



Review

Toward a Standardized Strategy of Clinical Metabolomics for the Advancement of Precision Medicine

Nguyen Phuoc Long ^{1,†} , Tran Diem Nghi ^{2,†}, Yun Pyo Kang ³, Nguyen Hoang Anh ¹, Hyung Min Kim ¹, Sang Ki Park ² and Sung Won Kwon ^{1,*} 

¹ College of Pharmacy, Seoul National University, Seoul 08826, Korea; phuoclong@snu.ac.kr (N.P.L.); 2018-23140@snu.ac.kr (N.H.A.); snuhmkim04@snu.ac.kr (H.M.K.)

² Department of Life Sciences, Pohang University of Science and Technology, Pohang 790-784, Korea; diemngi0208@postech.ac.kr (T.D.N.); skpark@postech.ac.kr (S.K.P.)

³ Department of Cancer Physiology, Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA; yunpyo.kang@moffitt.org

* Correspondence: swkwon@snu.ac.kr; Tel.: +82-2-880-7844

† These authors contributed equally to this work.

Received: 4 December 2019; Accepted: 21 January 2020; Published: 29 January 2020



Abstract: Despite the tremendous success, pitfalls have been observed in every step of a clinical metabolomics workflow, which impedes the internal validity of the study. Furthermore, the demand for logistics, instrumentations, and computational resources for metabolic phenotyping studies has far exceeded our expectations. In this conceptual review, we will cover inclusive barriers of a metabolomics-based clinical study and suggest potential solutions in the hope of enhancing study robustness, usability, and transferability. The importance of quality assurance and quality control procedures is discussed, followed by a practical rule containing five phases, including two additional “pre-pre-” and “post-post-” analytical steps. Besides, we will elucidate the potential involvement of machine learning and demonstrate that the need for automated data mining algorithms to improve the quality of future research is undeniable. Consequently, we propose a comprehensive metabolomics framework, along with an appropriate checklist refined from current guidelines and our previously published assessment, in the attempt to accurately translate achievements in metabolomics into clinical and epidemiological research. Furthermore, the integration of multifaceted multi-omics approaches with metabolomics as the pillar member is in urgent need. When combining with other social or nutritional factors, we can gather complete omics profiles for a particular disease. Our discussion reflects the current obstacles and potential solutions toward the progressing trend of utilizing metabolomics in clinical research to create the next-generation healthcare system.

Keywords: adaptive metabolomics; lipidomics; multi-omics; precision medicine; systems biology; machine learning

1. Introduction

Metabolomics offers a pragmatic and robust framework for the comprehensive measurement and identification of the endogenous and exogenous low-molecular-weight metabolites in biological systems [1]. Together with other omics platforms, e.g., genomics and proteomics, metabolomics offers a systematic assessment of the interactions between genetic variations, central metabolism, and environmental exposures to explain the initiation and progression of a disease [2]. It is understood that the causation in systems biology should determine how and why a process occurs in a certain way rather than identify separate molecular components [3]. In this regard, measuring and modeling the

metabolome can shed light on the pathophysiological mechanisms and provide an essential piece of information for precision medicine.

In clinical medicine and epidemiology, together with the development of machine learning and artificial intelligence, omics analyses empower systems biology-based methods for precision health monitoring and treatment [4]. Typical applications of omics technologies in precision medicine encompass the assessment of biomarker panel to estimate disease risk, diagnose disease, quantify progression, and optimize treatment strategies [5]. Even so, the achievements of metabolomics in large-scale research are somewhat limited [6–10]. Up to the present time, a significant part of the metabolomics-based epidemiological and clinical studies is cross-sectional or non-prospective case-control, which compares the metabolic phenotypes of participants by disease or exposure status [11,12]. These study designs are intricate in establishing the temporal precedence and based on prevalent rather than incident cases; thus, they are rather the preliminary step of the process of diagnostic biomarker discovery. The majority of diagnostic biomarker discovery studies are conducted with the lack of external validation, leading to the high false-positive rates and insufficient data on the predictive value of biomarkers in pre-diagnostic samples [13]. The commonly applied strategies of metabolomics biomarkers in clinical and epidemiological studies are not adequate to provide sufficient predictive evidence to be clinically useful, as observed in the study aiming at identifying biomarkers for early pancreatic cancer diagnosis [14]. Other clinical fields, such as obstetrics, are also in urgent need of large-scale studies using a well-defined sampling strategy to determine predictive biomarkers for preterm birth given the current heterogeneity in methodologies and identification methods across different studies [15].

The Consortium of Metabolomics Studies has discussed experience gained from current epidemiological metabolomics studies worldwide and suggested guidelines concerning important aspects of a large-scale metabolomics study [16]. A recently published review from the Consortium of Metabolomics Studies Statistics Working Group delineated a framework involving study and bio-sampling design for studies that integrate metabolomics with other omics data as well as considerations on multi-omics data processing techniques at the population level [17]. Extending the boundaries of these perspectives, our review will cover all-inclusive barriers in a metabolomics study at the clinical and epidemiological scale, primarily focusing on mass spectrometry (MS)-based approaches. We will investigate analytical and computational challenges, thus proving the need for the establishment of a standardized biobank, quality assurance (QA) and quality control (QC) standards, the establishment of an automated system for sample handling and data acquisition, standards for metabolite and lipid identification, as well as the usability of machine learning. We will provide several discussions on fundamental requirements on standards for reporting metabolomics studies at the clinical and epidemiological scale. Significantly, automated systems of machine learning and artificial intelligence platforms are predicted to be developed soon and integrated into various steps of a metabolomics study [18]. Machine learning methods such as random forests and deep neural networks, which are reliable in prediction, have a “black box” nature, causing a substantial debate, the so-called “cascade of black box data manipulations, processing, and analyses” in omics research. The trade-off between model accuracy and interpretability can be regarded as an emerging crisis in the field. Collectively, we propose a comprehensive and adaptive multi-omics workflow demonstrating the indispensable position of metabolomics in the omics universe along with a proper reporting guideline. This practice, if deeply implemented, may potentially extend the translation of omics science into clinical settings for the next-generation healthcare system. Additionally, we have tried to raise current concerns that are open for active discussion throughout the paper.

