diseases

MDPI

Review

Tracking the Metabolites of Health and Disease Using Artificial Intelligence

Ahmed Fadiel [1,*], Kenneth D. Eichenbaum [2,3], Aya Hassouneh [4] and Kunle Odunsi [5,6]

[1] Computational Oncology Unit, University of Chicago Medicine Comprehensive Cancer Center, 900 E 57th St, KCBD Bldg., Chicago, IL 60637, USA
[2] Department of Anesthesiology, Oakland University William Beaumont School of Medicine, Rochester, MI 48309, USA; keichenbaum@wayne.edu
[3] Department of Anesthesiology, Wayne State University, Detroit, MI 48341, USA
[4] Electrical and Computer Engineering, Western Michigan University, 1903 W. Michigan Ave., Kalamazoo, MI 49008, USA; aia.hassouneh@gmail.com
[5] University of Chicago Medicine Comprehensive Cancer Center, 5841 South Maryland Avenue, MC1140, Chicago, IL 60637, USA; odunsia@uchicago.edu
[6] Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL 60637, USA
* Correspondence: afadiel01@uchicago.edu

Abstract

Using AI to analyze metabolite profiles provides critical insights into health, aging, and disease. Metabolomic signatures reveal how lifestyle and therapy impact organ function and cancer progression. This review highlights emerging toolkits for high-throughput data analysis, emphasizing their integration with other omics. Advanced AI approaches facilitate metabolic pathway mapping and accelerate biomarker discovery. By combining AI with multi-omics, researchers can optimize interventions and enhance precision medicine. This article serves as a resource demonstrating AI's potential in diagnostics and drug discovery.

Keywords: metabolomics; artificial intelligence; multi-omics; cancer; bioinformatics; drug development pipelines

1. Introduction & Background

Metabolomics is the study of how cells process materials into end products. It involves the systematic analysis of molecular intermediates and small products weighing less than 1500 Da[1]. Generating comprehensive datasets allows machine learning to identify molecular patterns in health and disease states. This review provides an overview of the workflow engines used in translational research.

A wealth of information can be collected and accessed to characterize further key information housed inside the metabolome for targeted and untargeted data. Global untargeted data harvesting can be designed to identify as many metabolites as possible within a given matrix [1–4]. Over time, by tracking metabolite profiles, distinct organ systems can be evaluated, and a composite picture using biomarkers that can provide information about which exposures impact and improve or harm an organ system. With advances in instrumentation and analytical toolkits, it is now possible to develop metabolomic signature patterns that describe physiological and pathological conditions across the spectrum from healthy individuals to patients with advanced disease. While LC–MS is the most commonly used platform in AI-driven metabolomics studies due to its broad analytical versatility and sensitivity across a range of polar and non-polar compounds, other platforms such as GC–MS and CE–MS offer complementary advantages that are underutilized

Check for updates

Academic Editors: Charat Thongprayoon and Omar Cauli

Received: 28 November 2025
Revised: 30 January 2026
Accepted: 15 February 2026
Published: 25 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

in many AI workflows. LC–MS has become widely adopted because it can separate and detect a wide spectrum of chemical species without requiring chemical derivatization. It also provides high sensitivity in complex biological matrices, making it suitable for many metabolomics applications [5]. In contrast, GC–MS excels at separating and identifying volatile and thermally stable metabolites, often with extensive spectral libraries and reproducible fragmentation patterns, thereby enhancing confidence in the identification of small molecules [6]. Capillary electrophoresis–mass spectrometry (CE-MS) provides high-resolution separation for charged and highly polar compounds that are difficult to retain or separate on conventional LC columns, making it particularly effective for amino acids, nucleotides, and other ionogenic metabolites [7]. Each of these platforms also presents unique analytical challenges that impact downstream AI/ML analyses. For example, LC–MS is susceptible to matrix effects and ion suppression arising from co-eluting endogenous compounds, which can alter ionization efficiency and complicate quantitation [8]. These technical factors can introduce non-biological variability that machine learning models may misinterpret as biological signal if not properly controlled. A critical understanding of these platform-specific limitations is essential for interpreting metabolomic patterns accurately and for designing AI-driven workflows that are both robust and biologically meaningful.

Importantly, these metabolomic signatures can inform drug discovery and development. By identifying disease-specific metabolites and altered pathways, researchers can uncover potential drug targets, evaluate the biochemical effects of candidate therapeutics, and prioritize compounds for further development. Integrating metabolomic data with AI and machine learning enhances the ability to predict drug efficacy, potential toxicity, and mechanisms of action, thereby accelerating the translation of basic metabolic insights into therapeutic interventions.

While high-dimensional metabolomics datasets combined with AI/ML offer powerful pattern recognition and discovery capabilities, translating these approaches into clinical or real-world settings remains challenging. Standardization and reproducibility issues, including differences in instrumentation, sample handling, and analytical protocols, can lead to inconsistent metabolite identification, quantitation, and poor reproducibility across studies, which limit clinical utility and biomarker validation efforts [9]. Comprehensive metabolome analysis often relies on expensive, high-end instruments (e.g., LC–MS) with complex workflows that require significant technical expertise, while routine clinical laboratories typically depend on robust, targeted assays with simpler instrumentation and standardized protocols [10]. Moreover, the lack of broadly accepted best practices for untargeted metabolomics, variability in study design and data processing, and limited validation across diverse cohorts present additional barriers to implementation in precision medicine and clinical diagnostics [11]. Addressing these practical and translational considerations, which include cost, reproducibility, standardization, and validation, can help achieve the full potential of AI-driven metabolomics in precision medicine and diagnostic applications.

Ultimately, tracking metabolomic changes can serve as a dynamic monitor for organ wellness, systemic health, and treatment response. Over time, applying artificial intelligence (AI) to these molecular patterns not only improves disease understanding but also supports the discovery and optimization of new therapies.

Recent Advances in AI-Driven Metabolic Pathway Modeling

Recent years have witnessed a rapid expansion of deep learning methodologies for metabolomics data interpretation, particularly in the prediction of metabolic pathways and flux distributions [12]. Since 2023, transformer-based architectures, graph neural networks (GNNs), and variational autoencoders (VAEs) have been increasingly applied to model complex biochemical reaction networks and to infer pathway activity directly from high-

dimensional metabolomic profiles. These approaches enable the identification of latent metabolic states and improve the resolution of pathway-level biomarker discovery beyond traditional multivariate statistics. Graph neural networks have shown particular promise for pathway inference by representing metabolites as nodes and enzymatic reactions as edges, thereby allowing biologically constrained learning of metabolic network topology and activity [13,14]. Recent studies have demonstrated that GNN-based metabolomics models can accurately predict pathway dysregulation in cancer metabolism, mitochondrial disorders, and inflammatory diseases, outperforming conventional enrichment-based approaches. Transformer models have also been adapted for metabolomics to capture long-range dependencies among metabolites and to predict metabolic flux alterations under pharmacological or nutritional perturbations.

Importantly, these deep learning strategies are increasingly being integrated into drug discovery pipelines, enabling in silico prioritization of metabolic enzymes as therapeutic targets, prediction of off-target metabolic liabilities, and patient stratification based on pathway activity signatures. The integration of AI-based pathway modeling with multi-omics data is thus accelerating translational research by bridging metabolomic pattern recognition with mechanistic interpretation and therapeutic development.

In this review, we apply a hierarchical framework for computational analysis: Artificial Intelligence (AI) serves as the broad field of systems mimicking human intelligence; Machine Learning (ML) refers to the specific algorithms that learn from embedded data to identify patterns without explicit programming; and Deep Learning (DL) is a specialized subset of ML that utilizes multi-layered neural networks to manage the most complex, high-dimensional datasets.

2. Metabolomics Landscape and Foundational Research

There are many emerging opportunities for metabolomics to investigate pathological disorders in the context of small molecules. This growing field, like proteomics and genomics, can stitch together different molecular cascades that help characterize how metabolites form and are utilized in complex biological systems and disease processes. Wishart's review [15] discussed several recent technological developments in metabolomics, emphasizing how to use metabolomics to find the root causes of complicated disorders. Pang and coworkers [16] highlighted developments in metabolomics technology and methodologies and how they apply to the study of clinical pharmacology. They addressed many of the field's present difficulties and prospective future paths. Yang and coworkers highlighted new applications of metabolomics biotechnology in revealing metabolic anomalies and related underlying disease mechanisms [17]. The application of metabolomics to cancer and chemotherapy assessment was reported by Gao et al. [18]. They highlighted the importance of metabolomics in mass spectrometry applications and in clinical and laboratory gastrointestinal cancer investigations. Tumor growth can stem from both interwoven epigenetic and metabolomic changes. Metabolites have experimentally been shown to directly modulate chromatin factors and epigenetic modifications alter transcriptional regulation of metabolic enzymes leading to metabolic reprogramming [19,20].

Fu et al. [21] performed a metabolomics analysis to review the study findings on the interventional effects of functional meals on metabolic disorders. These analyses demonstrate the wide range of applications for metabolomics that include animal model preclinical studies, drug screening, drug efficacy and toxicity studies, and comprehensive clinical evaluation. Collectively, these studies illustrate how metabolomics can inform drug screening, evaluate drug efficacy and toxicity, and guide the development of novel therapeutics.

Our understanding of the mechanisms of molecules linked to many metabolic illnesses is at an early stage. Targeting precise downstream effects of functional biologics,

genetic targeting, and dietary interventions on metabolic illnesses can be enhanced using metabolomics with other "omics" technologies. The swift development of metabolomic techniques has made it possible to quickly detect and identify tiny molecules in biological and environmental materials. These tools are increasingly capable of uncovering novel metabolites and pathways that can serve as potential drug targets, inform mechanism-of-action studies, and accelerate therapeutic discovery pipelines. Kaddurah-Daouk and co-workers [22] emphasized the use of metabolomics in the analysis of nervous system disorders. They share ideas on metabolomics, metabolic profiling technologies, preliminary research on neuropsychiatric disorders, and medications used to treat these disorders. Increasing usage of metabolomics in medicine facilitates the identification of biomarkers for diagnosing illness and treatment, and therapeutic intervention design, as discussed by the review of Aderemi and coworkers [23]. Recently, Abdelhamid and coworkers discussed a machine learning algorithm analysis of 8500 biomarkers using metabolomics, proteomics, and lipidomic to predict survival outcomes in trauma patients [24], demonstrating how AI-driven multi-omics analysis can prioritize targets and guide drug development strategies.

3. Inside the Metabolomics Workflow Engine: Software Tools

Software tools designed specifically for metabolomics data are rapidly being developed, with hundreds of tools already available in the literature. Many of these tools are open-source and freely available. Metabolomic tools offer the advantages of being fast, cheap, and sensitive. However, the data generated by these tools mean little unless they are adequately analyzed for the construction of biochemical pathways and for understanding how they interact in both diseased and non-diseased states. A full Metabolomic path encompasses the workflow process spanning preprocessing, data collection, and postprocessing and is illustrated in Figure 1. Initially, the biological sample is processed, in which molecules of interest are extracted and any interfering substances are removed. Adaptive sampling to determine improved resolution for scanning and optimized LC-MS settings can be introduced with AI techniques. Raw data must be converted to a specific format, often open source, to be accepted by the platform tools. Then, the data will be processed using those tools to find the Peak, or unique identifying feature, that stands out from background noise and is then aligned and normalized. Alignment allows for matching the same molecule across datasets. The samples can be corrected for unwanted variation, adjusting for concentration and setting up internal standards with quality control tools. Software can compare peaks against external datasets. In-silico prediction can be applied to identified peaks. Once the Dataset is organized, principal component analysis can be used to see which data clusters are found together with the help of graphical neural networks. Additional analysis can identify changes and pathological aberrancies with increases and decreases in certain molecules.

