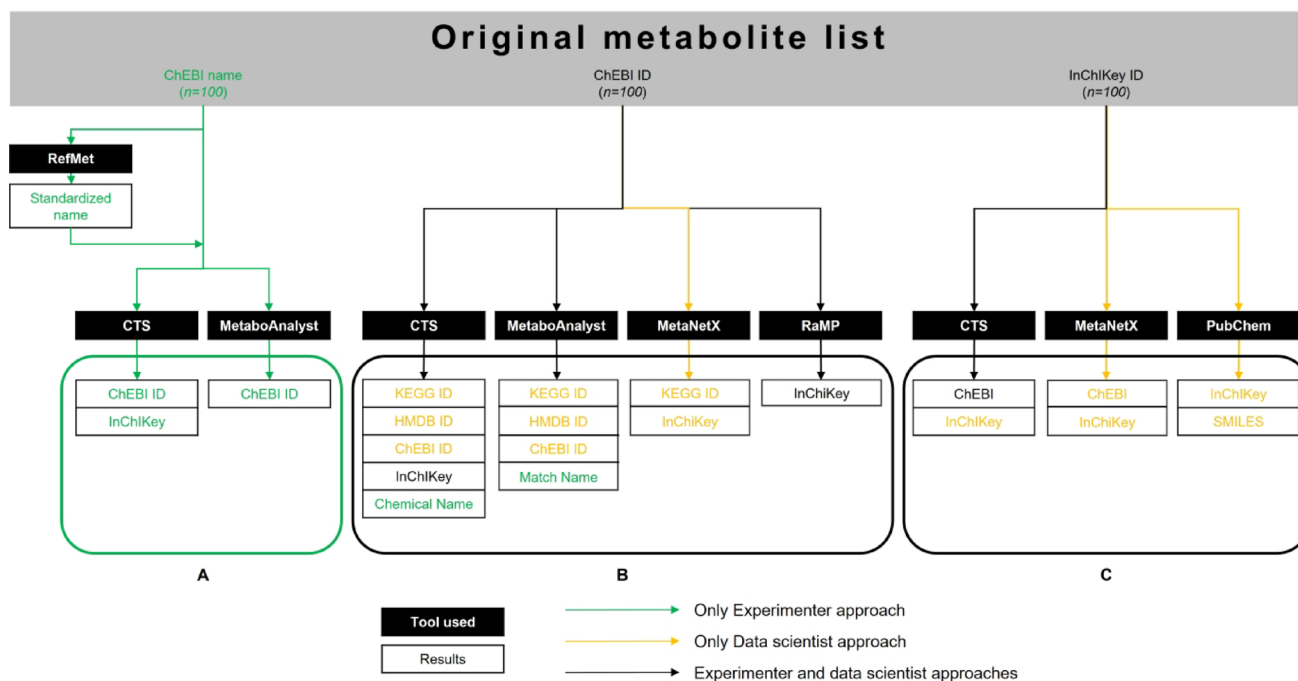


Fig. 1 Overall tested workflow. Overview of the global workflow applied in this study for metabolite ID conversion from a user or a bioinformatic approach, based on one original metabolite list. **A:** Use of ChEBI names (followed or not by RefMet standardization), **B:** Use of ChEBI IDs, and **C:** Use of InChIKeys using CTS Proxy ‘Batch

Conversion’, MetaboAnalyst ‘Metabolite ID Conversion’, MetaNetX MNXref ID Mapper, RaMP ‘Metabolites to Properties’, and PubChem ‘Identifier Exchange Service’ tools. For the sake of simplicity, only the ID that was used for the comparisons were listed as an output result

modification. Since MetaNetX is primarily designed for bioinformatics applications, it was only evaluated within this specific context.

First, to assess the effectiveness of each tool, we checked both the stability of the converted data within the same tool, and the consistency of the results from one tool to another. For this objective, the comparison of the output results was directly performed using a python script of direct comparison that checks whether the identifiers returned were the same or not. Second, for IDs not present in the initial metadata (SMILES, HMDB IDs, KEGG IDs), we examined the consistency of the results, either using the more frequent converted IDs or by building informatic loops that allowed returning to the initial IDs (see Supplemental Fig. 3). The options offered by the selected tools in terms of Input and Output IDs, and their monitoring are presented in Supplemental Table 2. Of the 6 tools selected in this study, chemoinformatic identifier formats are supported as input by only 3 of them for InChIKey and 2 for SMILES, and are provided as output by 4 tools for InChIKey and 5 for SMILES. Although these types of identifiers are considered universal, they are underused in the conversion services studied. Furthermore, when a tool accepts an identifier from a reference database as input, in 83% of cases the same type of identifier is proposed in the results. Despite the reputation and openness of the chosen databases (ChEBI, KEGG (Kyoto Encyclopedia of Genes and Genomes) and HMDB (Human metabolome database)), their identifiers are only supported by half of the tools studied. While this result can be explained by the specialisation of HMDB to the human model, this is not the case for the other resources. Nevertheless, integration of the KEGG database is subject to obtaining an academic service provider licence.

It is important to emphasize that each of these tools has a different update schedule, which can significantly impact the quality of the results and their overall usability. It is also important to differentiate between updates to the overall tool and updates to the converter - it is much more common for the overall tool to be updated without any specific updates to the conversion tool. In Supplemental Table 2, we can see that, although the global tools are regularly updated, the conversion tools are rarely affected by these updates, and this is particularly apparent when we looked at the latest database updates. These differences in database versions can lead to differences in the results obtained, and can sometimes mean that the latest metabolites added to the databases are not taken into account, which can restrict the range of additional analyses that can be carried out. However, similarly to the experimenter approach, the data were retrieved three times, on 23th April 2024, on 17th July 2024, and on 8th August 2025, and the results remained strictly identical for all the tools.