2. Representative Achievements of Metabolomics in Public Health and Clinical Research

Following emerging instrumental and logistics technologies, analytical methods, and computational tools, research on the application of hydrophilic and lipophilic metabolite biomarkers in epidemiological studies has been increasing over the last decade, particularly on cancer, cardiovascular diseases, and diabetes, but it is still relatively sparse compared with other omics approaches [8]. MS is the primary

analytical method on which the emerging omics fields such as proteomics and metabolomics are based. Owing to its potential as a multi-platform analytical method and its diverse applicability from diagnosis and prognosis to the discovery of novel biomarkers for targeted therapies, MS has been routinely applied in omics studies, especially cancer and metabolism-related omics research, for a better insight into human diseases [19–26]. Table 1 displays the main findings of some major studies among the initial achievements of the field.

Table 1. Recent representative achievements of metabolomics in public health and clinical research.

| Country | Year | Metabolomics/ Lipidomics/ Multi-omics | Clinical Field | Platform | Main Conclusions | References |
|--------------------|------|---|--------------------------|-------------------|---|----------------------|
| The United Kingdom | 2011 | Steroidomics | Cancer | GC-MS | This study combined population-based steroidomics research with machine learning analysis, aiming at identifying urine steroid biomarkers for differentiating adrenocortical carcinoma from benign adenoma. | Arlt et al. [27] |
| The United States | 2011 | Metabolomics | Diabetes | LC-MS | A large nested group of 2422 normoglycemic subjects in the Framingham Offspring Study was followed for 12 years to help identify a panel of three amino acids (isoleucine, phenylalanine, and tyrosine) which might serve as novel predictors for future diabetes. | Wang et al. [28] |
| The Netherlands | 2015 | Steroidomics | Cancer | GC-MS | Urinary steroid signatures were established for the discrimination of adrenal cortical carcinoma from other adrenal conditions. | Kerkhofs et al. [29] |
| The United Kingdom | 2015 | Metabolomics | Physiological state | GC-MS and LC-MS | The Husermet project has applied untargeted MS to investigate the comprehensive hydrophilic and lipophilic metabolome of serum biospecimen obtained from the phenotyping of 1200 healthy subjects. | Dunn et al. [30] |
| The United States | 2015 | Steroidomics | Cancer | LC-MS | The novel LC-MS/MS assay used in this study enabled the examination of all estrogen metabolites for epidemiological and clinical research on hormone-related diseases. | Ziegler et al. [31] |
| Multinational | 2018 | Multi-omics | Diabetes | N/A | The Environmental Determinants of Diabetes in the Young (TEDDY) study followed more than 12,000 children to explore metabolic pathways related to type 1 diabetes. | Rewers et al. [32] |
| Multinational | 2018 | Multi-omics | Obesity | NMR and LC-MS | Along with fecal metagenomes, plasma and urine metabolome revealed molecular pathways uniting the gut microbiome and the human phenome to hepatic steatosis in two cohorts of non-diabetic obese women in the FLORINASH consortium. | Hoyles et al. [33] |
| Japan | 2018 | Metabolomics | Physiological state | CE-MS (and LC-MS) | The Tsuruoka large-scale cohort study has gathered plasma metabolomics data from more than 10,000 individuals as an innovative model for preventive medicine. | Harada et al. [34] |
| Multinational | 2019 | Metabolomics | Long-term mortality risk | NMR | From the metabolic profile of 44,168 individuals, 14 biomarkers have been associated with all-cause mortality. The combination of this set of metabolite predictors and sex considerably improves mortality risk prediction compared to traditional risk factors such as age, body mass index, systolic blood pressure, or total cholesterol. | Deelen et al. [35] |
| The United Kingdom | 2019 | Metabolomics | Parkinson's disease | TD-GC-MS | This study applied an unbiased method, taking the use of volatile sebum metabolites to diagnose Parkinson's disease. | Trivedi et al. [36] |
| The United Kingdom | 2019 | Metabolomics | Cardiovascular disease | NMR | This study with over 7000 participants on the metabolic profile of atherosclerosis revealed that this condition was associated with perturbations of multiple interconnected pathways related to lipid, fatty acid, and amino acid metabolism, and displayed a considerably similar model between coronary and carotid atherosclerosis. | Tzoulaki et al. [37] |
| The United States | 2019 | Metabolomics | Obesity | LC-MS | This study revealed the disturbance of the metabolome in obese versus healthy individuals. Approximately a third of compounds followed changes in body mass index, suggesting the role of metabolome profiling in participant recruitment for clinical trials related to obesity. | Cirulli et al. [38] |
| The United States | 2019 | Metabolomics | Diabetes | LC-MS | A total of 69 out of 331 plasma metabolites extracted from more than 2000 samples from the Diabetes Prevention Program were linked to type 2 diabetes regardless of treatment randomization. | Chen et al. [39] |
| Sweden | 2019 | Metabolomics | Physiological state | LC-MS | Not just stopping at the cross-sectional quantification of urinary eicosanoid metabolites, this recently published study extended the scope by focusing on the long-term repeatability and stability of the method used in order to serve the large-scale analysis of multiple cohorts. | Gomez et al. [40] |

The lack of independent validation cohorts remains challenging in most early metabolomics studies at population-based level, which is gradually being improved in later large-scale projects [41–44]. Prospective longitudinal studies with larger sample size, longer follow-up period, and more efficient

nested design (e.g., nested case-control, case-cohort, and clinical trials) have been conducted recently to assist the biomarker establishment [11,45–47]. Despite being cost- and time-consuming, prospective study designs may serve both biomarker discovery and validation purposes and can be equipped with a standardized biobank as a platform for biomarker research to overcome the limitations of cross-sectional studies [13]. Future research can benefit from this model to provide clinically robust results and eventually help innovate the healthcare system.

3. Analytical and Computational Challenges in Clinical Metabolomics and Lipidomics Studies

The general workflow of untargeted metabolomics research is described in Figure 1. Furthermore, large-scale multi-site studies require stringent awareness of systematic and random errors, batch variation, differences between analytical instruments, heterogeneity across centers regarding methods for storage, handling, and transportation [48]. We have made attempts to inspect and address these issues throughout this section sequentially. Particular challenges of lipidomics will also be discussed in detail.