Commercial metabolomics platforms have transformed mass spectrometry-based metabolomics workflows by integrating data-processing algorithms into vendor-specific, user-friendly software environments [4,15]. Platforms such as MassProfiler Professional, Compound Discoverer, PeakView, MarkerLynx, MetaboScape, and Profiler AM+ (Table 1) deliver end-to-end solutions for handling raw data, peak detection, feature alignment, normalization, statistical modeling, and metabolite annotation [25,26]. These systems have modernized metabolomics capabilities for laboratories lacking dedicated bioinformatics infrastructure [27].

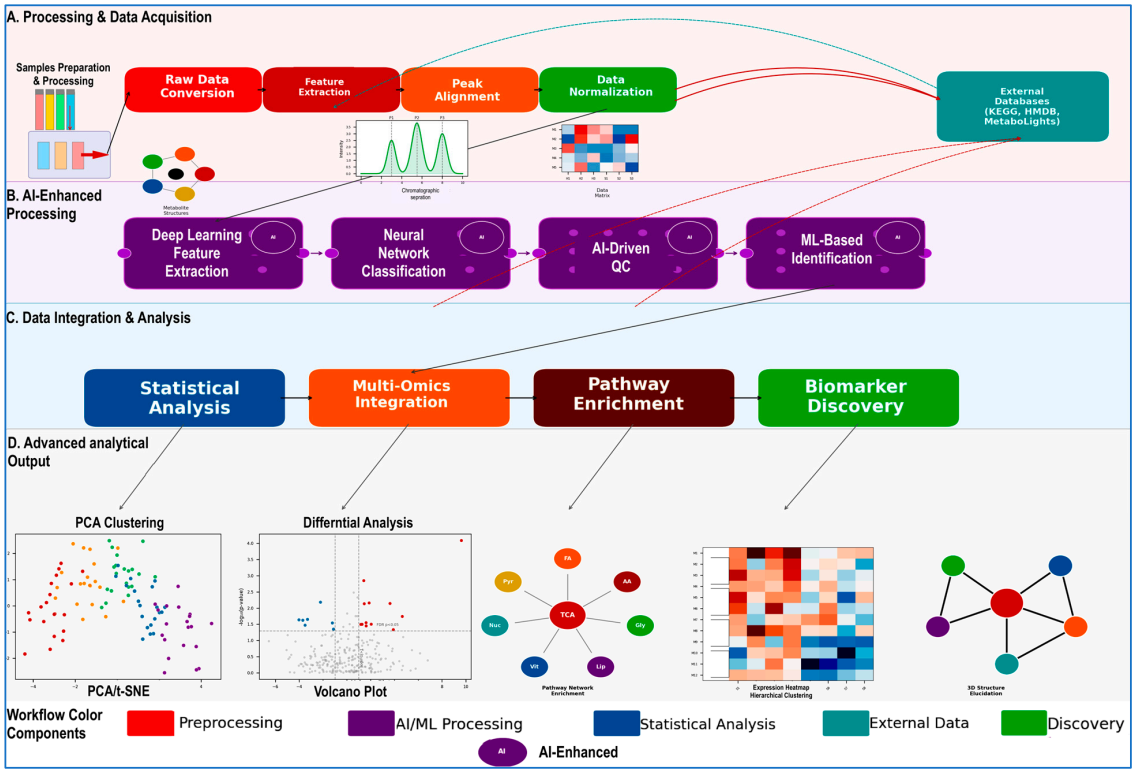


Figure 1. Integrated metabolomics analysis pipeline combining conventional preprocessing with artificial intelligence (AI) and machine learning (ML)–assisted analysis. (**A**) Preprocessing and data acquisition: Biological samples are extracted and analyzed by liquid chromatography–tandem mass spectrometry (LC–MS/MS). Chromatographic separation generates retention time–resolved ion signals that are converted into structured data matrices (m/z–retention time features). Subsequent steps include raw data conversion, feature extraction (peak detection), retention time alignment, and normalization. External databases such as HMDB, KEGG, and MetaboLights support metabolite annotation and biological interpretation. (**B**) AI/ML-assisted processing layer: This layer comprises four interconnected computational modules (purple boxes): representation learning (e.g., deep learning–based feature extraction), classification (e.g., supervised models), data quality control that includes outlier detection and batch correction, and ML-assisted metabolite annotation. These modules enhance robustness and scalability while supporting reproducible downstream analysis. (**C**) Data integration and analysis: Statistical analysis includes univariate and multivariate approaches with appropriate multiple-testing correction (e.g., false discovery rate [FDR]). Data integration incorporates correlation networks and data fusion across complementary datasets such as transcriptomics or proteomics where available. Pathway enrichment analysis uses curated resources that may include KEGG, Reactome, and WikiPathways to contextualize metabolic alterations. Biomarker discovery applies feature selection and receiver operating characteristic (ROC) analysis with appropriate validation strategies. Red arrows indicate external database queries and iterative annotation refinement. (**D**) Analytical outputs and disease-relevant interpretation: Outputs include dimensionality reduction such as principal component analysis [PCA] and t-distributed stochastic neighbor embedding [t-SNE] as well as volcano plots for differential abundance analysis (FDR-adjusted $p < 0.05$; $|\log_2 \text{fold change}| > 1$). Pathway network visualization highlights the central role of the tricarboxylic acid (TCA) cycle and its connections to glycolysis, amino acid metabolism, lipid metabolism, nucleotide metabolism, and cofactor/vitamin pathways. Additional outputs include hierarchical clustering heatmaps and molecular structure annotation or elucidation that are supported by spectral evidence. AI/ML-assisted components improve analytical scalability and support reproducible identification of disease-associated metabolic signatures while maintaining rigorous statistical control. Color coding: red/coral (preprocessing), purple (AI/ML processing), blue (statistical analysis), orange (data integration), brown (pathway analysis), green (biomarker discovery), teal (external databases).

The principal strength of commercial platforms lies in their seamless integration with specific instrument ecosystems, enabling management of large-scale datasets with stable file-format support and access to continuously updated spectral libraries and vendor technical assistance [28,29]. Guided workflows and intuitive graphical interfaces facilitate standardization and implementation of quality-control procedures, which are requisite for high-throughput studies and regulated research environments [30]. These features are valuable in clinical metabolomics and pharmaceutical applications where reproducibility and regulatory compliance are mandated [31,32].

There are several limitations that constrain universal applicability for commercial platforms. Licensing and maintenance costs can be onerous and may impede access in academic and developing-world settings [33,34]. The proprietary nature of underlying algorithms and data structures presents challenges for transparency and independent validation, often complicating efforts to reproduce published analytical workflows [35,36]. Vendor-specific data formats create "data lock-in" scenarios that impede integration with outside software ecosystems and obstruct long-term cross-platform analysis or multi-omics integration [37,38]. For custom experimental designs, limited flexibility may prevent implementation of bespoke statistical models or cutting-edge machine-learning approaches [39,40].

Some metabolomics laboratories have adopted hybrid analytical strategies that involve using commercial platforms for data acquisition, initial preprocessing, and quality control, while exporting curated feature tables and spectral data to open-source environments [41,42]. This allows for transparent, extensible downstream analysis. Balancing the reliability of commercial tools with the algorithmic transparency and flexibility of community-developed software enables reproducibility, interoperability, and sustainable data stewardship in metabolomics research [43,44].

The approach to generating data with metabolomics requires a preprocessing assessment of which platform and database tools (Tables 2 and 3) should be used: NMR-based, MS-based, or FT-IR Spectroscopy-based. A recently designed software website: MSCAT (Metabolomics Software CATalog) database of metabolomics software tools, provides an overview of available tools and assists researchers in choosing a data analysis workflow for metabolomics studies according to their specific needs [45]. MSCAT's database can be used to identify the output of the preprocessing tools. There are several approaches to data harvesting in metabolomics, using either mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, or Fourier-transform infrared (FTIR) spectroscopy techniques. Examples of such approaches include metabolomic fingerprinting, metabolic profiling, metabolic foot printing, target analysis, and flux analysis, each playing significant roles in understanding toxicological mechanisms and disease processes in living organisms and identifying potential therapeutic targets for drug discovery. Platform technologies such as Nuclear Magnetic Resonance have been used specifically in applications to detect novel chemicals and clarify structures. Mass Spectrometry allows for highly sensitive separation techniques and is becoming more widely used.

Table 4 illustrates the advantages and disadvantages of standard metabolomics tools. Available software tools are listed in sequential order in characterizing a metabolomics study. First metabolomic tools are used on samples to obtain data and results are generated by the instruments. Then, output data is preprocessed (Table 5) and investigated using quality control tools (Table 6). Subsequently, annotation tools (Table 7), are used to define the success and failure of untargeted metabolomics. Finally, statistical and visualization tools (Table 8) are used to plot metabolite data sets, convert, and clean up the data. Additionally, these instruments conduct data normalization for sample labeling and list metabolite names within the experimental conditions. These workflows can also be integrated with AI to

accelerate biomarker discovery and identify druggable metabolic pathways, bridging basic metabolomics insights to therapeutic development (Tables 9 and 10).

The software tools detailed in Tables 1–9 do not function in isolation but comprise a synchronized 'metabolomics workflow engine.' For example, the MSCAT database serves as more than a catalog; it is an intelligent assistant that helps researchers choose a specific data analysis workflow tailored to the unique sensitivities of MS or NMR platforms. By integrating these tools, researchers can 'stitch together' different molecular cascades, transforming raw data into a composite picture of organ wellness and systemic health.

Table 1. Commercial Metabolomics Software Platforms: Comparative Overview.

Platform	Vendor	Primary Functions	Key Strengths	Notable Limitations	Target Instruments
MassProfiler Professional	Agilent	Chemometric analysis, feature extraction, multivariate statistics, pathway mapping	Robust Agilent integration, comprehensive statistical visualization, database-linked annotation	Limited cross-vendor compatibility, closed algorithmic architecture	Agilent GC/MS, LC/MS, CE/MS
Compound Discoverer	Thermo Fisher	Untargeted/semi-targeted workflows, preprocessing, library identification (mzCloud, mzVault)	Node-based workflow design, extensive library ecosystem, multi-modal MS support	Premium licensing costs, restricted extensibility	Thermo Fisher LC–MS, high-resolution MS
PeakView	SCIEX	Qualitative LC–MS/MS analysis, spectral visualization, structural elucidation	Intuitive spectrum-centric interface, batch processing capabilities	Primarily qualitative focus, SCIEX-specific formats	SCIEX TripleTOF, LC–MS/MS systems
MarkerLynx XS	Waters	Automated feature detection, data matrix generation, multivariate analysis (PCA)	Optimized Waters UPLC–MS integration, robust peak alignment	Proprietary environment, limited external interoperability	Waters UPLC–MS (MassLynx)
MetaboScape	Bruker	Discovery metabolomics/ lipidomics, T-ReX feature extraction, CCS-aware workflows	Advanced multidimensional visualization, ion-mobility integration	Optimized for Bruker formats, costly library licensing	Bruker high-resolution, ion-mobility MS
Profiler AM+	Shimadzu	LC/MS, GC/MS analysis, metabolic pathway mapping, quantitative visualization	Seamless method package integration, metabolic map visualization	Shimadzu-centric ecosystem, limited algorithmic transparency	Shimadzu LC–MS, GC–MS platforms

Table adapted and synthesized from vendor documentation and published platform comparisons [25,29,46].

Table 2. Overview of Software Tools.