2.4 Evaluation of metabolite cross-reference quality in reference databases

To assess the quality of metabolite cross-referencing across databases, we chose to use the HMDB bank (developed by the Wishart Lab and provided by the Metabolomics Innovations Inc company) as a reference, and for each of the inventoried metabolites, to extract the associated InChIKey and the potential knapsack, ChEBI and PubChem IDs reported. In conventional metabolomic pipelines, the HMDB database provides comprehensive annotations on human metabolites, but researchers wish to propagate these identifications to metabolic pathway or metabolic network analysis tools that require other semantic database identifiers. We therefore studied the reliability of cross-references by comparing the identifiers reported by the HMDB (ChEBI, PubChem and KnapSack) to the InChIKeys in the target database. The quantification of these discrepancies aims to highlight the pitfalls of annotation when assuming that equivalent identifiers imply structural identity. This study was carried out on version V2024-05–22 of HMDB, the version of December 2024 of KnapSack and the August 2025 versions of PubChem and ChEBI. Of the 248,079 entries in HMDB, 13,820 have a ChEBI ID, 105,695 a PubChem ID and 8,021 a KnapSack ID. We then retrieved the InChIKey corresponding to the cross-referenced identifiers from the original databases and carried out a layer-by-layer comparison of the two InChIKeys, theoretically corresponding to the same molecule. This analysis was run on the Big Data-oriented ‘Metabolomics Semantic DataLake’ infrastructure (Apache Spark) and realised tasks were created using Scala programming language. The entire analysis is available via a public gist page (<https://gist.github.com/fgiacomoni/148cac13723aad07a354ca6a4bb07f4>). Supplemental Fig. 4 presents the data flow used to assess the cross-linking between metabolites reported in HMDB v2024-05–22 and their references to KnapSack, PubChem and ChEBI.

3 Results

3.1 Feedback from both users’ experience

In this section, we describe global feedback from both users’ experience. This includes, among others, the ease of use of the tools, the flexibility of input data, informatic accessibility, query duration, and information regarding the output data (see Table 1).

Results showed a global heterogeneity, mainly in terms of flexibility of input data and informatic accessibility. Conversion times were found to be appropriate, except when running the query using the chemical name as input.

Table 1 Characteristics of the used conversion tools

	CTS	Metabo Analyst	RaMP	RefMet	MetaNetX	Pub Chem
Input data						
Import data from a file	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Name	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Type of ID	chemo informatic or semantic	semantic	semantic	chemo informatic or semantic	chemo informatic or semantic	chemo informatic
Use of mixing ID sources	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Case sensitive (Short/Upper/Lower)	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	NA	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>
Input/output identifiers						
ChEBI ID	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>
HMDB ID	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>
InChIKey	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>
KEGG ID	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>
Isomeric SMILES	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> / <input checked="" type="checkbox"/>
Informatic accessibility						
API	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> (seems to work only with metabolite names)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Package	<input checked="" type="checkbox"/> (R)	<input checked="" type="checkbox"/> (R)	<input checked="" type="checkbox"/> (R)	<input checked="" type="checkbox"/> (R)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Output data						
Download format	CSV or TXT (max 500 lines)	CSV	TSV	TXT	JSON, HTML or TSV	TXT or compressed format Gzip or Bzip2
Conversion Time	8 min using names 4 min using IDs	Few seconds	A minute	Few seconds	Few seconds	Few seconds
Specificity of output data	Possibility of multiple hits A score of 0 or 1 is attributed to InChIKey values, according to confidence in the results.	Possibility of multiple hits. A score of 1 indicates a perfect match. Possibility to manually select the correct match.	NA	NA	Availability of an additional mnx_id column corresponding to MetaNetX's identifiers.	NA

Finally, the output data formats were found quite similar, despite some specificities in reporting the precision of data matching.

Following this first evaluation, the output data were selected more precisely, to be included in the present test and were: chemical name, HMDB ID, InChIKey, and ChEBI ID. Only hits with a score of 1 were considered for CTS. Concerning SMILES, one limitation of this format is that there is no standard way to generate a canonical representation (O'Boyle, 2012). Indeed, canonical SMILE is often chosen in many converter tools with the objective of aggregating database sources whereas isomeric SMILES (including geometry and chirality) is only proposed by PubChem Identifier Exchange Service (Hähnke et al., 2018).

For MetaboAnalyst's Metabolite ID Conversion tool, when using compound names as input, and none of the suggested hits were the correct match based on the comparison

of ChEBI IDs, the closest compounds in CheBI ontology was selected.

Concerning MetaNetX, a metabolite identifier first, refers to a set of metabolites in external databases (reported in a cross-referencing column, labelled 'Xref'), which are merged together because they are assumed to be the same biochemical entity. Consequently, we only selected this reference ID as result.

3.2 Results using the experimental user approach

Figure 2 gives an overview of the conversion results using the experimenter approach. The number of matches found in the different conversion tools range from 92 to 100% depending of the nature of the input data. The only case to get 100% exact matches was found using CTS and InChIKeys as input data. When looking at conversion