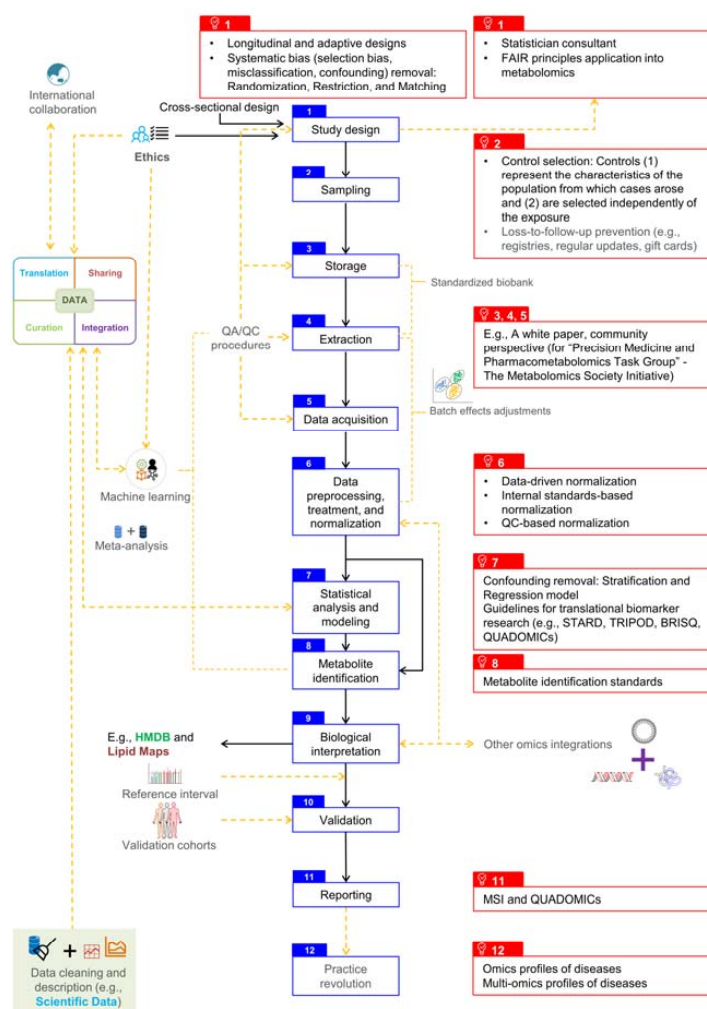


Figure 1. The standard untargeted metabolomics workflow at clinical and epidemiological scale. Blue boxes constitute the backbone of the workflow. Black arrows indicate current standards. Orange dashed arrows imply that the issues have been raised but have not received enough attention or have lacked communication. Red boxes represent current and/or potential solutions to these issues. BRISQ: Biospecimen Reporting for Improved Study Quality. FAIR: Findability, Accessibility, Interoperability, and Reusability. MSI: Metabolomics Standards Initiative. QUADOMICS: Quality Assessment of studies on the diagnostic accuracy of OMICs-based technologies. STARD: Standards for the Reporting of Diagnostic Accuracy Studies. TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis.

3.1. The Need for Standardized Biobank Establishment

The preanalytical steps in sample acquisition and processing are among the most decisive for producing reliable analytical results [49]. Currently, it is challenging to collect and optimally handle samples because of the collaboration of multiple research laboratories and long shipping to main hubs and freezers [50]. It is also of importance to emphasize that the number of biobanks worldwide has experienced a rapid escalation, giving rise to the issue of interbiobank variability [51]. The majority of biobanks established for epidemiological purposes benefit either genomics or more traditional epidemiologic analyses, whereas the collection of DNA requires profoundly different criteria than that of metabolites. The metabolome is dynamically changing, and metabolism can continue in tissues or biofluids after collection even when being carried on with great care [49]. Collectively, the establishment of biobanks dedicated to metabolomics along with other omics is currently needed, and harmonization in the methods of sample collection, storage, analytical methods protocols should be defined within the study design to secure the requirement of reproducible research as well as reduce the cost of future research. Significant attempts have been made by International Agency for Research on Cancer Biobank, Promoting Harmonization of Epidemiological Biobanks in Europe, and Biobanking and Biomolecular Resources Research Infrastructure, for instance, on the way to reach a standardized, unified, fully validated protocol where all steps have been tested for reproducibility and robustness [52–54]. With that same goal, recommendations on the processing and biobanking of serum, plasma, tissues, urine, feces, as well as other biofluids have been discussed at great length in the paper of “Precision Medicine and Pharmacometabolomics Task Group” from the Metabolomics Standards Initiative (MSI) [49]. Furthermore, primary aspects of biobank harmonization have been comprehensively unveiled in this white paper.

Under adverse sample handling conditions and delayed freezing, only 62% of metabolites had a stable level and good to excellent concordance [55]. At a long-term epidemiological scale, however, the impact of different storage conditions on metabolite stability needs to be more profoundly examined since the plasma concentrations of specific metabolites can be altered under five-year storage at $-80\text{ }^{\circ}\text{C}$ [56]. Potential solutions to ensure sample stability in the long run would be the exclusion or better adjustment coefficient of profoundly altered compounds prior to the establishment of biomarker panel, development of more accurate analytical techniques for unstable but biologically essential metabolites, development of better long-term storage strategies such as liquid nitrogen storage, and avoidance of repeated freeze-thaw cycles. Efforts have been recorded; for instance, the European Federation of Clinical Chemistry and Laboratory Medicine Working Group on Preanalytical Phase is currently working on some recommendations for a study focusing on sample stability to enable standardization and reproducibility [57]. Another problem appears in case each protocol of the biobank has a different influence on the metabolite changes. Thus, research for better recommendations on how samples should be interpreted may be a potential topic.