Tool Name	Type	Platform	Dependency/ Implementation	Software Availability	Validation Reference (Example)	Performance Metrics (Appropriate Wording)	Typical Use Cases
XCMS Online	GUI	Web	Cloud-hosted web application; backend built on R/XCMS with custom web services and job scheduling	https://xcmsonline.scripps.edu (accessed on 15 September 2023)	Smith et al., 2006; Tautenhahn et al., 2012 [47,48]	Performance varies by dataset and parameters; validated for reproducible feature detection across studies	Peak detection, retention time alignment, comparative metabolomics
MZmine	GUI	Cross-platform	Java (platform-independent); supports mzML, mzXML, NetCDF; modular processing pipeline	http://mzmine.github.io (accessed on 15 September 2023)	Pluskal et al., 2010 [49]	No single sensitivity metric; benchmarked for comprehensive feature detection and visualization	End-to-end LC–MS data processing and visualization
MetaboAnalyst	GUI	Web	R backend with Shiny web framework; relies on multiple CRAN/Bioconductor packages	https://www.metaboanalyst.ca (accessed on 15 September 2023)	Chong et al., 2019 [50]	Statistical validity depends on selected normalization and models; widely benchmarked	Statistical analysis, pathway analysis, metabolite profiling
MS-DIAL	GUI	Windows	C# (.NET Framework); includes in-house spectral libraries; not VBA-dependent	http://prime.psc.riken.jp/Metabolomics/MS-DIAL (accessed on 15 September 2023)	Tsugawa et al., 2015 [42]	High coverage and annotation rate for untargeted LC–MS/MS in benchmarking studies	Non-targeted LC–MS/MS, spectral deconvolution, compound annotation
XCMS	R package	Cross-platform	R with C/C++; Bioconductor package; integrates MSnbase, BiocParallel	https://bioconductor.org/packages/xcms (accessed on 15 September 2023)	Smith et al., 2006; Tautenhahn et al., 2008 [47,51]	Robust and reproducible feature detection when parameters are optimized	Peak picking, retention time correction, alignment

Table 2. *Cont.*

Tool Name	Type	Platform	Dependency/ Implementation	Software Availability	Validation Reference (Example)	Performance Metrics (Appropriate Wording)	Typical Use Cases
RAMClustR	R package	Cross-platform	Pure R implementation; operates downstream of XCMS/MZmine outputs	https://cran.r-project.org/package=RAMClustR (accessed on 15 September 2023)	Broeckling et al., 2014 [52]	Demonstrated improvement in feature clustering coherence (no universal accuracy metric)	Spectral feature clustering into compounds
CAMERA	R package	Cross-platform	R (Bioconductor); designed to work with XCMS objects	https://bioconductor.org/packages/CAMERA (accessed on 15 September 2023)	Kuhl et al., 2012 [53]	Evaluated for isotope/adduct annotation performance	Peak annotation, isotope and adduct grouping
MetAlign	GUI	Windows	C++; historically 32-bit executable (runs on 64-bit Windows); optimized for large datasets	https://www.plantmetabolomics.org/metalign (accessed on 15 September 2023)	Lommen, 2009 [54]	Demonstrated high alignment sensitivity in GC–MS and LC–MS datasets	Signal alignment, normalization
MetaboLyzer	R package	Cross-platform	R; integrates statistical testing and metabolite identification workflows	https://cran.r-project.org/package=MetaboLyzer (accessed on 15 September 2023)	Mak et al., 2013 [55]	Emphasis on statistical robustness rather than raw detection sensitivity	Differential metabolomics, statistical analysis

Table 3. Comprehensive List of Database Tools.

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
COCONUT	Vendor-neutral (open formats); Cloud-based	Web	https://coconut.naturalproducts.net (accessed on 15 September 2023)	[56]
MUTLIN MS2 Molecular standards database	LC-MS/MS	Web	http://metlin.scripps.edu/ (accessed on 15 September 2023)	[57]
CSMDB	NMR	MATLAB R2027b	https://github.com/cibionnmrlab/CSMDB-with-ConQuer-ABC (accessed on 15 September 2023)	[58]
EMBL-MCF	LC-MS	NA	https://curatr.mcf.embl.de/ (accessed on 15 September 2023)	[59]

Table 4. Advantages and Disadvantages of Standard Metabolomics Tools.

Platform	Strengths	Weaknesses	Sensitivity Rating (1–5) *	Library Availability Rating (1–5) *
XCMS Online	User-friendly web interface; robust peak detection	Limited to web access; server dependency	4	5
MZmine	Comprehensive feature set; good visualization tools	Requires Java installation	4	4
MetaboAnalyst	Extensive statistical and visualization modules	Web-based; requires stable internet connection	5	5
MS-DIAL	Strong for non-targeted LC–MS/MS workflows	Windows-only desktop application	4	4
XCMS	Well-established peak picking and alignment	Steeper learning curve; R scripting needed	4	4
RAMclustR	Effective clustering of features into spectra	Requires R proficiency	4	3
CAMERA	Powerful peak annotation (isotopes, adducts)	Limited to R environment	4	4
MetAlign	Accurate alignment and baseline correction	Windows-only GUI; legacy interface	4	4
MetaboLyzer	Supports metabolite identification and statistics	Requires R knowledge; more limited ecosystem	3	3

* Ratings are qualitative and provided for comparative illustration across commonly used platforms, not as absolute performance measures.

Table 5. Comprehensive List of Reviewed Preprocessing Tools.

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
CROP	LC–MS/MS	R 3.5.3	https://github.com/rendju/CROP (accessed on 15 September 2023)	[60]
ncGTW	LC–MS/MS	R 3.5.3, C++	https://github.com/ChiungTingWu/ncGTW (accessed on 15 September 2023)	[61]
TidyMS	LC–MS/MS	Python 3.8	https://github.com/griquelme/tidyms (accessed on 15 September 2023)	[62]
AutoTuner	LC–MS/MS	R 4.0.1	https://github.com/crmclean/Autotuner (accessed on 15 September 2023)	[63]
hRUV	LC–MS/MS	R 4.1.0	https://shiny.maths.usyd.edu.au/hRUV/ (accessed on 15 September 2023)	[64]
MetumpX	Any	R 3.5.3	https://github.com/hasaniqbal777/MetumpX-bin (accessed on 15 September 2023)	[65]

Table 6. Comprehensive List of Reviewed QC Tools.

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
MetaQuac	Targeted LC–MS	R 3.4.4	https://github.com/bihealth/metaquac (accessed on 15 September 2023)	[65]
dbnorm	Vendor-neutral (open formats); Desktop (local)	R 4.0.1	https://github.com/NBDZ/dbnorm (accessed on 15 September 2023)	[66]
MetaClean	LC–MS/MS	R 4.1.0	https://cran.r-project.org/src/contrib/Archive/MetaClean/ (accessed on 15 September 2023)	[67]
NeatMS	LC–MS/MS	Python 3.8	https://github.com/bihealth/ (accessed on 15 September 2023)	[68]

Table 7. Comprehensive List of Reviewed Annotation Tools.

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
MESSAR	LC–MS/MS	Web	https://messar.biodatamining.be/ (accessed on 15 September 2023)	[69,70]
SMART 2.0	2D NMR	Web	https://smart.ucsd.edu/classic (accessed on 15 September 2023)	[71]
MetFID	MS/MS data	Python-based artificial neural network (ANN) model implemented using Keras/TensorFlow	Method described in literature; no dedicated public web server reported	[72]

Table 7. *Cont.*

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
CPVA	Vendor-neutral (open formats); Cloud-based	Web	https://academic.oup.com/bioinformatics/article/36/12/3913/5809525 (accessed on 15 September 2023)	[73]
NRPro	LC–MS/MS	Java, Web	https://bioinfo.cristal.univ-lille.fr/nrpro/ (accessed on 15 September 2023)	[74]
MetENP/MetENPWeb	LC–MS/MS	R, Web	https://www.metabolomicsworkbench.org/data/analyze.php (accessed on 15 September 2023)	[75]
CANOPUS	LC–MS/MS	Standalone	https://bio.informatik.uni-jena.de/software/canopus/ (accessed on 15 September 2023)	[71]
MolDiscovery	LC–MS/MS	Python 3.8	https://github.com/mohimanilab/molDiscovery (accessed on 15 September 2023)	[76]
MetIDfyR	LC–MS/MS	R 4.0.0	https://github.com/agnesblch/MetIDfyR (accessed on 15 September 2023) MetIDfyR	[77]
Qemistree	LC–MS/MS	Python 3.8	https://github.com/biocore/q2-qemistree (accessed on 15 September 2023)	[78]
IIMN	LC–MS/MS	GNPS, Web	https://ccms-ucsd.github.io/GNPSDocumentation/fbmn-iin/ (accessed on 15 September 2023)	[79]
FOBI	Vendor-neutral (open formats); Hybrid (cloud + local)	R, Web	https://github.com/pcastellanoescuder/FOBI_Visualization_Tool (accessed on 15 September 2023)	[80]
Biodendro	LC–MS/MS	Python 3.8	https://github.com/ccdmb/BioDendro (accessed on 15 September 2023)	[81]
AllCCS atlas	IM-MS	Web	https://github.com/ZhuMetLab/AllCCS; http://allccs.zhulab.cn/ (accessed on 15 September 2023)	[82]
Binner	LC–MS/MS	Java	https://binner.med.umich.edu/ (accessed on 15 September 2023)	[83]
MS-CleanR	LC–MS/MS	R <=4.2	https://github.com/eMetaboHUB/MS-CleanR (accessed on 15 September 2023)	[84]
Retip	LC–MS/MS	R 4.0.0	https://www.retip.app/ (accessed on 15 September 2023)	[85]

Table 7. *Cont.*

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
QSRR Automator	LC–MS/MS	Python 3.8	https://github.com/UofUMetabolomicsCore/QSRR_Automator/releases/tag/v1_exe (accessed on 15 September 2023)	[86]
MFAssignR	LC–MS/MS	R, HTML	https://github.com/skschum/MFAssignR (accessed on 15 September 2023)	[87]
McSearch	LC–MS/MS	R 4.0.0	https://github.com/HuanLab/McSearch (accessed on 15 September 2023)	[88]
REDU	LC–MS/MS	GNPS, Web	https://ccms-ucsd.github.io/GNPSDocumentation/ReDU/ (accessed on 15 September 2023)	[89]
MASST	LC–MS/MS	GNPS, Web	https://masst.gnps2.org/ (accessed on 15 September 2023)	[90,91]
NPClassifier	Any	Web	https://npclassifier.ucsd.edu/ (accessed on 15 September 2023)	[92]
patRoon	HR MS/MS	R 4.3.3	https://github.com/rickhelmus/patRoon (accessed on 15 September 2023)	[93]
LipidLynxX	LC–MS/MS	Python 3.8, Standalone	http://lipidmaps.org/lipidlynxx/ (accessed on 15 September 2023)	[94]

Table 8. Comprehensive List of Reviewed Statistical and Visualization Tools.

Software Tool	Analysis Type/Optimization	Platform Dependency	Implementation	Software Availability	Ref.
Epimetal	Visualization/ Exploration	Vendor-neutral (open formats); Cloud-based	JavaScript, Web	https://github.com/amergin/epimetal; https://github.com/amergin/epimetal (accessed on 15 September 2023)	[95]
Metabolite AutoPlotter	Quantitative/ Targeted	Quantitative metabolomics data, any	R, Web	https://mpietzke.shinyapps.io/AutoPlotter/ (accessed on 15 September 2023)	[96]
Metabolite-Investigator	Quantitative/ Targeted	LC-MS	R, Web	https://github.com/cfbeuchel/Metabolite-Investigator (accessed on 15 September 2023)	[97]
VIIME	Visualization/ Exploration	Vendor-neutral (open formats); Cloud-based	Web	https://viime.org/#/ (accessed on 15 September 2023)	[98]
struct	Quantitative/ Targeted	Vendor-neutral (open formats); Desktop (local)	R based	https://bioconductor.org/packages/release/bioc/html/struct.html (accessed on 15 September 2023)	[99]
lipidr	Untargeted LC-MS	LC-MS/ MS	R based	https://www.lipidr.org/ (accessed on 15 September 2023)	[100]

Table 8. *Cont.*

Software Tool	Analysis Type/Optimization	Platform Dependency	Implementation	Software Availability	Ref.
NOREVA (Statistics Only)	Quantitative/ Statistical Analysis	Vendor-neutral (open formats); Hybrid (cloud + local)	Web, R, Standalone	http://idrblab.cn/noreva/ (accessed on 15 September 2023)	[101]
%polynova_2way (Statistics Only)	Quantitative/ Statistical Analysis	Processed Data	SAS	https://doi.org/10.1371/journal.pone.0244013 (accessed on 15 September 2023)	[102]
rawR (Visualization)	Visualization/ Exploration	LC-MS	R 4.0.0, C++	https://github.com/fgcz/rawR (accessed on 15 September 2023)	[103]
MEtaboverse (Visualization)	Visualization/ Exploration	Vendor-neutral (open formats); Desktop (local)	Java, HTML, Stanalone	https://github.com/Metaboverse (accessed on 15 September 2023)	[104]
JS-MS 2.0 (Visualization)	Visualization/ Exploration	LC-MS/MS	Java, JavaScript, HTML	https://github.com/optimusmoose/jsms (accessed on 15 September 2023)	[105]
Metpropagate (Visualization)	Untargeted LC-MS	Untargeted LC-MS	R package, Python > 3.4	https://github.com/emmag (accessed on 15 September 2023)	[106]

Table 9. Comprehensive List of Isotopic Tools.