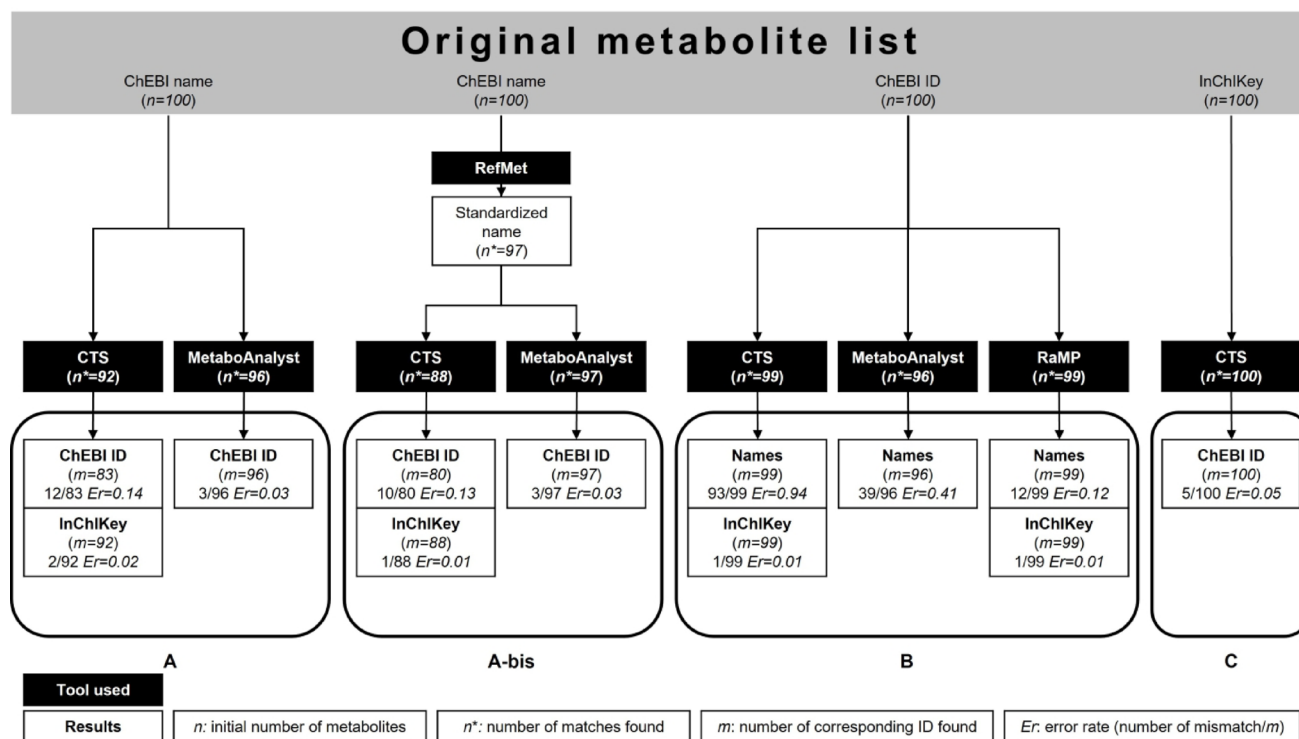


Fig. 2 Conversion results using the experimenter approach. Results of the user approach query based on the original list. **A.** ChEBI names, **A-bis.** ChEBI name followed by RefMet standardization; **B.** ChEBI

IDs; **C.** InChIKey using CTS Proxy ‘Batch Conversion’, MetaboAnalyst ‘Metabolite ID Conversion’, and RaMP ‘Metabolites to Properties’ tools

rates, they were found to be full, except using CTS with ChEBI names as input data. The error rates in conversion results were found to range from 1% to 94% highlighting the importance of choosing the adequate input associated with the various tools. Globally, all conversion into names were found to induce a high level of mismatches. Standardized names with RefMet provided similar results than using the initial ChEBI names. The best results ($\leq 5\%$ errors) were obtained either using MetaboAnalyst with ChEBI names as input, or using CTS or RaMP with ChEBI IDs as input, or using CTS with InChIKeys.

3.3 Results using the data scientist approach

In this section, the consistency of results between the different converters was first evaluated, either to retrieve semantic identifiers (HMDB, KEGG IDs), or to obtain cheminformatics identifiers (InChIKeys or SMILES). In a second step, the stability of results through informatic loops as well as the consistency of input and output when returning to the initial IDs.

Starting from the ChEBI IDs, CTS and MetaboAnalyst were found the only tools from our selection that provide the semantic identifiers (HMDB, KEGG). As illustrated in Figs. 3A and 100% of ChEBI were matches using the two tools with HMDDBs as output data, with 92% of convergence

and 8% converted only with one tool. Regarding conversion into KEGG IDs Figs. 3B and 93% of the ChEBI IDs were matched, with 2% of divergence, and 7% of conversion with only one tool. Indeed, out of the nine IDs that do not present consensus, 5 did not have results with CTS and only 2 presented a conflict between the 2 converters. To obtain cheminformatics identifiers, starting from the ChEBI IDs, CTS allowed obtaining only InChIKeys, MetaboAnalyst only SMILES, whereas MetaNetX and RAMP both types of IDs.

By retrieving the InChIKey, 100% of the ChEBI IDs were associated by all converters, and 49% of the conversion results converged in all tools (Fig. 3C). Results also revealed differences for 49% of the metabolites when using MetaNetX. Indeed, these differences mainly reflect the characteristics of this tool to ignore distinctions in protonation states (at 92% in our results), or isomer forms (at 8%) in the corresponding database, in line with the objective to facilitate the use of the results in molecular networks.

Starting from the InChIKeys, the only semantic identifiers we can convert are ChEBI IDs which can be obtained by CTS and MetaNetx. In this conversion 98% of IDs obtained at least one ChEBI ID. Regarding convergence of results, only 26.5% ($n=26$) were found identical using the two tools and 48% were specific to CTS (Fig. 3D). All results are provided in Supplemental Material 2.

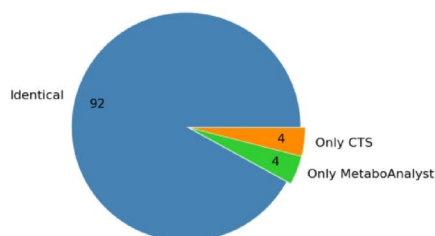
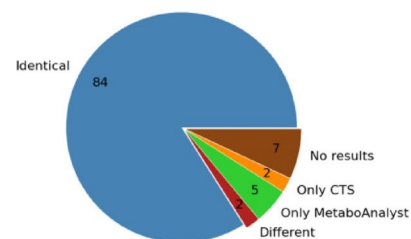
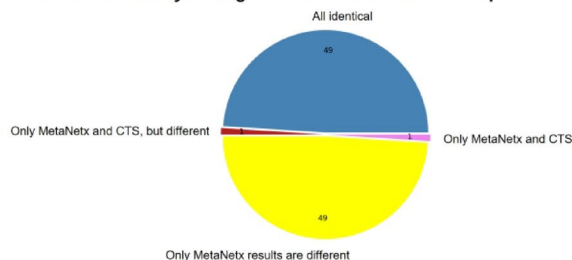
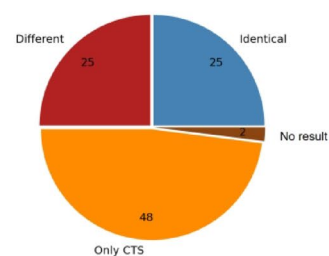
A: ChEBI to HMDB using CTS and MetaboAnalyst**B: ChEBI to KEGG using CTS and MetaboAnalyst****C: ChEBI to InChIKeys using CTS and MetaNetX and Ramp****D: InChIKey to ChEBI using CTS and MetaNetX**