3.2. The High Demand for Quality Assurance and Quality Control Standards

Quality assurance (QA) and quality control (QC) are two fundamental quality management procedures ensuring the success of any omics studies. At clinical and epidemiological scale, QA activities in metabolomics study are established to guarantee that data are managed by good clinical and epidemiological practice, and that the study fulfills predetermined requirements for quality. QC activities are undertaken within the QA system to verify the fulfillment of all quality requirements [58]. Existing guidelines, e.g., “Guidance for Industry: Bioanalytical Method Validation” from the Food and Drug Administration, provide QA/QC processes, which may be well adapted to targeted and semi-targeted metabolomics assays. However, they were not designed with metabolomics in mind, not routinely applied, and not readily translated into untargeted metabolomics [58]. QA/QC protocols should be ready as early as at the hypothesis formation and experimental study design step. The implementation of different QA/QC processes has been reported for MS-based metabolomics and lipidomics yet varies in both fields [58–61]. The lack of well-established QA/QC procedures for untargeted metabolomics may

impede the harmonization across laboratories and multi-center studies [62]. Reducing heterogeneity across all sampling sites involves sample selection and preparation, constant cleaning of instruments, calibration, or maintenance, quality checking, data inspection at predetermined time points, and regular personnel training [48,63]. Furthermore, patient-specific variability should be raised and controlled instantly in the study design step, including, but not limited to, age, gender, ethnicity, body mass index, diet, circadian rhythms, smoking habits, physical activity, health status, and concomitant medications [64]. Taken as an example, excess exposure to glucocorticoids, which are drugs that are strictly associated with metabolic control, may affect plasma lipidome [65]. Experts in metabolomics and QA/QC procedures gathered in a meeting entitled “Think Tank on Quality Assurance and Quality Control for Untargeted Metabolomic Studies” to identify appropriate reporting standards of QA/QC activities for publications and databases, leading to the foundation of the metabolomics Quality Assurance and Quality Control Consortium [62,66]. Pioneering scientists among those who participated in the Think Tank meeting collaborated to publish guidelines that focused on the use of system suitability and QC samples in MS-based untargeted metabolomics to collect high-quality data [58]. Developing from the traditional three phases (pre-analytical, analytical, and post-analytical), we will initiate a model adopting two more steps “pre-pre-” analytical, associated with the initial sample collection, handling, and storage; and “post-post-” analytical, associated with the final interpretation to the end-users such as patients, readers of the publications, or the whole public, which have been proven to be more fallible than pre- and post-analytical actions [67]. In metabolomics, QA/QC can cover all activities from pre-pre-, pre-analytical, analytical to post- and post-post-analytical steps to ensure the acquisition of high quality and analytically reproducible data in any high-throughput study center. At clinical scale, metabolomics deals with thousands of samples, which urgently requires an appropriate experimental design and a universal QA/QC protocol. Concerning major factors that have been suggested, we will abridge current challenges based on the suggestion of Plebani et al. as well as provide firm recommendations in Figure 2 [58,67]. However, more efforts should be made to improve the current practice of metabolic phenotyping research. For instance, it is of importance to establish a platform to enhance the interpretation and comprehension of users (post-post phase) such as clinical doctors and public health specialists. Not limited to metabolomics, conventional protocols and current practice are being diversified among lipidomics laboratories, according to a recent comprehensive survey [68]. While diversity is a driving force for development, standardization and harmonization are extremely important to provide a robust outcome. The impact of this heterogeneity on metabolite and lipid quantitative measurement is currently unclear and needs to be explored further. Subsequently, all laboratories, regardless of experience, should collaborate to define community-level best practice guidelines for executing metabolomics and lipidomics experiments. This can be achieved by conducting interlaboratory studies, organizing symposiums, workshops, training sessions, or creating laboratory networks/focus groups aimed at resolving methodological issues. Another concern is the pressing need to extend the depth of metabolome coverage to prevent missing data and loss of valuable metabolites, and save cost [69]. Hence, the reference range for various kinds of biofluids and samples should be established. Once the reference interval of commonly detected metabolites is established and validated at an interlaboratory scale, it will benefit the QA/QC process and help orientate and validate other methods in the future. Additional recommendations that may be beneficial for metabolomics study are given in Table 2.