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
isoSCAN	GC-MS	R package	https://github.com/jcapelladesto/isoSCAN (accessed on 15 September 2023)	[107]
MIAMI	GC-MS	Compiled C++ source code	https://md.tu-bs.de/ (accessed on 15 September 2023)	[108]

Table 10. Comprehensive List of Lipidomics, MSI and Multiomics Tools.

Software Tool	Platform Dependency	Implementation	Software Availability	Ref.
LiPydomics	Lon Mobility, Lipidomics	Python, HTML	https://github.com/dylanhross/lipydomics (accessed on 15 September 2023)	[109]
LipidCreator	LC-MS	C#, HTML, Skyline plugin	https://github.com/lifs-tools/lipidcreator (accessed on 15 September 2023)	[110]
Lipid Annotator	LC-MS/MS	Raman imaging/MSI (LC-MS/MS imaging)	R and C++ (open-source R package)	[111]
Raman2imzML	Raman	C++, R package	https://github.com/LlucSF/Raman2imzML (accessed on 15 September 2023)	[112]
SUMMER	LC-MS/Lipidomics	R, Web	http://igc1.salk.edu:3838/summer/ and https://bitbucket.org/salkigc/summer/src/master/ (accessed on 15 September 2023)	[112]

4. Analysis of Metabolomic Data Using Machine Learning and Artificial Intelligence

Machine learning algorithms are required to map disease states from complex datasets. These systems learn from embedded data to identify unique patterns without explicit programming. Deep learning models provide opportunities to individualize diagnosis and predict drug responses. This facilitates the advancement of precision medicine and therapeutic discovery. Furthermore, integrating pathway predictions with AI helps researchers identify drug targets and off-target effects. Sampling data from healthy and diseased individuals can distinguish between the two groups based on the patterns of metabolites present (Figure 2). It is important to distinguish the type of insight AI provides: most applications identify statistical associations between metabolites and phenotypes, some models predict pathway membership using network-based approaches, and others can generate mechanistic hypotheses, which require experimental validation. For example, researchers can use metabolomic data to map complications during pregnancy [113] and monitor solid tumor progression throughout a treatment course [114].

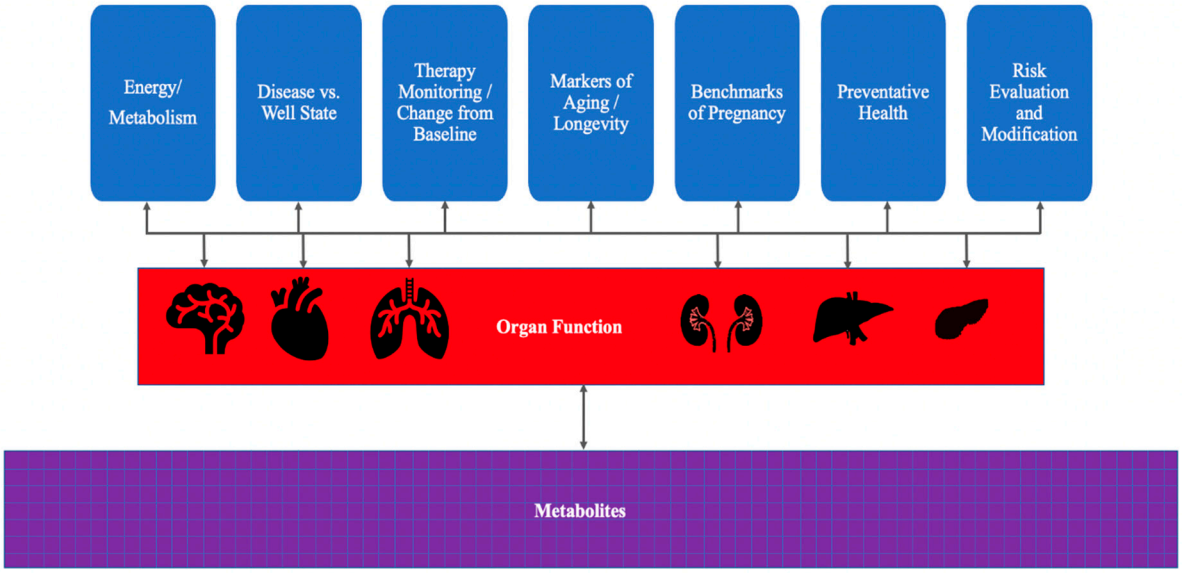


Figure 2. Developing Metabolomics Applications in Health and Disease.

It is important to distinguish the type of insight AI provides in metabolomics. Most AI applications in this field primarily detect statistical associations; as differentiating healthy from diseased individuals based on metabolite profiles. Some models, particularly graph neural networks and pathway-based deep learning approaches can predict pathway membership and identifying which metabolic routes are likely active or dysregulated. While these predictions can suggest possible intervention points, true mechanistic insight; the causal understanding of how specific metabolites influence biological processes; requires experimental validation. Therefore, AI serves as a powerful tool for pattern recognition and hypothesis generation, but translating these findings into mechanistic understanding relies on complementary biochemical and molecular studies.

Another application of machine learning in metabolomics is the prediction of metabolic pathways and networks that regulate routine homeostasis. Machine learning contributes to pathway and network analysis through pathway enrichment, network inference and hybrid knowledge-driven approaches where prior biochemical knowledge is integrated with high-dimensional metabolomic data to improve pathway prediction and suggest potential mechanistic relationships. By integrating metabolomic pathway predictions with AI, researchers can identify potential drug targets, therapeutic intervention points,

and off-target effects, supporting drug development and mechanism-of-action studies. Metabolism involves reactions that convert raw materials into energy and other useful valuable molecules. Understanding these pathways and networks can provide insight into the functioning of cells and tissues. Machine learning algorithms can be used to analyze metabolomics data and infer the underlying pathways and networks, which can help to identify potential targets for therapeutic intervention or to understand the mechanisms of diseases. Assays that measure metabolites for endpoint targeting, such as cardiovascular event risk, trauma outcomes [24], glucose tolerance [115], body composition correlation with health, kidney and liver health [116], and the impact of lifestyle on health, are presently being developed [117,118]. A recent comprehensive review details software and informatics tools specifically dedicated to lipid metabolism and the intricacies of lipidomics [119].

4.1. From Sample Data to Intelligent Mapping of Pathways

AI capabilities are categorized by their complexity and application. Machine learning algorithms are utilized to map health states from large datasets and improve analysis methods through automated processing. In contrast, Deep learning networks, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), rely on decision trees and neural architectures to develop complex automated detection capabilities, extending from raw data preprocessing to feature selection and taxonomy. Machine learning algorithms are designed for computers to improve data analysis methods without explicit programming. Over time they develop improved accuracy and precision in learning complex patterns, and their automated processing permits improved task-oriented speed and efficiency. They also facilitate the interpretation of results. Deep learning networks rely on neural networks and decision trees to develop complex automated analysis and detection capabilities. They can be implemented into each data harvesting step beginning with data preprocessing and extending to feature selection and taxonomy. Rather than providing a direct map, AI-driven analysis infers pathological states by identifying complex signatures within high-dimensional data. For example, machine learning algorithms have been utilized to analyze 8500 multi-omic biomarkers to predict 30-day survival and recovery in trauma patients. Similarly, AI has been applied to plasma profiles to predict the response to neoadjuvant chemotherapy in triple-negative breast cancer and to detect intrauterine growth restriction by analyzing multi-platform metabolomics data. In renal health, these tools (Table 11) facilitate the prediction of chronic kidney disease progression, moving from general observation to specific prognostic indicators.

Table 11. Machine Learning and Deep Learning Algorithms in Metabolomics Data Analysis.

Algorithm	Strengths	Weaknesses	Applications in Metabolomics	Metabolomics Tools
Random Forest	High accuracy and robustness to overfitting	Slow training and prediction times	Identifying biomarkers in disease diagnosis	MetaboAnalyst, KPIC2, W4M
Support Vector Machine	Good at handling high-dimensional data	Sensitive to outliers	Classifying samples based on metabolic profiles	MetaboAnalyst, apLCMS
Neural Network	High degree of accuracy, able to manage large datasets	Requires large data amounts for training	Identifying metabolic pathways and prediction of metabolic fluxes	CRANK-MS, MetaboLabPy

Table 11. *Cont.*

Algorithm	Strengths	Weaknesses	Applications in Metabolomics	Metabolomics Tools
Deep Belief Network	Good at feature extraction and unsupervised learning	Requires large data amounts for training	Identifying unknown metabolic compounds and clustering of samples based on metabolic profiles	Custom Python/ R pipelines
Convolutional Neural Network	Good at image recognition and handling high-dimensional data	Requires large data amounts for training	Identifying metabolic images and prediction of metabolic fluxes	DeepMet, SIRIUS (for CSI:FingerID)
k-Nearest Neighbors	Simple and easy to implement	Sensitive to the choice of k and the presence of outliers	Classifying samples based on metabolic profiles	MetaboAnalyst, MZmine
Principal Component Analysis (PCA)	Good at dimensionality reduction and visualization of data	Can lose important information during dimensionality reduction	Identifying metabolic pathways and clustering of samples based on metabolic profiles	MetaboAnalyst, XCMS, GNPS
Linear Discriminant Analysis (LDA)	Good at feature extraction and classification	Assumes that the data maps to a Gaussian distribution	Identifying biomarkers in disease diagnosis and classification of samples based on metabolic profiles	MetaboAnalyst (as PLS-DA)
Artificial Neural Network (ANN)	High degree of accuracy, able to manage large datasets	Requires large data amounts for training and can be sensitive to overfitting	Identifying metabolic pathways and prediction of metabolic fluxes	CRANK-MS, MetaboAnalyst (via R)
Recurrent Neural Network (RNN)	Good at handling sequential data	Requires large data amounts for training and can be sensitive to overfitting	Identifying metabolic pathways and prediction of metabolic fluxes in time-series data	Custom time-series scripts

4.2. Examples of Artificial Intelligence in Metabolomics Data Analysis

4.2.1. Classical Machine Learning Methodologies

Classical approaches involve both supervised methods (using labeled data to predict group membership) and unsupervised methods (identifying patterns without predefined labels).

- Random Forest (Supervised): This algorithm is most appropriate for identifying biomarkers in disease diagnosis due to its high accuracy and robustness against overfitting.
- Support Vector Machine (SVM) (Supervised): SVM is highly effective at classifying samples based on metabolic profiles, particularly when handling high-dimensional data.
- k-Nearest Neighbors (k-NN) (Supervised): This is a simple methodology primarily used for the classification of samples according to their metabolic profiles.
- Linear Discriminant Analysis (LDA) (Supervised): LDA is used for both biomarker identification and sample classification, assuming the data follows a Gaussian distribution.
- Principal Component Analysis (PCA) (Unsupervised): PCA is widely utilized for dimensionality reduction, data visualization, identifying metabolic pathways, and clustering samples.

4.2.2. Deep Learning Methodologies

Deep learning relies on neural networks and decision trees to perform complex automated analysis across various stages of the metabolomics workflow, from preprocessing to feature selection1. These models are specifically used to individualize diagnosis, treatment, and drug response prediction.