Fig. 3 Results of conversions of ChEBI IDs and InChIKey IDs to different IDs. Results of ChEBI ID conversion to **A.** HMDB IDs, and **B.** KEGG IDs, using CTS and MetaboAnalyst; Results from ChEBI ID

conversion to **C.** InChIKeys; and Results of InChIKey ID conversion to **D.** ChEBI using CTS and MetaNetX

In a second step, informatic loops allowed us to explore the stability of each tool and also to go deeper onto the understanding about their internal functioning.

Globally, our results showed perfect conversions with all the tools (Supplemental Table 3), when starting from InChIKey. Concerning results using PubChem, it is important to note that the 100 InChIKeys were converted into 147 SMILES. Indeed, this can be explained by the intermediate conversion into CID, generating multiple hits. Regarding the results starting from ChEBIs, MetaboAnalyst did not present any mismatch on the loop, whereas CTS presented 3 errors and MetaNetX only 26 good matches, for the same reason than explained before.

Overall, we can see that the stability of the results is related to the algorithm used by the tool to give the results. Unfortunately, most of the time these algorithms are not available to the public.

3.4 Biases from databases

Cross-references provided across databases presented some incoherent information regarding nomenclatures, optical isomerism, stereochemistry of asymmetric carbons, and molecular structure (acid/base; zwitterionic or canonical forms, molecules with a permanent charge) in addition to a mismatch between two structurally different compounds.

Based on the InChIKeys, we were able to identify all incorrect metabolite matches in HMDB, PubChem and

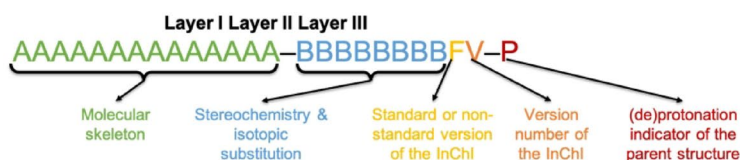
ChEBI and to categorize them into “structurally different compounds”, “isomers” or “protonation state difference”, in reference with the structure of the InChIKey layers (Fig. 4A).

All HMDB entries with ambiguous cross-links to ChEBI, PubChem, and KnapSack were identified. The percentage of mismatches was independent of the total number of cross-links provided.

From the 13,820 entries cross-referenced between HMDB and ChEBI, 4% of mismatch were found (Fig. 4B, central part). From the 105,695 HMDB entries with external reference to PubChem, 1.7% of mismatch were also found (Fig. 4B, right part). From the 8,021 HMDB entries with external reference to KnapSack, 65% of the InChIKeys mismatch (Fig. 4B, left part). Mismatching was classified according to the unmatched InChIKey layer(s): the majority of them were found related to (i) molecular skeleton or (ii) isomerism and stereochemistry (respectively, 50% and 47% in ChEBI cross-references; 27% and 72% for PubChem and 19% and 80% in KnapSack cross-references). Less than 1% for PubChem and KnapSack and 3% for ChEBI were found to be related to (de)protonation of the core parent structure.

The higher percentage of mismatch observed for the cross-referencing of HMDB to KnapSack was found to be related to the use of non-standardized InChIKey format on KnapSack. This is probably explained by the fact that is a plant-metabolite database, and some compound differences can only be translated by the on-standardized InChIKey

A. The overall length of an InChIKey is fixed to 27 characters (dashed included)



B. Effective databases mirroring (based on identifiers associations and InChIKey matching)

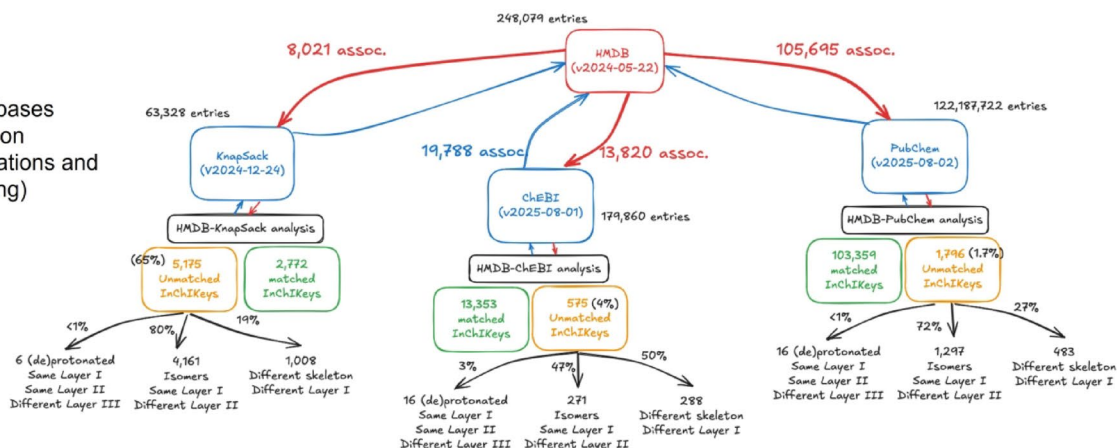


Fig. 4 InChIKey structure and HMDB cross-references to external databases. **A** Structure of the InChIKey; **B** Distribution of HMDB entries with external references to ChEBI, PubChem and KnapSack, and respective unmatched InChIKeys

format. Recently, some tools were introduced such as the InChI Trust initiative, to facilitate conversion between standard and non-standard version of InChI(Key) (<https://iupac-inchi.github.io/InChI-Web-Demo>).