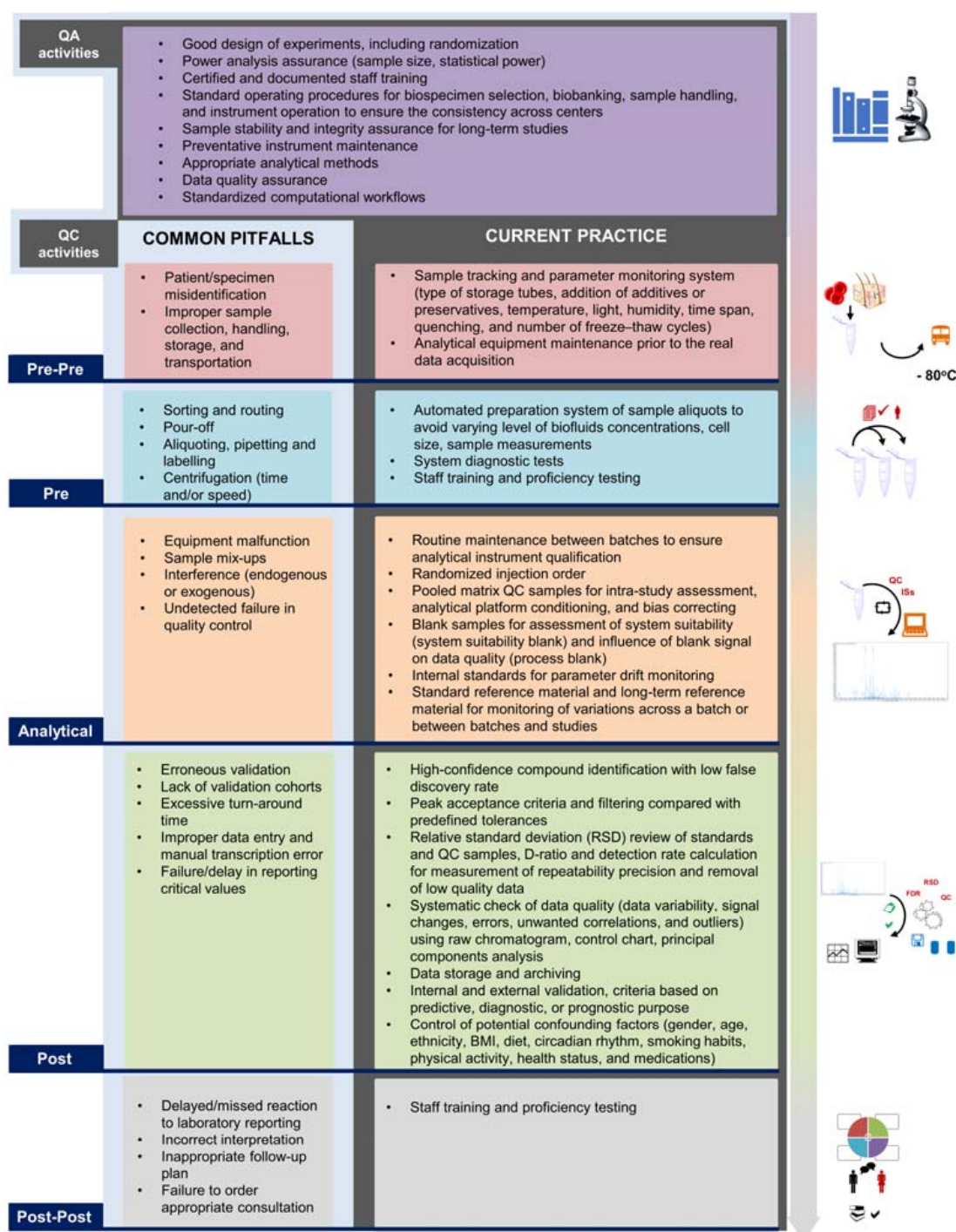


Figure 2. Quality assurance (QA)/quality control (QC) procedures following the novel five-step classification: pre-pre-analytical, pre-analytical, analytical, post-analytical, and post-post-analytical phases. QA activities are considered before while QC activities are undertaken during and after sample collection. The first column of QC is based on the content of five-step laboratory errors suggested by Plebani et al. [67]. The second column displays current QC techniques and activities recommended to be carried out in clinical metabolomics.

Table 2. Additional recommendations to ensure QA/QC procedures in metabolomics and lipidomics study at clinical and epidemiological scale.

| Phases ¹ | Recommendations ² |
|-----------------------------|---|
| Pre-pre-analytical | <ul style="list-style-type: none"> - Need of universal and reproducible study protocols for different tissue and biofluid types to reduce the effects of extrinsic exposures (e.g., enzymatic activities, oxygen, UV light, or temperature) and intrinsic factors (e.g., age, gender, health status, body mass index, circadian rhythm), or to standardize exposure of samples to these effects when inevitable. - Need for well-managed metabolomics biobanks with the proper collection, processing, storage, and tracking processes. - Need for an automated system for sample handling and preparation. - Consultation with an experienced statistician in the omics field. |
| Pre-analytical | <ul style="list-style-type: none"> - Application of a single validated standard operating procedure (SOP) for the pre-analytical phase across all sampling sites. - Need of a standardized protocol to preserve samples, avoid prolonged storage at room temperature and multiple numbers of freeze and thaw cycles. - Assessment of the suitability of the analytical platform using system suitability samples and blank samples. - Personalized training and education, different for novice and experienced researchers. |
| Analytical | <ul style="list-style-type: none"> - Adoption and calibration of every apparatus used under a quality assurance system, verification for unexpected variations, and routine maintenance. - Need for a pooled QC sample or when impossible, an alternative QC sample from the first batch of samples randomly collected. - Need for reporting standards/SOPs (at every step of the analytical process). - Establishment of minimum acceptance criteria for analyzed samples. |
| Post-analytical | <ul style="list-style-type: none"> - Establishment of a database of authentic standards using the same analytical condition to prevent misidentification (e.g., Fiehn's library that has more than 1000 authentic standards). - Use of data visualization tools to evaluate the analytical run quality and check for systematic and random errors. - Establishment of a strategy for statistical modeling. - Consideration of blindness if needed. |
| Post-post-analytical | <ul style="list-style-type: none"> - Report of QC metadata (e.g., sample order, QC sample, reference materials used). - Education of the community about QC procedures. - Standardized data sharing on a public repository. - Personalized training and education. |

Note: ¹ Based on the content of five-step laboratory errors suggested by Plebani et al. [67]. ² In reference to [49,58,62,70–73].

3.3. The Role of Establishing an Automated System for Sample Handling, Data Acquisition, and Automated Modeling

The shortage of standardized methods for sample preparation prior to analysis has been observed in metabolomics. Sample collection and processing is a major logistical problem and the largest source of errors in laboratory medicine, affecting downstream analyses [74]. Evidence-based practices to minimize pre-analytical mistakes are urgently needed and have been developed based on conventional biochemical assessments, but more research is required to determine the best method for omics techniques, including metabolomics [75]. Moreover, some metabolites are stable, whereas others are very susceptible to perturbations and improper sample handling [56]. In the clinical and epidemiological research settings, the establishment of an automated or semi-automated system for sample handling, acquisition, and automated modeling is needed to ensure robust assay performance regardless of

the experience of the operator. To overcome previous methods that still necessitate some manual intervention, a fully robotic sample preparation program for LC-MS bioanalysis has recently been proven to have significantly higher efficiency, with more sample throughput, superior data quality, and promising bioanalytical time- and cost-savings [71]. Laboratory automation, by lessening hands-on activities, also minimizes possible human errors [57]. In turn, automated sample preparation still poses a considerable challenge through the additional cost and time spent on instrument calibration and software validation for robotic liquid handlers before implementation. It will get harder to figure out errors or recognize them right after they occur when the system becomes more automated [57]. Moreover, the reliability and stability of metabolites in human biofluids against all the various factors of sampling, storage, and use should be measured and known before any further applications.