- Artificial Neural Networks (ANN)/General Neural Networks: These are most appropriate for identifying metabolic pathways and predicting metabolic fluxes because they can manage very large datasets with high accuracy.
- Convolutional Neural Networks (CNN): Beyond image recognition, CNNs are used for identifying metabolic images, predicting metabolic fluxes, and performing regression and pattern categorization on high-dimensional spectra like MS or NMR.
- Recurrent Neural Networks (RNN): RNNs are specialized for sequential and time-series data, making them ideal for identifying metabolic pathways and predicting fluxes over time4. They are also used to analyze the sequential features of high-dimensional spectra.
- Deep Belief Networks (DBN): These networks excel at unsupervised learning and feature extraction, making them suitable for identifying unknown metabolic compounds and clustering samples.

Organizing biomarker samples into clusters reduces the data dimensionality and permits easier classification. Complex, high-dimensional metabolomics data, such as mass spectrometry or nuclear magnetic resonance spectra, are most effectively analyzed using convolutional neural networks (CNNs) for spatial and spectral feature extraction, whereas recurrent neural networks (RNNs) are specifically applied to sequential metabolomics datasets, including longitudinal, time-course, dynamic fluxomic, and treatment-response studies. Recurrent neural networks are primarily designed to model ordered sequences and temporal dependencies. Therefore they are not applied to static metabolomic spectra in a generic manner. In metabolomics, RNN architectures, such as long short-term memory (LSTM) and gated recurrent unit (GRU) networks, are implemented for time-resolved metabolomics, metabolic flux analysis, and dynamic perturbation experiments to capture delayed and non-linear metabolite responses across sampling time points. These models are particularly useful for predicting treatment response trajectories, circadian metabolic oscillations, and disease progression based on longitudinal metabolomic measurements. These algorithms learn and extract unique elements for pattern categorization and regression. These AI methods can also highlight candidate metabolites for drug targeting and evaluate compound effects in silico, supporting early-phase drug development. Reinforcement learning algorithms may guide the design of a metabolomics study. These algorithms learn from known data and issue recommendations on sample selection and desired measurements with an eye toward efficient resource utilization.

4.3. Limitations of Metabolomics Tools and Strategies for Data Integration

Although many robust data management tools and new applications for data utilization have been discussed in this review, many limitations of these metabolomics tools still need to be addressed [5,6]. First, the set of compounds which can be targeted is limited. Similarly, compounds that are not targeted will remain undetected. Increasingly complex diagnostic tools are in development that permit detection of all possible trackable components in the sample, a significant improvement from serial single-metabolite analysis. There is further complexity in sample processing when targeted compounds require purification for proper calibrated detection and compounds of interest must be preprogrammed for machine recognition. Other limitations include low throughput speeds. Having corroborative information accessible from alternate sources such as genomics or other 'omics may help

mediate these issues. Further limitations lie in streamlining real-time detection, which is exacerbated by the complexity of the multistep process [120].

While high-dimensional untargeted metabolomics combined with machine learning has become pervasive in discovery research, this "many metabolites × ML" paradigm carries inherent challenges that merit critical evaluation. Untargeted approaches generate very large feature spaces that often require complex statistical and computational methods to interpret. The identities of many detected signals remain unknown due to gaps in spectral libraries and reference standards, complicating downstream biological interpretation [121]. Machine learning models trained on such extensive datasets may tend to overfit when sample sizes are limited and can function as "black boxes." This makes it difficult to extract actionable mechanistic insight or validate causal relationships. These are limitations that constrain clinical adoption [122]. In contrast, targeted metabolomics assays focus on a defined subset of known metabolites, enabling absolute quantification, improved reproducibility, and clearer biological interpretation, thereby facilitating clinical validation and regulatory acceptance [123]. Targeted methods are routinely used to validate candidate biomarkers identified in untargeted screens, monitor specific pathways, and quantify low-abundance metabolites with high precision [15]. Biologically informed feature selection and dimensionality-reduction strategies can reduce the complexity of ML models. These capabilities enhance interpretability by prioritizing metabolites with known relevance to physiological pathways. This strategy is increasingly discussed in the literature as providing more robust and clinically meaningful metabolomics analyses [124]. Together, these considerations emphasize that while high-dimensional ML is a powerful discovery tool, integrating it with targeted, biologically constrained, and platform-optimized approaches advances metabolomics toward reproducible, interpretable, and clinically actionable outcomes.

High-throughput analysis can be extremely helpful in the intraoperative setting, such as evaluating tissue margins, tissue composition and tissue pathology or in the critical care setting. It is important to balance the cost-effectiveness of various evaluation techniques such as ELISA and immunochromatography given systemic economic constraints.

Since metabolomics data can involve thousands of metabolites, the high dimensionality of data can make it difficult to achieve statistical power. In the preprocessing stage, there is further risk of instrument drift and sample processing variation. This can often be improved with data normalization and batch effect correction. Integrating this data with other genomic and proteomic markers and their associated variability and potential biases in measurement and processing can bring additional complexity. Statistical methods can also vary with attendant risks of limited knowledge of metabolite behavior and overfitting of data. Univariate, multivariate statistics and machine learning approaches may be used and cross-referenced with data from known metabolite databases to identify metabolites' functional and regulatory roles. Decision tree and random forest models are commonly used for feature selection by ranking metabolites according to their predictive importance. In contrast, unsupervised clustering approaches (e.g., hierarchical clustering and k-means) are applied as a separate analytical step to group metabolites or samples based on similarity in their abundance profiles.

Platform-Specific Machine Learning Challenges in Metabolomics

While machine learning algorithms are widely applied in metabolomics, their performance is strongly influenced by platform-specific data characteristics [125]. In NMR-based metabolomics, relatively low analytical sensitivity and extensive spectral overlap limit the number of reliably detected metabolites. They can also introduce multicollinearity, which may reduce feature stability and compromise model interpretability. These limitations

necessitate spectral deconvolution, region exclusion, and peak alignment strategies prior to model training [84]. In mass spectrometry-based metabolomics, additional challenges arise from adduct formation, in-source fragmentation, batch effects, and annotation uncertainty. Multiple ion species for a single metabolite inflate feature dimensionality and introduce redundancy, which can create bias in feature importance estimates in tree-based and regression-based models. Uncertainties in compound annotation propagate into training labels and can degrade supervised learning performance [84]. Missing values are prevalent in metabolomics due to detection limits and stochastic ion suppression. As a result, advanced preprocessing strategies, such as probabilistic quotient normalization, LOESS-based signal correction, k-nearest neighbor and random forest-based imputation along with batch effect correction are essential for preserving biological signal and ensuring reliable downstream machine learning performance. Inadequate normalization and imputation can substantially alter classification accuracy, biomarker ranking, and pathway enrichment outcomes, highlighting the critical importance of preprocessing pipelines in metabolomics-specific ML workflows [101].

4.4. The Future of Metabolomics Processing and Data Analysis

Future metabolomics tools will increasingly integrate AI to improve throughput, identify druggable pathways, and accelerate translational research. Open-source, robust, and user-friendly software platforms will make it easier to link metabolomic insights to therapeutic development, supporting precision medicine and drug discovery. By analyzing complex datasets, these tools can pinpoint key elements of cellular processing and track sequential metabolic byproducts, revealing patterns that suggest potential therapeutic targets. While many existing tools focus primarily on annotations, challenges remain in processing LC-MS/MS and multi-omics data efficiently. Web-based platforms, often with associated GitHub repositories, provide growing datasets that improve model accuracy over time. Programming languages such as R, Python, and Java are increasingly integrated, enabling diverse datasets to be combined and analyzed more effectively. As these tools become faster, more computationally efficient, and increasingly robust, researchers will be able to uncover meaningful metabolic patterns more reliably, directly translating software improvements into actionable insights for drug discovery and translational research.

4.5. Ethical and Privacy Considerations in Clinical AI

While this review identifies technical limitations such as high dimensionality and statistical power, the clinical implementation of AI-driven metabolomics also faces significant privacy and ethical challenges. As workflows move toward individualizing diagnosis and treatment, the datasets generated contain sensitive biological information. Ensuring data security is critical, particularly when integrating multi-omics data (genomics, proteomics, and lipidomics) that could potentially identify individual participants. Future research must prioritize the development of privacy-preserving AI architectures to ensure that the transition from 'metabolomic insight' to 'translational therapeutic discovery' remains compliant with ethical standards and regulatory frameworks.

4.6. Critical Evaluation of AI and Multi-Omics Integration

Although AI-enhanced workflows and multi-omics integration are frequently presented as key innovations in metabolomics, several challenges and trade-offs limit their mechanistic insight and scientific impact. Multi-omics datasets are heterogeneous and often incomplete, complicating integration and interpretation of results. Rigorous preprocessing and careful handling of missing data are important to avoid introducing bias or obscuring biological signals [126]. Untargeted metabolomics produces extensive raw data, including many unidentified features. There is limited coverage of comprehensive

spectral libraries, and uncertain feature annotation weakens AI models' ability to map computational patterns to actual biochemical mechanisms [127]. Machine learning models applied to high-dimensional metabolomics frequently face the curse of dimensionality, in which the number of variables exceeds the number of samples, increasing the risk of overfitting and reducing generalizability across cohorts [127]. Deep learning and other complex models may achieve high predictive performance but often lack transparency, making it difficult to interpret which metabolic features drive model outputs and how these relate to underlying biology. Similarly, multi-omics integration holds promise for linking metabolism with upstream regulatory layers, but differences in data scale, noise profiles, and statistical properties across omics types remain obstacles to reliable integration and mechanistic inference [128].

5. Conclusions

The integration of AI-driven automation into metabolomic workflows represents a transformative shift in our ability to interpret the complex biochemical interactions of living organisms. By synchronizing diverse software toolkits, from preprocessing to post-analysis, researchers can create a more comprehensive "workflow engine" that transforms high-dimensional data into actionable biological insights. This approach moves beyond simple metabolite identification to a dynamic understanding of how interventions alter systemic health and disease states.

Ultimately, the synergy between AI and multi-omics strategies is the key to bridging the current gap between basic metabolomic research and translational therapeutic discovery. By addressing the challenges of high-throughput data analysis and providing a scalable framework for biomarker discovery, these computational approaches provide the foundation for the next generation of precision medicine and individualized clinical interventions.