Globally, structural mismatches (e.g., 575 HMDB metabolites with ChEBI Compound IDs but differing InChIKeys; Fig. 4B) highlight the need for structural validation beyond identifier matching, as uncritical use risks erroneous metabolite identification and biological misinterpretation. Users should prioritize InChIKey/SMILES canonicalization and cross-check against primary structures.

4 Discussion

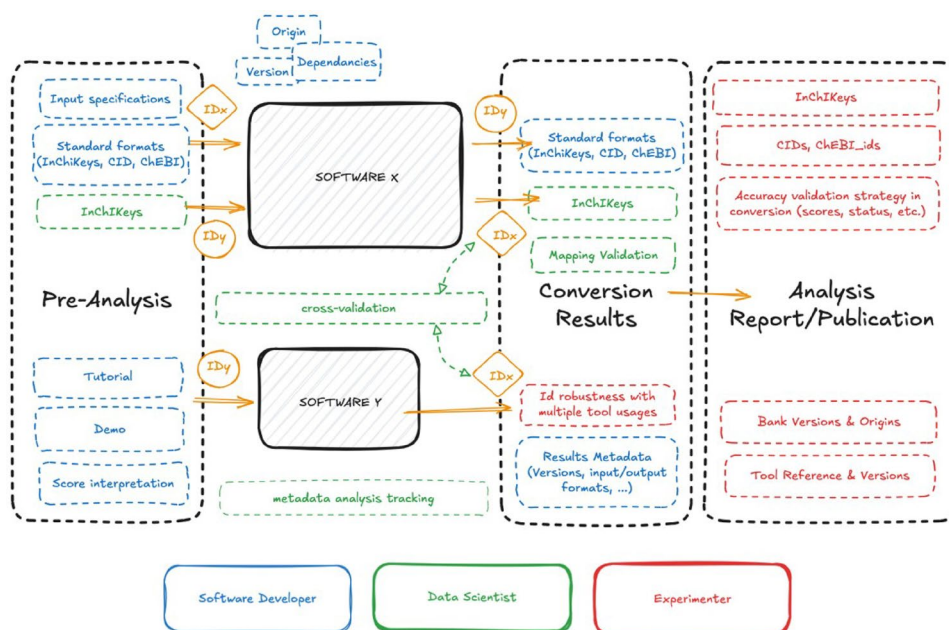
In a context of F.A.I.R principles implementation in metabolomics, the aim of the present work was to explore, from an operational point of view, the potential ambiguities that may be introduced during metabolite contextualization and reporting. The present results highlighted the multiplicity of tools and resources for conversion of metabolite names and identifiers, that are adding to the diversity in representing structure over the various scientific fields. Notably, workflows based on curated identifiers such as InChIKeys, HMDB, ChEBI, and PubChem CID have enabled greater overall coherence and reproducibility across metabolomics studies. The important point to raise to users was, in fact, the complexity of this step, part of the whole metabolomics

workflow, and the necessity of traceability and transparency for its reports.

4.1 Challenges of database/bank origins

Beyond these positive points, our study draws attention to several challenges that remain in this area. One of the main problems is the lack of transparency regarding the origin of databases or banks: publications often omit essential information about the origin of the database and the specific version used when annotating metabolites and assigning identifiers (information sometimes available under the name of in-house bank). Given that database updates, such as identifier transformations (e.g. HMDB identifiers changing from five to seven digits in version 4.0), can have a direct impact on conversion results, the omission of metadata compromises the traceability and reproducibility of annotations and makes it impossible to perform conversions. In addition, database philosophies can differ considerably: the HMDB focuses on metabolites relevant to humans, while ChEBI and PubChem cover a wider variety of isomers and chemical forms. This discrepancy between databases can lead to inconsistencies in mapping and conflicting identifications, particularly when it comes to rare or plant-specific compounds.

Fig. 5 Overview of proposed operational recommendations (minimum requirements) tailored to user type. The diagram maps each stakeholder group (e.g. Software Developer, Experimental Researcher or Data Scientist) to its corresponding set of minimum reporting items. Arrows illustrate how fulfilling these minimum requirements supports data flow across the results and publication workflow



4.2 Multiplicity of identifier formats and chemoinformatic limitations

The multiplicity of identifier formats and the ambiguity inherent in their use constitute another major challenge. Current tools often accept or generate distinct syntactic variants for the same identifier (e.g., CHEBI:16113, 16113, or chebi:16113). When input and output formats are not harmonised or do not have explicit mapping rules, interoperability between tools is compromised, and automated scripts may fail unexpectedly. Chemoinformatic identifiers such as SMILES and InChIKeys, although theoretically powerful, are not all uniformly supported by available converters. In particular, differences in the third block of InChIKeys, which reflect protonation state and isomerism, highlight the sensitivity of these identifiers to chemical context, sometimes creating new sources of discordance. These challenges require both technical vigilance and improved software documentation to ensure interoperability.

4.3 Operational validation and reporting standards

To overcome these limitations, the study emphasises operational validation and robust reporting practices. Thus, the workflows implemented must incorporate inter-database and inter-tool validation, verifying that output identifiers and associated names remain consistent using strategies such as comparing IUPAC names or chemical structures. The use of multiple conversion tools is strongly recommended, with automation via APIs to support large-scale studies and reduce manual effort. Finally, minimum reporting standards should be adopted: at least one common name

and two identifiers (one semantic and one chemoinformatic) should be provided for each metabolite, along with transparent documentation on the versions of the databases and tools used.