3.4. The Assessment of Systematic Errors in Clinical Metabolomics Studies

Systematic errors of data acquisition are current challenges in MS-based metabolomics; especially, these are more problematic in the big data era because secondary data analysts may not have access to knowledge about potential data artifacts. For example, metabolomics datasets are susceptible to measurement errors caused by the timing of sample collection and decrease in instrument sensitivity [76]. However, if meta-data are not included in the dataset, the analyst cannot accurately assess the impact of these errors. Possible causes of batch effects, another type of systematic error, involve differences in instrument performance, sample handling, differences in the preparation of batches, and other unwanted environmental and technical errors [77]. Unadjusted batch variation, even subtle, may predispose to the unexpected correlation of significant features with only the batches and tend to positively bias cross-validation results [78,79]. Correction of batch variation becomes particularly essential when analyzing or meta-analyzing datasets to both increase robustness and improve reproducibility across different metabolomics studies. Data normalization for within-batch and properly between batches can be divided into three categories: data-driven normalization, internal standards-based normalization, and QC-based normalization, but standard protocols are still lacking. Several different algorithms are available to perform the batch correction, among them the common principal component analysis (CPCA) combining with median fold change was reported to perform better than solely CPCA, component correction, median fold change, or ComBat [80]. However, how to normalize data from various sources for integration purposes is still challenging, requiring additional work on data processing. Scientists need to regularly test the instrument performance and employ QC samples to adjust systematic errors and estimate variations between analytical methods [58]. QC-based normalization is becoming more popular since it regresses unwanted variation, but retains essential biological variation of interest [77]. There are multiple types of QC samples corresponding to numerous purposes, as shown in Figure 2 (the analytical step). The use of QC samples for batch correction can lead to good results when enough quality controls, e.g., one QC sample for every five biological samples, are available [81]. However, a small but quite common practice is that only one or two pooled QC samples are prepared in each batch of analysis, which may introduce another source of error since the evaporation of the employed organic solvents over time may be significant in the analysis involving many samples. Regarding computational solutions, the new version of the *statTarget* package with a user-friendly graphical interface integrating a QC-based random forests signal correction algorithm has been released to remove inter- and intra-batch variations in omics study [82]. *NormalizerDE*, an R package, provides 12 different normalization methods for omics data to enable researchers the correct technical variation, and decide the most suitable strategy for particular datasets [83]. *NOREVA* is another powerful online tool that implements various normalization methods of which five well-established criteria were supported for the comparison of different normalization approaches [84]. Future research should focus on whether these adjustment methods introduce unwanted effects, such as the difference between the feature matrix and identified matrix, variations in the protocol used for QC samples analysis and the number of QC samples. Of note, batch-effect removal seems not beneficial when the batch variation is incompletely or improperly corrected, or batch-group design is unbalanced [85].

3.5. The Advancement and Limitations of Metabolite Annotation/Identification Standards

Protein misidentification is still a problem affecting reproducibility in proteomics, a field that is known to be more mature than metabolomics; then, it goes without question that misidentification will continue to be a bottle-neck in the more problematic area such as metabolomics [86]. Misidentification can induce false positive and/or false negative results during translation. Generally, the chemical analysis working group defined four different levels of confidence in metabolite identification observed in published metabolomics literature. Level 1 (identified compounds) requires comparison with more than one orthogonal property of a chemical reference standard acquired within the same laboratory and same technical setting. Level 2 (putatively annotated compounds) relies on the physicochemical property or spectral similarity to a known metabolite while level 3 (putatively annotated compound classes) relies on the physicochemical property or spectral similarity to those of a known chemical class of metabolites. Level 4 is about unknown compounds, but can still be extracted and quantified based on spectral data [87,88]. Lipidomics encounters a similar problem yet has distinct characteristics as discussed later in this text. In case known and unknown identification features are used together for the statistical analysis, proper information on these features, e.g., exact mass, MS/MS spectrum, retention time associated with the mobile phase, and its correlation with other elements, should be given to facilitate the later identification process. It is recommended that all authors report the level of confidence, common name, and structural code in their publication and submission to repositories. However, the employment of these standards still has not received adequate attention in peer-reviewed publications [89].

Matching the identified feature with an authentic chemical standard run on the same equipment and analytical method as the experimental samples or with MS/MS spectral databases is currently considered as the gold standard, which is impractical to interpret manually in case of enormous datasets. The major hindrance in metabolomics has now switched from identifying known MS/MS spectra molecules to the unknowns that cannot be found in the databases or that are available but do not contain experimental MS/MS data [90]. In-house spectral databases storing reference spectra have been intensively reviewed, such as The Human Metabolome Database (HMDB) and MassBank [91]. Another concern is that none of the existing metabolomics analysis software enables semi-automated search against public data in metabolomics repositories. Imitating the model of WebBLAST server for public sequencing data, a web-based MS/MS spectra data search engine called MASST has been introduced. MASST will allow users to search against all shared data on three central repositories (Metabolomics Workbench, MetaboLights, and GNPS) by converting data to a uniform open format and providing all library hits including the origin of the matched MS/MS spectral data, all dataset matches, and sample information or other metadata associated with that dataset [92]. This unique approach, called meta-annotation, will significantly improve the precision of metabolite annotation. Of note, a recently developed tool Peak Annotation and Verification Engine (PAVE) based on isotopic labeling and computational analysis indicated that only around 4% of all LC/MS peaks are annotated as putative metabolites [93].

3.6. Ongoing Efforts for a Universal Workflow Covering all Computational Steps of Metabolomics Research

The metabolomics community is initiating a great effort to arrive at a universal metabolomics tool that can simultaneously handle data preprocessing, processing, and visualization, as well as classification, network enrichment, and integrative analysis. Having such a platform would improve the depth of upcoming studies effortlessly and improve the reproducibility of the results. MZmine 2, MS-DIAL, and the “all-in-one” cloud-based platform of XCMS/METLIN and its variants including XCMS-MRM are among the most typical examples of the openness and complement of computational pre-processing procedures in metabolomics [90,94–102]. Notwithstanding, despite being the most popular statistical analysis tool used by the omics community, it still lacks functionality and is not particularly suitable for large-scale clinical metabolomics studies. Recently, developers released MetaboAnalystR 2.0, which additionally implemented raw spectral processing and functional interpretation in the endeavor to

handle large datasets, support end-to-end, and to report reproducible analysis of the data [103]. Among these integration efforts, potential universal solutions have evolved but are currently under development, such as Workflow4Metabolomics, PhenoMeNal, and MetaboFlow [104,105]. Workflow4Metabolomics is based on the Galaxy environment and provides multiple metabolomics processing workflows from raw data to metabolite annotation while PhenoMeNal and MetaboFlow offer an extensive suite of open-source, standardized, interoperable, and sharable metabolomics data pipelines as a complete data analysis solution. Galaxy is an endeavor toward harmonizing and standardizing metabolomics workflows, and is much more familiar with scientists than any other platform [106]. However, these initial achievements have not reached expectations, and their influence on the metabolomics community has been unimpressive.