Author Contributions: A.F. conceived the idea and wrote the article with K.D.E. A.H. contributed to writing and organizing information into tables and figures. K.O. supervised the work and provided overall guidance. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s). Availability of Data and Materials (all web accessed on 15 September 2023):

QC Tools (Table 2):
https://github.com/bihealth/metaquac
https://github.com/NBDZ/dbnorm
https://cran.r-project.org/src/contrib/Archive/MetaClean/
https://github.com/bihealth/
Annotated Tools (Table 3):
https://messar.biodatamining.be/
https://smart.ucsd.edu/classic
https://academic.oup.com/bioinformatics/article/36/12/3913/5809525
https://bioinfo.cristal.univ-lille.fr/nrpro/
https://www.metabolomicsworkbench.org/data/analyze.php
https://bio.informatik.uni-jena.de/software/canopus/
https://github.com/mohimanilab/molDiscovery
https://github.com/agnesblch/MetIDfyR
https://github.com/biocore/q2-qemistree

https://ccms-ucsd.github.io/GNPSDocumentation/fbmn-iin/
https://github.com/pcastellanoescuder/FOBI_Visualization_Tool
https://github.com/ccdmb/BioDendro
https://github.com/ZhuMetLab/AllCCS
http://allccs.zhulab.cn
https://binner.med.umich.edu/
https://github.com/eMetaboHUB/MS-CleanR
https://www.retip.app/
https://github.com/UofUMetabolomicsCore/QSRR_Automator/releases/tag/v1_exe
https://github.com/skschum/MFAssignR
https://github.com/HuanLab/McSearch
https://redu.ucsd.edu/
https://masst.ucsd.edu/
https://npclassifier.ucsd.edu/
https://github.com/rickhelmus/patRoon
http://lipidmaps.org/lipidlynxx/

Preprocessing Tools (Table 4)

https://github.com/rendju/CROP
https://github.com/ChiungTingWu/ncGTW
https://github.com/griquelme/tidyms
https://github.com/crmclean/Autotuner
https://shiny.maths.usyd.edu.au/hRUV/
https://github.com/hasaniqbal777/MetumpX-bin

Statistical and Visualization Tools (Table 5)

https://github.com/amergin/epimetal
http://epimetal.computationalmedicine.fi/
https://mpietzke.shinyapps.io/AutoPlotter/
https://github.com/cfbeuchel/Metabolite-Investigator
https://viime.org/#/
https://bioconductor.org/packages/release/bioc/html/struct.html
https://www.lipidr.org
http://idrblab.cn/noreva/
https://doi.org/10.1371/journal.pone.0244013
https://github.com/fgcz/rawR
https://github.com/Metaboverse
https://github.com/optimusmoose/jsms
https://github.com/emmag

Platform Tools (Table 6)

https://github.com/sipss/AlpsNMR
https://github.com/BEKZODKHAKIMOV/SigMa_Ver1
https://github.com/stefhk3/nmrfilterprojects
https://bitbucket.org/iAnalytica/mshub_process/src/master/
https://github.com/CCMS-UCSD/GNPS_Workflows/tree/master/mshub-gc/tools/mshub-gc/proc

Database Tools (Table 7)

https://coconut.naturalproducts.net
http://metlin.scripps.edu/
https://github.com/cibionnmrlab/CSMDB-with-ConQuer-ABC
https://curatr.mcf.embl.de/

Isotopic Tools (Table 8)

https://github.com/jcapelladesto/isoSCAN
https://md.tu-bs.de/

Lipidomics, MSI and Multomics tools (Table 9)

https://github.com/dylanhross/lipydomics
https://github.com/lifs-tools/lipidcreator

https://github.com/LlucSF/Raman2imzML
http://igc1.salk.edu:3838/summer/
https://bitbucket.org/salkigc/summer/src/master/

Conflicts of Interest: The authors declare that they have no competing interests.

Abbreviations

AI	Artificial Intelligence
MSCAT	Metabolomics Software Catalog
MS	Mass spectroscopy
NMR	Nuclear Magnetic Resonance
FTIR	Fourier-transform infrared

References

1. Fiehn, O. Metabolomics—The link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155–171. [CrossRef] [PubMed]
2. Nicholson, J.K.; Lindon, J.C.; Holmes, E. "Metabonomics": Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, *29*, 1181–1189. [CrossRef]
3. Goodacre, R.; Vaidyanathan, S.; Dunn, W.B.; Harrigan, G.G.; Kell, D.B. Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends Biotechnol.* **2004**, *22*, 245–252. [CrossRef]
4. Patti, G.J.; Yanes, O.; Siuzdak, G. Metabolomics: The apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263–269. [CrossRef]
5. Shulaev, V. Metabolomics technology and bioinformatics. *Brief. Bioinform.* **2006**, *7*, 128–139. [CrossRef]
6. Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **2012**, *137*, 293–300. [CrossRef]
7. Drouin, N.; Ramautar, R. Capillary Electrophoresis-Mass Spectrometry for Metabolomics: Possibilities and Perspectives. *Adv. Exp. Med. Biol.* **2021**, *1336*, 159–178. [CrossRef] [PubMed]
8. Zhu, P.; Harms, A.C.; Maas, P.; Bakas, M.; Whien, J.J.; Dubbelman, A.C.; Hankemeier, T. Matrix effects in untargeted LC-MS metabolomics: From creation to compensation with post-column infusion of standards. *J. Chromatogr. A* **2025**, *1765*, 466508. [CrossRef] [PubMed]
9. Zhou, J.; Zhong, L. Applications of liquid chromatography-mass spectrometry based metabolomics in predictive and personalized medicine. *Front. Mol. Biosci.* **2022**, *9*, 1049016. [CrossRef]
10. Pinu, F.R.; Goldansaz, S.A.; Jaine, J. Translational Metabolomics: Current Challenges and Future Opportunities. *Metabolites* **2019**, *9*, 108. [CrossRef] [PubMed]
11. Cochran, D.; NourEldein, M.; Bezdekova, D.; Schram, A.; Howard, R.; Powers, R. A Reproducibility Crisis for Clinical Metabolomics Studies. *Trends Anal. Chem.* **2024**, *180*, 117918. [CrossRef] [PubMed] [PubMed Central]
12. Gomari, D.P.; Schweickart, A.; Cerchietti, L.; Paietta, E.; Fernandez, H.; Al-Amin, H.; Suhre, K.; Krumsiek, J. Variational autoencoders learn transferrable representations of metabolomics data. *Commun. Biol.* **2022**, *5*, 645. [CrossRef] [PubMed] [PubMed Central]
13. Baranwal, M.; Magner, A.; Elvati, P.; Saldinger, J.; Violi, A.; Hero, A.O. A deep learning architecture for metabolic pathway prediction. *Bioinformatics* **2020**, *36*, 2547–2553. [CrossRef] [PubMed]
14. Wang, C.; Yuan, C.; Wang, Y.; Chen, R.; Shi, Y.; Zhang, T.; Xue, F.; Patti, G.J.; Wei, L.; Hou, Q. MPI-VGAE: Protein-metabolite enzymatic reaction link learning by variational graph autoencoders. *Brief. Bioinform.* **2023**, *24*, bbad189. [CrossRef] [PubMed] [PubMed Central]
15. Wishart, D.S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* **2016**, *15*, 473–484. [CrossRef]
16. Pang, H.; Jia, W.; Hu, Z. Emerging applications of metabolomics in clinical pharmacology. *Clin. Pharmacol. Ther.* **2019**, *106*, 544–556. [CrossRef]
17. Yang, Q.; Zhang, A.-H.; Miao, J.-H.; Sun, H.; Han, Y.; Yan, G.-L.; Wu, F.-F.; Wang, X.-J. Metabolomics biotechnology, applications, and future trends: A systematic review. *RSC Adv.* **2019**, *9*, 37245–37257. [CrossRef]
18. Gao, P.; Huang, X.; Fang, X.-Y.; Zheng, H.; Cai, S.-L.; Sun, A.-J.; Zhao, L.; Zhang, Y. Application of metabolomics in clinical and laboratory gastrointestinal oncology. *World J. Gastrointest. Oncol.* **2021**, *13*, 536–549. [CrossRef] [PubMed]
19. Huo, M.; Zhang, J.; Huang, W.; Wang, Y. Interplay among metabolism, epigenetic modifications, and gene expression in cancer. *Front. Cell Dev. Biol.* **2021**, *9*, 793428. [CrossRef]

20. Verma, A.; Lindroth, A.M. The emerging intertwined activities of metabolism and epigenetics unveils culprits and prospects in cancer. *Exp. Mol. Med.* **2025**, *57*, 1928–1939. [CrossRef]
21. Fu, J.; Zhang, L.-L.; Li, W.; Zhang, Y.; Zhang, Y.; Liu, F.; Zou, L. Application of metabolomics for revealing the interventional effects of functional foods on metabolic diseases. *Food Chem.* **2022**, *367*, 130697. [CrossRef]
22. Kaddurah-Daouk, R.; Krishnan, K. Metabolomics: A global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology* **2009**, *34*, 173–186. [CrossRef]
23. Aderemi, A.V.; Ayeleso, A.O.; Oyedapo, O.O.; Mukwevho, E. Metabolomics: A scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases. *Metabolites* **2021**, *11*, 418. [CrossRef] [PubMed]
24. Abdelhamid, S.S.; Scioscia, J.; Vodovotz, Y.; Wu, J.; Rosengart, A.; Sung, E.; Rahman, S.; Voinchet, R.; Bonaroti, J.; Li, S.; et al. Multi-Omic Admission-Based Prognostic Biomarkers Identified by Machine Learning Algorithms Predict Patient Recovery and 30-Day Survival in Trauma Patients. *Metabolites* **2022**, *12*, 774. [CrossRef]
25. Broadhurst, D.; Goodacre, R.; Reinke, S.N.; Kuligowski, J.; Wilson, I.D.; Lewis, M.R.; Dunn, W.B. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018**, *14*, 72. [CrossRef]
26. Dunn, W.B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J.D.; Halsall, A.; Haselden, J.N.; et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6*, 1060–1083. [CrossRef]
27. Misra, B.B.; Langefeld, C.D.; Olivier, M.; Cox, L.A. Integrated omics: Tools, advances, and future approaches. *J. Mol. Endocrinol.* **2019**, *62*, R21–R45. [CrossRef]
28. Alseekh, S.; Aharoni, A.; Brotman, Y.; Contrepois, K.; D'auria, J.; Ewald, J.; Ewald, J.C.; Fraser, P.D.; Giavalisco, P.; Hall, R.D.; et al. Mass spectrometry-based metabolomics: A guide for annotation, quantification and best reporting practices. *Nat. Methods* **2021**, *18*, 747–756. [CrossRef] [PubMed]
29. Viant, M.R.; Kurland, I.J.; Jones, M.R.; Dunn, W.B. How close are we to complete annotation of metabolomes? *Curr. Opin. Chem. Biol.* **2017**, *36*, 64–69. [CrossRef] [PubMed]
30. Siskos, A.P.; Jain, P.; Römisch-Margl, W.; Bennett, M.; Achaintre, D.; Asad, Y.; Marney, L.; Richardson, L.; Koulman, A.; Griffin, J.L.; et al. Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma. *Anal. Chem.* **2016**, *89*, 656–665. [CrossRef]
31. Johnson, C.H.; Ivanisevic, J.; Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 451–459. [CrossRef]
32. Beger, R.D.; Dunn, W.B.; Bandukwala, A.; Bethan, B.; Broadhurst, D.; Clish, C.B.; Dasari, S.; Derr, L.; Evans, A.; Fischer, S.; et al. Towards quality assurance and quality control in untargeted metabolomics studies. *Metabolomics* **2019**, *15*, 4. [CrossRef]
33. Considine, E.C.; Thomas, G.; Boulesteix, A.L.; Khashan, A.S.; Kenny, L.C. Critical review of reporting of the data analysis step in metabolomics. *Metabolomics* **2018**, *14*, 7. [CrossRef]
34. Stanstrup, J.; Broeckling, C.D.; Helmus, R.; Hoffmann, N.; Mathé, E.; Naake, T.; Nicolotti, L.; Peters, K.; Rainer, J.; Salek, R.M.; et al. The metaRbolomics Toolbox in Bioconductor and beyond. *Metabolites* **2019**, *9*, 200. [CrossRef]
35. 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.; et al. Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211–221. [CrossRef]
36. Salek, R.M.; Steinbeck, C.; Viant, M.R.; Goodacre, R.; Dunn, W.B. The role of reporting standards for metabolite annotation and identification in metabolomic studies. *GigaScience* **2013**, *2*, 13. [CrossRef] [PubMed]
37. Haug, K.; Cochrane, K.; Nainala, V.C.; Williams, M.; Chang, J.; Jayaseelan, K.V.; O'dOnovan, C. MetaboLights: A resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **2019**, *48*, D440–D444. [CrossRef] [PubMed]
38. Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K.S.; et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **2015**, *44*, D463–D470. [CrossRef]
39. Gromski, P.S.; Muhamadali, H.; Ellis, D.I.; Xu, Y.; Correa, E.; Turner, M.L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant analysis—A marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. [CrossRef] [PubMed]
40. Liebal, U.W.; Phan, A.N.T.; Sudhakar, M.; Raman, K.; Blank, L.M. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites* **2020**, *10*, 243. [CrossRef]
41. Domingo-Almenara, X.; Guijas, C.; Billings, E.; Montenegro-Burke, J.R.; Uritboonthai, W.; Aisporna, A.E.; Chen, E.; Benton, H.P.; Siuzdak, G. The METLIN small molecule dataset for machine learning-based retention time prediction. *Nat. Commun.* **2019**, *10*, 5811. [CrossRef]
42. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523–526. [CrossRef]