4.4 Solutions and proposed operational recommendations

Although only five different types of identifiers were tested in this study, there is considerable interest in disseminating operational guidelines that could establish the best way to ensure conversion stability and, therefore, conversion consistency for future biological interpretation (links to pathways or phenotypes) and reporting of metabolites. Given that the choice of an appropriate conversion tool depends strongly on the user profile (e.g., experimental researcher or data scientist) and the specific objectives pursued, we propose a set of recommendations formulated as “minimum requirements” (*) (Fig. 5) and “best practices” (**) tailored to these two groups. In addition, because the feasibility of implementing several of these recommendations is intrinsically linked to the functionalities and design of the conversion software itself, a third stakeholder group—software developers—was also considered.

4.4.1 Recommendations from the software developer perspective

- Design workflows to consistently use the chemoinformatic identifier InChIKeys and stable semantic

identifiers (ChEBI, PubChem CID) as primary exchange formats. (*)

- Ensure version transparency for used tool, package dependencies and bank(s), and if the integrated bank is a home-made solution, ensure content origin (name, version and licence of aggregated banks), and maintain normalized version numbering and detailed changelogs. (*)
- Embed in results' metadata, the tool version, database versions, and input/output ID formats actually used. (*)
- Define clear input specifications and provide example datasets showcasing valid input ID formats (e.g., uppercase vs. lowercase, prefixes like CHEBI). (*)
- Provide tutorials, demos, and if a score is computed, explicit matching score interpretation guides. (*)
- Implement format normalization routines (accept all known variants and harmonize them internally) to minimize input/output mismatches. (**)
- Support updates synchronously between conversion tool and major database/bank releases (e.g., adapt ID format during databases updates). (**)
- Offer APIs/packages that reflect the results of the online service to enable batch conversion. (**)

4.4.2 Recommendations from the data scientist perspective

- Use InChIKeys as the primary entry point in data curation and keep name-based searches secondary, clearly indicating their lower reliability; (*)
- Implement automated version tracking for each mapping task: log tool/database versions as metadata alongside the conversion results. (*)
- Add error-checking routines to detect failed mappings or mismatches, returning warnings, and document the solving strategies used when identifiers are ambiguous, multiples or unresolved. (**)
- Use multiple converters and include cross-validation pipelines across databases and tools to detect inconsistencies (aggregation step). (*)
- Cross-compare both semantic (HMDB, ChEBI IDs) and chemoinformatic identifiers (InChIKeys, SMILES) to validate mappings. (*)
- Automate queries across multiple converters via APIs to lower the manual burden on end-users. (**)
- Employ looped consistency checks: verify if the name linked to the output ID matches the original input (using InChIKey or IUPAC name if needed). (**)
- Quantify identifier stability across database versions through iterative curation cycles. (**)

4.4.3 Recommendations from the experimenter perspective

- Report at least two identifiers (semantic such as PubChem CID or ChEBI and chemoinformatic such as InChIKey) in addition to one common name in experimental records, publications and supplementary tables. (*)
- Provide database origin and version for all reported identifiers, including cases where the source is an in-house database. (*)
- Always document the conversion tool and version used (e.g., "MetaboAnalyst v6.0 identifier conversion"), including the original identifier type and format. (*)
- The accuracy of conversion results has to be reported and in case of multiple results, a classification or additional decision-making element (charge state or isotopy) should be documented as a free text in the report or with a tag system indicating clearly conflict or ambiguity in the results. (*)
- Use multiple conversion tools where possible to guarantee robustness of reported identifiers. (*)
- Prefer using identifiers generated by these tools over manual name assignments. (*)
- Follow discipline-specific curation guidelines to minimize mapping errors and support FAIR sharing. (**)
- Be cautious with plant metabolite annotation, acknowledging the under-representation in current databases and transparently reporting any limitations. (**)

5 Conclusion

To summarize, exact conversion of metabolite names/identifiers is mandatory for interoperability of metabolomics results. Despite the multiplicity of tools, this step remains complex within the workflow, due to the fact that metabolomics is intrinsically at the frontier of analytical chemistry, bioinformatics and biology, with associated challenges in metabolite nomenclature and structural representation. Traceability and transparency from the design of conversion tools to the report of their use, should limit potential ambiguities in metabolomics contextualisation and publication. Operational guidelines were proposed here, according to user profile, but automated cross-validation pipelines and consistency checks should be further developed and made available for adoption by the community involved in metabolomics. The present work was restricted to formally identified metabolites. However, a similar work on partial structures (in particular in lipidomics) would be of great interest for future perspectives.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11306-026-02404-w>.

Author contributions EPG, BC and PP conceived and supervised research. FG, BC and EPG designed research. GH, FG, MU, MC, ML, BC and EPG wrote the manuscript. GH and MU analysed data. All authors read, provided feedback, and approved the manuscript.