For further data processing and model development, missing data may introduce unintentional bias. Imputation approaches have been reported in metabolomics studies, yet many of them borrow ideas from the field of microarray gene expression [107–109]. Dedicated to metabolomics, an integrative model employing two unique characteristics of metabolomics data, the known metabolic network information and the relationship between features within a metabolite such as mass-to-charge ratio (m/z) and retention time, was generated for missing data imputation [110]. A novel multiple factor analysis framework has been developed for multi-omics studies, which estimates and imputes plausible values to the missing data by integrating information from multiple layers [111].

4. Specific Notes on Clinical Lipidomics Studies

In lipidomics, there have been publications, from an epidemiological perspective, investigating lipid panels of human diseases. As in other omics fields, machine learning has also gradually proved its robust role in lipidomics studies, illustrated by its application in identifying the plasma lipidome for obesity prediction in a large population-based cohort [112]. However, these current measures do not depict the entire spectrum of the human plasma lipidome [10]. Despite remarkable progress in mass spectrometric technologies, there is still a lack of consensus in lipidomics measurements between different laboratories, and the quality of lipidomics output has not satisfied the criteria of common laboratory diagnostics [113]. In this context, an initial workflow with agreed recommendations has been recently suggested to facilitate the translation of quantitative MS-based lipidomics data into clinical settings [114]. Furthermore, 31 independent laboratories jointly participated in an interlaboratory study with the purpose to discover current gaps and harmonize lipid measurements across the lipidomics community [115]. Following this interlaboratory test, the consensus location estimates for several lipid categories were robustly established, which can serve as references for quality control and quality assessment. The Lipidomics Standards Initiative Consortium has also fostered lipidomics researchers to work closely to develop common minimum acceptable guidelines and standards toward harmonized lipidomics data [116]. Lipidomics is facing the same issues as metabolomics, though some of these aspects are technically and computationally different. Excellent discussions on the technical aspects of lipidomics have occurred [113,117,118]. We hereby discuss two significant lipidomics challenges that appear to be technically divergent from typical metabolomics.

4.1. QA/QC and Batch Effect Removal for Large-Scale Lipidomics Studies

As mentioned before, batch effects can alter the true biological signal. Several pioneering works have attempted to capture the unique characteristics of batch effects adjustment in lipidomics research. The population-based study within the Alzheimer Disease Neuroimaging Initiative cohort is a typical example of the role of QA/QC and technical variance removal in large-scale lipidomics studies. The authors employed QC samples to verify the high reproducibility of lipidomics data and the relative standard deviation (RSD) in QC samples as an inclusion criterion to establish the final lipid dataset for further statistical analysis [119]. In the subsequent data processing, they used the QC-based LOESS normalization method for batch effects adjustment to minimize the analytical variations [119]. In addition, a novel QC-based correction method using random forests (called SERRF) has been

developed. This method outperformed other commonly used normalization methods and was reported to have lowered unwanted systematic errors to 5% RSD in a large-scale lipidomics study [120].

4.2. The Need for High-Confidence Lipid Identification and Quantification

The top five challenges raised by the lipidomics community were (1) the shortage of standardization of methods and protocols, (2) the lack of reference standards, (3) the lack of data handling and quantification software, (4) over-reporting, and (5) the issue of false positives [68].

In lipid identification, level 1 of confidence is mostly not possible for a large number of lipid species. Fortunately, similar to peptides, the MS/MS patterns of significant lipid classes are predictable, which makes the *in silico* annotation and prediction of major lipids possible when consulting the MS/MS with biologically plausible fatty acid chains. Noticeably, the identification of lipid subclass compositions such as phosphatidylcholines (PCs), lysoPCs, and sphingomyelins at the fatty acid chain resolution using the AbsoluteIDQ™ p150 kit and Lipidizer™ platform provides evidence of the odd-chain fatty acids in PCs, which is quite essential and acknowledged. This method or other similar ones may be used in large-scale lipidomics research [121]. In-source fragmentation generates precursor ions whose masses are similar to those of endogenous lipids, which may induce misannotation. Therefore, a test for in-source fragments is recommended as an essential step in lipidomics method development [122]. Moreover, the advancement in separation techniques (e.g., ion mobility) and resolution of the mass spectrometer for lipid identification is needed to differentiate between isobaric species belonging to different lipid categories, resolve isotope effect, as well as allow researchers to directly quantify them in total lipid extracts [123].

Under the Lipidomics Standard Initiative, good practices and recommendations toward the standardization of crucial steps in lipid species identification have been discussed [124–126]. Concerning *in silico* lipidomics libraries, the major ones include LIPID MAPS and LipidBlast, followed by prospective executable packages such as MS-DIAL, LipidMatch, and LipiDex. *In silico* libraries allow users to identify lipids through a broad set of high-resolution tandem mass spectrometry experiments, hence improving biological interpretation [127–129]. Following the fast progress in lipidomics over the last two decades, new tools are needed to get more insights into this field. An R-based, open-source tool called Lipid Mining and Ontology (Lipid Mini-On) was later developed, which can split individual lipids into multiple ontology classes with Lipid Ontology terms (similar to Gene Ontology) based on LIPID MAPS classification as well as other molecular characteristics for the enrichment analysis of lipids currently absent from databases [130]. LION, a new and freely accessible lipid ontology, is able to assign more than 50,000 lipid species to various biophysical, chemical, and cell biological processes [131]. This web-based resource with a complementary enrichment analysis tool empowers researchers to investigate the complex lipidomes in biological systems.