43. 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.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
44. Spicer, R.; Salek, R.M.; Moreno, P.; Cañueto, D.; Steinbeck, C. Navigating freely-available software tools for metabolomics analysis. *Metabolomics* **2017**, *13*, 106. [CrossRef]
45. Dekermanjian, J.; Labeikovsky, W.; Ghosh, D.; Kechris, K. MSCAT: A Machine Learning Assisted Catalog of Metabolomics Software Tools. *Metabolites* **2021**, *11*, 678. [CrossRef]
46. Castelli, F.A.; Rosati, G.; Moguet, C.; Fuentes, C.; Marrugo-Ramírez, J.; Lefebvre, T.; Volland, H.; Merkoçi, A.; Simon, S.; Fenaille, F.; et al. Metabolomics for personalized medicine: The input of analytical chemistry from biomarker discovery to point-of-care tests. *Anal. Bioanal. Chem.* **2021**, *414*, 759–789. [CrossRef]
47. Smith, C.A.; Want, E.J.; 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**, *78*, 779–787. [CrossRef] [PubMed]
48. Tautenhahn, R.; Patti, G.J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84*, 5035–5039. [CrossRef]
49. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visual-izing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]
50. Chong, J.; Wishart, D.S.; Xia, J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr. Protoc. Bioinform.* **2019**, *68*, e86. [CrossRef]
51. Tautenhahn, R.; Böttcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform.* **2008**, *9*, 504. [CrossRef] [PubMed]
52. Broeckling, C.D.; Afsar, F.A.; Neumann, S.; Ben-Hur, A.; Prenni, J.E. RAMClust: A novel feature clustering method enables spectral-matching based annotation for metabolomics data. *Anal. Chem.* **2014**, *86*, 6812–6817.
53. Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T.R.; Neumann, S. CAMERA: An integrated strategy for com-pound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* **2012**, *84*, 283–289. [CrossRef]
54. Lommen, A. MetAlign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spec-trometry data preprocessing. *Anal. Chem.* **2009**, *81*, 3079–3086. [CrossRef]
55. Mak, T.D.; Laiakis, E.C.; Goudarzi, M.; Fornace, A.J. MetaboLyzer: A Novel Statistical Workflow for Analyzing Postprocessed LC–MS Metabolomics Data. *Anal. Chem.* **2013**, *86*, 506–513. [CrossRef] [PubMed]
56. Sorokina, M.; Steinbeck, C. Review on natural products databases: Where to find data in 2020. *J. Cheminform.* **2020**, *12*, 20. [CrossRef] [PubMed]
57. Xue, J.; Guijas, C.; Benton, H.P.; Warth, B.; Siuzdak, G. METLIN MS2 molecular standards database: A broad chemical and biological resource. *Nat. Methods* **2020**, *17*, 953–954. [CrossRef]
58. Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P.A. Consecutive queries to assess biological correlation in NMR metabolomics: Performance of comprehensive search of multiplets over typical 1D 1H NMR database search. *J. Proteome Res.* **2020**, *19*, 2977–2988. [CrossRef]
59. Dudek, C.-A.; Reuse, C.; Fuchs, R.; Hendriks, J.; Starck, V.; Hiller, K. MIAMI—-a tool for non-targeted detection of metabolic flux changes for mode of action identification. *Bioinformatics* **2020**, *36*, 3925–3926. [CrossRef]
60. Quiroz-Moreno, C.; Furlan, M.F.; Belinato, J.R.; Augusto, F.; Alexandrino, G.L.; Mogollón, N.G.S. RGCxGC toolbox: An R-package for data processing in comprehensive two-dimensional gas chromatography-mass spectrometry. *Microchem. J.* **2020**, *156*, 104830. [CrossRef]
61. Kouřil, Š.; de Sousa, J.; Václavík, J.; Friedecký, D.; Adam, T. CROP: Correlation-based reduction of feature multiplicities in untargeted metabolomic data. *Bioinformatics* **2020**, *36*, 2941–2942. [CrossRef]
62. Wu, C.-T.; Wang, Y.; Wang, Y.; Ebbels, T.; Karaman, I.; Graça, G.; Pinto, R.; Herrington, D.M.; Wang, Y.; Yu, G. Targeted realignment of LC-MS profiles by neighbor-wise compound-specific graphical time warping with misalignment detection. *Bioinformatics* **2020**, *36*, 2862–2871. [CrossRef] [PubMed]
63. Riquelme, G.; Zabalegui, N.; Marchi, P.; Jones, C.M.; Monge, M.E. A Python-Based Pipeline for Preprocessing LC–MS Data for Untargeted Metabolomics Workflows. *Metabolites* **2020**, *10*, 416. [CrossRef]
64. Seyler, L.; Kujawinski, E.B.; Azua-Bustos, A.; Lee, M.D.; Marlow, J.; Perl, S.M.; Ii, H.J.C. Metabolomics as an Emerging Tool in the Search for Astrobiologically Relevant Biomarkers. *Astrobiology* **2020**, *20*, 1251–1261. [CrossRef] [PubMed]
65. Wajid, B.; Iqbal, H.; Jamil, M.; Rafique, H.; Anwar, F. MetumpX—A metabolomics support package for untargeted mass spectrometry. *Bioinformatics* **2020**, *36*, 1647–1648. [CrossRef] [PubMed]
66. Kuhring, M.; Eisenberger, A.; Schmidt, V.; Kränkel, N.; Leistner, D.M.; Kirwan, J.; Beule, D. Concepts and Software Package for Efficient Quality Control in Targeted Metabolomics Studies: MeTaQuaC. *Anal. Chem.* **2020**, *92*, 10241–10245. [CrossRef]

67. Bararpour, N.; Gilardi, F.; Carmeli, C.; Sidibe, J.; Ivanisevic, J.; Caputo, T.; Augsburger, M.; Grabherr, S.; Desvergne, B.; Guex, N.; et al. Visualization and normalization of drift effect across batches in metabolome-wide association studies. *bioRxiv* **2020**. [CrossRef]
68. Chetnik, K.; Petrick, L.; Pandey, G. MetaClean: A machine learning-based classifier for reduced false positive peak detection in untargeted LC–MS metabolomics data. *Metabolomics* **2020**, *16*, 117. [CrossRef]
69. Liu, Y.; Mrzic, A.; Meysman, P.; De Vijlder, T.; Romijn, E.P.; Valkenborg, D.; Bittremieux, W.; Laukens, K. MESSAR: Automated recommendation of metabolite substructures from tandem mass spectra. *PLoS ONE* **2020**, *15*, e0226770. [CrossRef]
70. Liu, K.H.; Nellis, M.; Uppal, K.; Ma, C.; Tran, V.; Liang, Y.; Walker, D.I.; Jones, D.P. Reference Standardization for Quantification and Harmonization of Large-Scale Metabolomics. *Anal. Chem.* **2020**, *92*, 8836–8844. [CrossRef]
71. Dührkop, K.; Nothias, L.F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M.A.; Petras, D.; Gerwick, W.H.; Rousu, J.; Dorrestein, P.C.; et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat. Biotechnol.* **2021**, *39*, 462–471. [CrossRef]
72. Fan, Z.; Alley, A.; Ghaffari, K.; Ressom, H.W. MetFID: Artificial neural network-based compound fingerprint prediction for metabolite annotation. *Metabolomics* **2020**, *16*, 104. [CrossRef]
73. Luan, H.; Jiang, X.; Ji, F.; Lan, Z.; Cai, Z.; Zhang, W. CPVA: A web-based metabolomic tool for chromatographic peak visualization and annotation. *Bioinformatics* **2020**, *36*, 3913–3915. [CrossRef] [PubMed]
74. Ricart, E.; Pupin, M.; Müller, M.; Lisacek, F. Automatic annotation and dereplication of tandem mass spectra of peptidic natural products. *Anal. Chem.* **2020**, *92*, 15862–15871. [CrossRef]
75. Choudhary, K.S.; Fahy, E.; Coakley, K.; Sud, M.; Maurya, M.R.; Subramaniam, S. MetENP/MetENPWeb: An R package and web application for metabolomics enrichment and pathway analysis in metabolomics workbench. *bioRxiv* **2020**. Available online: https://www.biorxiv.org/content/10.1101/2020.11.20.391912v1 (accessed on 15 October 2023). [CrossRef]
76. Cao, L.; Guler, M.; Tagirdzhanov, A.; Lee, Y.-Y.; Gurevich, A.; Mohimani, H. MolDiscovery: Learning mass spectrometry fragmentation of small molecules. *Nat. Commun.* **2021**, *12*, 3718. [CrossRef]
77. Delcourt, V.; Barnabé, A.; Loup, B.; Garcia, P.; André, F.; Chabot, B.; Trévisiol, S.; Moulard, Y.; Popot, M.-A.; Bailly-Chouriberry, L. MetIDfyR: An Open-Source R Package to Decipher Small-Molecule Drug Metabolism through High-Resolution Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 13155–13162. [CrossRef]
78. Tripathi, A.; Vázquez-Baeza, Y.; Gauglitz, J.M.; Wang, M.; Dührkop, K.; Nothias-Esposito, M.; Acharya, D.D.; Ernst, M.; van der Hooft, J.J.J.; Zhu, Q.; et al. Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat. Chem. Biol.* **2020**, *17*, 146–151. [CrossRef]
79. Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A.T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nat. Commun.* **2021**, *12*, 3832. [CrossRef] [PubMed]
80. Castellano-Escuder, P.; González-Domínguez, R.; Wishart, D.S.; Andrés-Lacueva, C.; Sánchez-Pla, A. FOBI: An ontology to represent food intake data and associate it with metabolomic data. *Database* **2020**, *2020*, baaa033. [CrossRef] [PubMed]
81. Rawlinson, C.; Jones, D.; Rakshit, S.; Meka, S.; Moffat, C.S.; Moolhuijzen, P. Hierarchical clustering of MS/MS spectra from the firefly metabolome identifies new lucibufagin compounds. *Sci. Rep.* **2020**, *10*, 6043. [CrossRef]
82. Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat. Commun.* **2020**, *11*, 4334. [CrossRef]
83. Kachman, M.; Habra, H.; Duren, W.; Wigginton, J.; Sajjakulnukit, P.; Michailidis, G.; Burant, C.; Karnovsky, A. Deep annotation of untargeted LC-MS metabolomics data with *Binner*. *Bioinformatics* **2019**, *36*, 1801–1806. [CrossRef]
84. Fraisier-Vannier, O.; Chervin, J.; Cabanac, G.; Puech, V.; Fournier, S.; Durand, V.; Amiel, A.; André, O.; Benamar, O.A.; Dumas, B.; et al. MS-CleanR: A feature-filtering workflow for untargeted LC–MS based metabolomics. *Anal. Chem.* **2020**, *92*, 9971–9981. [CrossRef]
85. Bonini, P.; Kind, T.; Tsugawa, H.; Barupal, D.K.; Fiehn, O. Retip: Retention time prediction for compound annotation in untargeted metabolomics. *Anal. Chem.* **2020**, *92*, 7515–7522. [CrossRef] [PubMed]
86. Naylor, B.C.; Catrow, J.L.; Maschek, J.A.; Cox, J.E. QSRR automator: A tool for automating retention time prediction in lipidomics and metabolomics. *Metabolites* **2020**, *10*, 237. [CrossRef]
87. Schum, S.K.; Brown, L.E.; Mazzoleni, L.R. MFAssignR: Molecular formula assignment software for ultrahigh resolution mass spectrometry analysis of environmental complex mixtures. *Environ. Res.* **2020**, *191*, 110114. [CrossRef] [PubMed]
88. Xing, S.; Hu, Y.; Yin, Z.; Liu, M.; Tang, X.; Fang, M.; Huan, T. Retrieving and Utilizing Hypothetical Neutral Losses from Tandem Mass Spectra for Spectral Similarity Analysis and Unknown Metabolite Annotation. *Anal. Chem.* **2020**, *92*, 14476–14483. [CrossRef] [PubMed]
89. Jarmusch, A.K.; Wang, M.; Aceves, C.M.; Advani, R.S.; Aguirre, S.; Aksenov, A.A.; Aleti, G.; Aron, A.T.; Bauermeister, A.; Bolleddu, S.; et al. ReDU: A framework to find and reanalyze public mass spectrometry data. *Nat. Methods* **2020**, *17*, 901–904. [CrossRef]