Funding The work was supported by the MetaboHUB French infrastructure (ANR-11-INBS-0010) and the ESFRI Research Infrastructure ELIXIR programme 2024–2028, Commissioned Service Project Plan (COS) and its WP5 (Next level of reproducible, comparable and integrable Metabolomics) – funded by the European Union’s Horizon Europe Programme under grant agreement number 101131096.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Akhondi, S. A., Kors, J. A., & Muresan, S. (2012). Consistency of systematic chemical identifiers within and between small-molecule databases. *Journal of Cheminformatics*. <https://doi.org/10.1186/1758-2946-4-35>
- Alygizakis, N., Lestremau, F., Gago-Ferrero, P., Gil-Solsona, R., Arturi, K., Hollender, J., Schymanski, E. L., Dulio, V., Slobodnik, J., & Thomaidis, N. S. (2023). Towards a harmonized identification scoring system in LC-HRMS/MS based non-target screening (NTS) of emerging contaminants. *TrAC - Trends in Analytical Chemistry*. <https://doi.org/10.1016/j.trac.2023.116944>
- Braisted, J., Patt, A., Tindall, C., Sheils, T., Neyra, J., Spencer, K., Eicher, T., & Mathé, E. A. (2023). RaMP-DB 2.0: A renovated knowledgebase for deriving biological and chemical insight from metabolites, proteins, and genes. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btac726>
- Chambers, J., Davies, M., Gaulton, A., Hersey, A., Velankar, S., Petryszak, R., Hastings, J., Bellis, L., McGlinchey, S., & Overington, J. P. (2013). UniChem: A unified chemical structure cross-referencing and identifier tracking system. *Journal of Cheminformatics*. <https://doi.org/10.1186/1758-2946-5-3>
- Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E. L., Sumner, L. W., Trengove, R., & Wolfender, J. L. (2014). Metabolite identification: Are you sure? And how do your peers gauge your confidence? *Metabolomics*, *10*(3), 350–353. <https://doi.org/10.1007/s11306-014-0656-8>
- Dambrova, M., Makrečka-Kuka, M., Kuka, J., Vilskersts, R., Nordberg, D., Attwood, M. M., Smesny, S., Sen, Z. D., Guo, A. C., Oler, E., Tian, S., Zheng, J., Wishart, D. S., Liepinsh, E., & Schiöth, H. B. (2022). Acylcarnitines: Nomenclature, biomarkers, therapeutic potential, drug targets, and clinical trials. *Pharmacological Reviews*, *74*(3), 506–551. <https://doi.org/10.1124/p harmrev.121.000408>
- Djombou Feunang, Y., Eisner, R., Knox, C., et al. (2016). ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics*, *8*, Article 61. <https://doi.org/10.1186/s13321-016-0174-y>
- Fahy, E., & Subramaniam, S. (2020). RefMet: A reference nomenclature for metabolomics. *Nature Methods*, *17*(12), 1173–1174. <https://doi.org/10.1038/s41592-020-01009-y>
- Ganter, M., Bernard, T., Moretti, S., Stelling, & Pagni, M. (2013). MetaNetX.org: A website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics*, *29*(6), 815–816. <https://doi.org/10.1093/bioinformatics/btt036>
- González-Domínguez, R., Jáuregui, O., Queipo-Ortuño, M. I., & Andrés-Lacueva, C. (2020). Characterization of the human exposome by a comprehensive and quantitative large-scale multi-analyte metabolomics platform. *Analytical Chemistry*, *92*(20), 13767–13775. <https://doi.org/10.1021/acs.analchem.0c02008>
- Hähnke, V. D., Kim, S., & Bolton, E. E. (2018). PubChem chemical structure standardization. *Journal of Cheminformatics*, *10*, Article 36. <https://doi.org/10.1186/s13321-018-0293-8>
- Hajjar, G., Barros Santos, M. C., Bertrand-Michel, J., Canlet, C., Castelli, F., Creusot, N., Dechaumet, S., Diémé, B., Giacomoni, F., Giraudeau, P., Guitton, Y., Thévenot, E., Tremblay-Franco, M., Junot, C., Jourdan, F., Fenaille, F., Comte, B., Pétriacq, P., & Pujos-Guillot, E. (2023). Scaling-up metabolomics: Current state and perspectives. *TrAC - Trends in Analytical Chemistry*. <https://doi.org/10.1016/j.trac.2023.117225>
- Kim, S., Thiessen, P. A., Bolton, E. E., & Bryant, S. H. (2015). PUG-SOAP and PUG-REST: Web services for programmatic access to chemical information in PubChem. *Nucleic Acids Research*, *43*(W1), W605–W611. <https://doi.org/10.1093/nar/gkv396>
- Koistinen, V., Kärkkäinen, O., Keski-Rahkonen, P., Tsugawa, H., Scalbert, A., Arita, M., Wishart, D., & Hanhineva, K. (2023). Towards a Rosetta stone for metabolomics: Recommendations to overcome inconsistent metabolite nomenclature. *Nature Metabolism*, *5*(3), 351–354. <https://doi.org/10.1038/s42255-023-00757-3>
- Kopczynski, D., Ejsing, C. S., McDonald, J. G., Bamba, T., Baker, E. S., Bertrand-Michel, J., Brügger, B., Coman, C., Ellis, S. R., Garrett, T. J., Griffiths, W. J., Guan, X. L., Han, X., Höring, M., Holčápek, M., Hoffmann, N., Huynh, K., Lehmann, R., Jones, J. W., ... Liebisch, G. (2024). The lipidomics reporting checklist a framework for transparency of lipidomic experiments and repurposing resource data. *Journal of Lipid Research*. <https://doi.org/10.1016/j.jlr.2024.100621>
- Liebisch, G., Fahy, E., Aoki, J., Dennis, E. A., Durand, T., Ejsing, C. S., Fedorova, M., Feussner, I., Griffiths, W. J., Köfeler, H., Merrill, A. H., Murphy, R. C., O’Donnell, V. B., Oskolkova, O., Subramaniam, S., Wakelam, M. J. O., & Spener, F. (2020). Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *Journal of Lipid Research*, *61*(12), 1539–1555. <https://doi.org/10.1194/jlr.S120001025>
- Lin, X., Gao, H., Ding, Z., Zhan, R., Zhou, Z., & Ming, J. (2021). Comparative metabolic profiling in pulp and peel of green and red pitayas (*Hylocereus polyrhizus* and *Hylocereus undatus*) reveals