90. Wang, M.; Jarmusch, A.K.; Vargas, F.; Aksenov, A.A.; Gauglitz, J.M.; Weldon, K.; Petras, D.; da Silva, R.; Quinn, R.; Melnik, A.V.; et al. Mass spectrometry searches using MASST. *Nat. Biotechnol.* **2020**, *38*, 23–26. [CrossRef]
91. Kim, H.W.; Wang, M.; Leber, C.; Nothias, L.-F.; Reher, R.; Kang, K.B.; Hooft, J.J.J.; Dorrestein, P.; Gerwick, W.; Cottrell, G. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J. Nat. Prod.* **2021**, *84*, 2795–2807. [CrossRef] [PubMed]
92. Kim, T.; Tang, O.; Vernon, S.T.; Kott, K.A.; Koay, Y.C.; Park, J.; James, D.E.; Grieve, S.M.; Speed, T.P.; Yang, P.; et al. A hierarchical approach to removal of unwanted variation for large-scale metabolomics data. *Nat. Commun.* **2021**, *12*, 4992. [CrossRef]
93. Helmus, R.; ter Laak, T.L.; van Wezel, A.P.; de Voogt, P.; Schymanski, E.L. patRoon: Open source software platform for environmental mass spectrometry based non-target screening. *J. Cheminform.* **2021**, *13*, 1. [CrossRef] [PubMed]
94. Ni, Z.; Fedorova, M. LipidLynxX: A data transfer hub to support integration of large scale lipidomics datasets. *bioRxiv* **2020**. [CrossRef]
95. Wolthuis, J.C.; Magnusdottir, S.; Pras-Raves, M.; Moshiri, M.; Jans, J.J.M.; Burgering, B.; van Mil, S.; de Ridder, J. MetaboShiny: Interactive analysis and metabolite annotation of mass spectrometry-based metabolomics data. *Metabolomics* **2020**, *16*, 99. [CrossRef]
96. Kutuzova, S.; Colaianni, P.; Röst, H.; Sachsenberg, T.; Alka, O.; Kohlbacher, O.; Burla, B.; Torta, F.; Schrübbers, L.; Kristensen, M.; et al. SmartPeak Automates Targeted and Quantitative Metabolomics Data Processing. *Anal. Chem.* **2020**, *92*, 15968–15974. [CrossRef]
97. Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; et al. A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.* **2020**, *38*, 1159–1163. [CrossRef]
98. Liang, D.; Liu, Q.; Zhou, K.; Jia, W.; Xie, G.; Chen, T. IP4M: An integrated platform for mass spectrometry-based metabolomics data mining. *BMC Bioinform.* **2020**, *21*, 444. [CrossRef]
99. Rosa, T.R.; Folli, G.S.; Pacheco, W.L.S.; Castro, M.P.; Romão, W.; Filgueiras, P.R. DropMS: Petroleomics Data Treatment Based in Web Server for High-Resolution Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1483–1490. [CrossRef]
100. Mohamed, A.; Molendijk, J.; Hill, M.M. Lipidr: A software tool for data mining and analysis of lipidomics datasets. *J. Proteome Res.* **2020**, *19*, 2890–2897. [CrossRef]
101. Yang, Q.; Wang, Y.; Zhang, Y.; Li, F.; Xia, W.; Zhou, Y.; Qiu, Y.; Li, H.; Zhu, F. NOREVA: Enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res.* **2020**, *48*, W436–W448. [CrossRef]
102. Manjarin, R.; Maj, M.A.; la Frano, M.R.; Glanz, H. % polynova_2way: A SAS macro for implementation of mixed models for metabolomics data. *PLoS ONE* **2020**, *15*, e0244013. [CrossRef]
103. Kockmann, T.; Panse, C. rawR-Direct access to raw mass spectrometry data in R. *bioRxiv* **2020**. [CrossRef]
104. Berg, J.A.; Zhou, Y.; Ouyang, Y.; Waller, T.C.; Cluntun, A.A.; Conway, M.E.; Nowinski, S.M.; Van Ry, T.; George, I.; Cox, J.E.; et al. Network-aware reaction pattern recognition reveals regulatory signatures of mitochondrial dysfunction. *bioRxiv* **2021**. [CrossRef]
105. Henning, J.; Smith, R. A web-based system for creating, viewing, and editing precursor mass spectrometry ground truth data. *BMC Bioinform.* **2020**, *21*, 418. [CrossRef] [PubMed]
106. Linck, E.J.G.; Richmond, P.A.; Tarailo-Graovac, M.; Engelke, U.; Kluijtmans, L.A.J.; Coene, K.L.M.; Wevers, R.A.; Wasserman, W.; van Karnebeek, C.D.M.; Mostafavi, S. metPropagate: Network-guided propagation of metabolomic information for prioritization of metabolic disease genes. *npj Genom. Med.* **2020**, *5*, 25. [CrossRef]
107. Ross, D.H.; Cho, J.H.; Zhang, R.; Hines, K.M.; Xu, L. LiPydomics: A Python Package for Comprehensive Prediction of Lipid Collision Cross Sections and Retention Times and Analysis of Ion Mobility-Mass Spectrometry-Based Lipidomics Data. *Anal. Chem.* **2020**, *92*, 14967–14975. [CrossRef]
108. Capellades, J.; Junza, A.; Samino, S.; Brunner, J.S.; Schabbauer, G.; Vinaixa, M.; Yanes, O. Exploring the Use of Gas Chromatography Coupled to Chemical Ionization Mass Spectrometry (GC-CI-MS) for Stable Isotope Labeling in Metabolomics. *Anal. Chem.* **2020**, *93*, 1242–1248. [CrossRef]
109. Peng, B.; Kopczynski, D.; Pratt, B.S.; Ejsing, C.S.; Burla, B.; Hermansson, M.; Benke, P.I.; Tan, S.H.; Chan, M.Y.; Torta, F.; et al. LipidCreator workbench to probe the lipidomic landscape. *Nat. Commun.* **2020**, *11*, 2057. [CrossRef]
110. Koelmel, J.P.; Li, X.; Stow, S.M.; Sartain, M.J.; Murali, A.; Kemperman, R.; Tsugawa, H.; Takahashi, M.; Vasiliou, V.; Bowden, J.A.; et al. Lipid Annotator: Towards Accurate Annotation in Non-Targeted Liquid Chromatography High-Resolution Tandem Mass Spectrometry (LC-HRMS/MS) Lipidomics Using a Rapid and User-Friendly Software. *Metabolites* **2020**, *10*, 101. [CrossRef]
111. Iakab, S.A.; Sementé, L.; García-Altares, M.; Correig, X.; Ràfols, P. Raman2imzML converts Raman imaging data into the standard mass spectrometry imaging format. *BMC Bioinform.* **2020**, *21*, 448. [CrossRef]
112. Huang, L.; Currais, A.; Shokhirev, M.N. SUMMER, a shiny utility for metabolomics and multiomics exploratory research. *Metabolomics* **2020**, *16*, 126. [CrossRef]
113. Bahado-Singh, R.O.; Yilmaz, A.; Bisgin, H.; Turkoglu, O.; Kumar, P.; Sherman, E.; Mrazik, A.; Odibo, A.; Graham, S.F. Artificial intelligence and the analysis of multi-platform metabolomics data for the detection of intrauterine growth restriction. *PLoS ONE* **2019**, *14*, e0214121. [CrossRef] [PubMed]

114. Irajizad, E.; Wu, R.; Vykoukal, J.; Murage, E.; Spencer, R.; Dennison, J.B.; Moulder, S.; Ravenberg, E.; Lim, B.; Litton, J.; et al. Application of Artificial Intelligence to Plasma Metabolomics Profiles to Predict Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Front. Artif. Intell.* **2022**, *5*, 876100. [CrossRef]
115. Moin, A.S.M.; Sathyapalan, T.; Atkin, S.L.; Butler, A.E. Diagnostic and Prognostic Protein Biomarkers of β-Cell Function in Type 2 Diabetes and Their Modulation with Glucose Normalization. *Metabolites* **2022**, *12*, 196. [CrossRef]
116. Govender, M.A.; Brandenburg, J.T.; Fabian, J.; Ramsay, M. The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review. *Front. Genet.* **2021**, *12*, 682929. [CrossRef]
117. Lehallier, B.; Gate, D.; Schaum, N.; Nanasi, T.; Lee, S.E.; Yousef, H.; Losada, P.M.; Berdnik, D.; Keller, A.; Verghese, J.; et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **2019**, *25*, 1843–1850. [CrossRef] [PubMed]
118. Williams, S.A.; Kivimaki, M.; Langenberg, C.; Hingorani, A.D.; Casas, J.P.; Bouchard, C.; Jonasson, C.; Sarzynski, M.A.; Shipley, M.J.; Alexander, L.; et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **2019**, *25*, 1851–1857. [CrossRef]
119. Ni, Z.; Wölk, M.; Jukes, G.; Espinosa, K.M.; Ahrends, R.; Aimo, L.; Alvarez-Jarreta, J.; Andrews, S.; Andrews, R.; Bridge, A.; et al. Guiding the choice of informatics software and tools for lipidomics research applications. *Nat. Methods* **2022**, *20*, 193–204. [CrossRef]
120. Ali, A.; Abouleila, Y.; Shimizu, Y.; Hiyama, E.; Emara, S.; Mashaghi, A.; Hankemeier, T. Single-cell metabolomics by mass spectrometry: Advances, challenges, and future applications. *TrAC Trends Anal. Chem.* **2019**, *120*, 115436. [CrossRef]
121. Scalbert, A.; Gunter, M.; Albanes, D.; Eliassen, A.H.; Krycer, J.R. 3.7 Metabolic change and metabolomics: Emerging approaches and new insights. In *World Cancer Report: Cancer Research for Cancer Prevention*; Wild, C.P., Weiderpass, E., Stewart, B.W., Eds.; International Agency for Research on Cancer: Lyon, France, 2020. Available online: https://www.ncbi.nlm.nih.gov/books/NBK606501/?utm_source=chatgpt.com (accessed on 15 September 2023).
122. Barberis, E.; Khoso, S.; Sica, A.; Falasca, M.; Gennari, A.; Dondero, F.; Afantitis, A.; Manfredi, M. Precision Medicine Approaches with Metabolomics and Artificial Intelligence. *Int. J. Mol. Sci.* **2022**, *23*, 11269. [CrossRef] [PubMed]
123. Qin, N.; Qin, M.; Shi, W.; Kong, L.; Wang, L.; Xu, G.; Guo, Y.; Zhang, J.; Ma, Q. Investigation of pathogenesis of hyperuricemia based on untargeted and targeted metabolomics. *Sci. Rep.* **2022**, *12*, 13980. [CrossRef]
124. Tian, L.; Yu, T. An integrated deep learning framework for the interpretation of untargeted metabolomics data. *Brief. Bioinform.* **2023**, *24*, bbad244. [CrossRef]
125. Viant, M.R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943–948. [CrossRef]
126. Ahmed, Z.; Wan, S.; Zhang, F.; Zhong, W. Artificial intelligence for omics data analysis. *BMC Methods* **2024**, *1*, 4. [CrossRef]
127. Zulfiqar, M.; Gadelha, L.; Steinbeck, C.; Sorokina, M.; Peters, K. MAW: The reproducible Metabolome Annotation Workflow for untargeted tandem mass spectrometry. *J. Cheminform.* **2023**, *15*, 32. [CrossRef]
128. Ouhmouk, M.; Baichoo, S.; Abik, M. Challenges in AI-driven multi-omics data analysis for oncology: Addressing dimensionality, sparsity, transparency and ethical considerations. *Inform. Med. Unlocked* **2025**, *57*, 101679. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.