- potential valorization in the pharmaceutical and food industries. *BioMed Research International*. <https://doi.org/10.1155/2021/6546170>
- Liu, K. H., Nellis, M., Uppal, K., Ma, C., Tran, V. L., Liang, Y., Walker, D. I., & Jones, D. P. (2020). Reference standardization for quantification and harmonization of large-scale metabolomics. *Analytical Chemistry*, 92(13), 8836–8844. <https://doi.org/10.1021/acs.analchem.0c00338>
- Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A., & Pagni, M. (2016). MetaNetX/MNXref - Reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research*, 44(D1), D523–D526. <https://doi.org/10.1093/nar/gkv1117>
- Moretti, S., Tran, V. D. T., Mehl, F., Ibberson, M., & Pagni, M. (2021). MetaNetX/MNXref: Unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Research*, 49(D1), D570–D574. <https://doi.org/10.1093/nar/gkaa992>
- O’Boyle, N. M. (2012). Towards a universal SMILES representation - A standard method to generate canonical SMILES based on the InChI. *Journal of Cheminformatics*, 4, Article 22. <https://doi.org/10.1186/1758-2946-4-22>
- Pence, H. E., & Williams, A. (2010). Chemspider: An online chemical information resource. *Journal of Chemical Education*, 87(11), 1123–1124. <https://doi.org/10.1021/ed100697w>
- Pham, N., van Heck, R. G. A., van Dam, J. C. J., Schaap, P. J., Saccenti, E., & Suarez-Diez, M. (2019). Consistency, inconsistency, and ambiguity of metabolite names in biochemical databases used for genome-scale metabolic modelling. *Metabolites*. <https://doi.org/10.3390/metabo9020028>
- Pietzner, M., Stewart, I. D., Raffler, J., Khaw, K. T., Michelotti, G. A., Kastenmüller, G., Wareham, N. J., & Langenberg, C. (2021). Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nature Medicine*, 27(3), 471–479. <https://doi.org/10.1038/s41591-021-01266-0>
- Poupin, N., Vinson, F., Moreau, A., Batut, A., Chazalviel, M., Colsch, B., Fouillen, L., Guez, S., Khoury, S., Dalloux-Chioccioli, J., Tournadre, A., Le Faouder, P., Pouyet, C., Van Delft, P., Viars, F., Bertrand-Michel, J., & Jourdan, F. (2020). Improving lipid mapping in genome scale metabolic networks using ontologies. *Metabolomics*. <https://doi.org/10.1007/s11306-020-01663-5>
- Qi, X., Ozsoyoglu, Z. M., & Ozsoyoglu, G. (2014). Matching metabolites and reactions in different metabolic networks. *Methods*, 69(3), 282–297. <https://doi.org/10.1016/j.ymeth.2014.06.007>
- Sitzmann, M., Ihlenfeldt, W. D., & Nicklaus, M. C. (2010). Tautomerism in large databases. *Journal of Computer-Aided Molecular Design*, 24(6–7), 521–551. <https://doi.org/10.1007/s10822-010-9346-4>
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., & Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Tevis, D. S., Flores, S. R., Kenwood, B. M., Bhandari, D., Jacob, P., Liu, J., Lorkiewicz, P. K., Conklin, D. J., Hecht, S. S., Goniewicz, M. L., Blount, B. C., & De Jesús, V. R. (2021). Harmonization of acronyms for volatile organic compound metabolites using a standardized naming system. *International Journal of Hygiene and Environmental Health*. <https://doi.org/10.1016/j.ijheh.2021.113749>
- Thomason, K., Babar, M. A., Erickson, J. E., Mulvaney, M., Beecher, C., & MacDonald, G. (2018). Comparative physiological and metabolomics analysis of wheat (*Triticum aestivum* L.) following post-anthesis heat stress. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0197919>
- Vinaixa, M., Schymanski, E. L., Neumann, S., Navarro, M., Salek, R. M., & Yanes, O. (2016). Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *TrAC - Trends in Analytical Chemistry*, 78, 23–35. <https://doi.org/10.1016/j.trac.2015.09.005>
- Vlahakis, C., Hazebroek, J., Beecher, C., & de Jong, F. (2019). Isotopic ratio outlier analysis improves metabolomics prediction of nitrogen treatment in maize. *Phytochemistry*, 164, 130–135. <https://doi.org/10.1016/j.phytochem.2019.05.006>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). Comment: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. <https://doi.org/10.1038/sdata.2016.18>
- Witting, M., Malik, A., Leach, A., Bridge, A., Aimo, L., Conroy, M. J., O’Donnell, V. B., Hoffmann, N., Kopczynski, D., Giacomoni, F., Paulhe, N., Gassiot, A. C., Poupin, N., Jourdan, F., & Bertrand-Michel, J. (2024). Challenges and perspectives for naming lipids in the context of lipidomics. *Metabolomics*. <https://doi.org/10.1007/s11306-023-02075-x>
- Wohlgenuth, G., Haldiya, P. K., Willighagen, E., Kind, T., & Fiehn, O. (2010). The chemical translation service—a web-based tool to improve standardization of metabolomic reports. *Bioinformatics*, 26(20), 2647–2648. <https://doi.org/10.1093/bioinformatics/btq476>
- Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., Méret, M., Keurentjes, J. B., Nikoloski, Z., Fernie, A. R., Willmitzer, L., & Brotman, Y. (2018). Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. *Molecular Plant*, 11(1), 118–134. <https://doi.org/10.1016/j.molp.2017.08.012>
- Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D., & Wishart, D. S. (2012). MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gks374>
- Yones, S. A., Csombordi, R., Komorowski, J., & Diamanti, K. (2021). Metafetcher: An r package for complete mapping of small-compound data. *Metabolites*. <https://doi.org/10.3390/metabo11110743>
- Yu, B., Zanetti, K. A., Temprosa, M., Albanes, D., Appel, N., Barrera, C. B., Ben-Shlomo, Y., Boerwinkle, E., Casas, J. P., Clish, C., Dale, C., Dehghan, A., Derkach, A., Eliassen, A. H., Elliott, P., Fahy, E., Gieger, C., Gunter, M. J., Harada, S., ... Moore, S. C. (2019). The Consortium of Metabolomics Studies (COMETS): Metabolomics in 47 prospective cohort studies. *American Journal of Epidemiology*, 188(6), 991–1012. <https://doi.org/10.1093/aje/kwz028>
- Zhang, B., Hu, S., Baskin, E., Patt, A., Siddiqui, J. K., & Mathé, E. A. (2018). RaMP: A comprehensive relational database of metabolomics pathways for pathway enrichment analysis of genes and metabolites. *Metabolites*. <https://doi.org/10.3390/metabo8010016>