5. Toward Reproducible Data Analysis and Interpretation in Clinical Metabolomics and Lipidomics

One of the possible future directions of computational metabolomics platforms is to diversify and standardize analytical processes for a variety of study designs, including cross-sectional and longitudinal studies, and clinical trials. In that context, the role of machine learning cannot be ignored. Machine learning has the power to process a large amount of data that exceeds human capacity, gather lessons and experiences from nearly all clinicians, and guide patients instantly [132]. A recent study compared the performance for phenotype differentiation in untargeted metabolomics among different machine learning techniques. Concerning all aspects including skewed error distributions, unbalanced phenotype allocation, biological and technical outliers, missing data, and dimensionality reduction, the two non-linear methods, support vector machine and random forests, act as the leading algorithms over naïve Bayes, neural networks, sPLS-DA, PLS-DA, and k-nearest neighbors classifiers [133]. Additionally, the opaque methods usually yield accurate predictions but are more complicated to interpret than the transparent ones. As a solution, algorithms such as LIME and iBreakDown show promise for

explaining the predictions of any complex, black-box model [134–137]. It is worth mentioning that the so-called explainable prediction models are most beneficial from the datasets that have a small number of features from input data. Feature selection/engineering in machine learning is an ongoing research topic. In that respect, automated and flexible feature engineering algorithms may be essential to increase the throughput, reproducibility, and transferability. Although several machine learning methods have been experimented, the interpretation of several complicated cases is still unfeasible and remains an open question [138]. Interestingly, an expert knowledge-collecting tool called tinderesting allows experts to rate the features, which will be subsequently used to train a machine learning model and may improve the overall performance of the prediction model in dynamic metabolomics studies [139].

Metabolomics research often recruits a small sample size, usually less than 300 per group for classification purposes, which may not ensure the study's reliability [140]. Hence, meta-analysis has evolved as a method of choice to enhance the study's robustness and minimize the risk of false discoveries [141]. Finally, the value of meta-analysis in metabolomics has been demonstrated to increase the probability of determining metabolites associated with the phenotype of interest [142,143]. Robust meta-analysis and network-based interaction visualization frameworks should be investigated for metabolomics datasets to bring the interpretation of metabolomics data to the next level. Of note, another method to increase the study robustness given small sample size is the application of appropriate study design such as prospective longitudinal study.

6. Fundamental Requirements on Standards for Reporting Metabolomics-Based Biomarker Studies at Clinical and Epidemiological Scale

The standard for reporting metabolomics-based biomarker discovery studies is still ambiguous and incomplete [144]. A thorough understanding of principles, adequate guidelines, and practical tools has minimized the pitfalls and fortified critical appraisal, validation, and replication of studies. The main characteristics of reporting principles are matching, exhaustive, skeptical, second-order, transparent, and reproducible [145]. In clinical metabolomics studies, authors are urged to report whether data appear reasonably matched to the hypothesis (data matching), whether there are alternative approaches to explore data (exhaustiveness), whether the analysis addresses any other premises or questions (skepticism), whether data pertain directly to the investigation or just contribute to relevant background (second-order), whether there are means such as data visualization or data summary included in the analysis that highlights the connection between data and results (transparency), and whether the code and data are made available in open sources (reproducibility) [145]. There are several reporting guidelines for clinical studies and multivariable prediction models, especially for diagnostic and prognostic purposes, such as the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statements, the Standards for the Reporting of Diagnostic Accuracy Studies (STARD), and the Biospecimen Reporting for Improved Study Quality (BRISQ) [146–148]. However, these guidelines are poorly incorporated in the omics field in general and in metabolomics in particular. Specifically, the assessment of metabolite and lipid biomarkers currently makes up a considerable portion of metabolomics research applications, whereas epidemiological studies have their unique features at the population level. Collectively, a reporting guideline combining biomarker diagnostic/prognostic characteristics with distinct epidemiological aspects in terms of study design, sample size, analytical approach, confounding, and bias should be developed to fully satisfy metabolomics study at clinical and epidemiological scales. The QUADOMICs tool, an adaptation for omics study of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS), assesses the methodological quality of omics-based diagnostic studies, and could be considered as a prototype to help develop an appropriate guideline for metabolomics [149]. Recently, a novel reporting guideline based on R Markdown template for the data analysis step of biomarker discovery metabolomics studies has been issued [150]. The new guideline, along with our suggested one in reference to STROBE, TRIPOD, and QUADOMICs, if applied, will improve the completeness, transparency, and reproducibility of clinical lipidomics and metabolomics study (Table 3). Metabolomics journals should obligate authors to share their code as a requirement for publication rather than recommend it.

Table 3. Proposed checklist for clinical metabolomics-based biomarker discovery and validation in reference to STROBE, TRIPOD, and QUADOMICs.

| Section/Topic | Item | Essential Topic | STROBE [151] | TRIPOD [146] | QUADOMICs [149] | Our In-House Assessment [143] | Checklist Item ¹ |
|----------------------------------|------|-----------------|--------------|--------------|-----------------|-------------------------------|---|
| 1. Title and abstract | | | | | | | |
| Title | 1a | Yes | 1 | 1 | | | Specify the study design with developing and/or validating purpose, the target population, and the outcome with simple and straightforward terms in the title. |
| Abstract | 1b | Yes | 1 | 2 | | | Depending on the target journal, provide in the abstract a precise and structured summary of objectives, study design, study participants, sample size, type of samples (e.g., plasma, serum, or urine), analytical platform, predictor variables, outcome, statistical method, results, and conclusions. |
| 2. Introduction | | | | | | | |
| Background | 2a | Yes | 2 | 3a | | | Provide the scientific and clinical background (including the diagnostic or prognostic purpose) and explain the rationale for developing and/or validating the multivariable prediction model, in regard to existing models. |
| Objectives | 2b | Yes | 3 | 3b | | | Determine the objectives and hypotheses, emphasizing if it is a development and/or validation research of the model. |
| 3. Methods | | | | | | | |
| Ethical approval | 3a | Yes | | | | | Clearly report ethics committee approval and participant consent. |
| Study design | 3b | Yes | 4 | 4a | | Item 1 | State the study design (e.g., case-control, cohort, randomized trial, or registry data) or source of data (e.g., biobank, public database). |
| Setting | 3c | Yes | 5 | 4b | | | Describe the settings, locations, and relevant dates where the data were collected, including periods of follow-up, if applicable. |
| Participants | 3d | Yes | 6 | 5a–5c | Item 1 and 2 | Item 2–6 | (a) Cross-sectional study: Give the inclusion and exclusion criteria, and the sources and methods of recruitment of study participants. (b) Longitudinal study: Give the inclusion and exclusion criteria, and the sources and methods of recruitment of study participants. Describe follow-up methods. Give diagnostic criteria and prior treatment, if applicable. For developing purpose, the spectrum of patients should represent those who will receive the test in practice. |
| Sample size | 3e | Yes | 10 | 8 | | | Explain how the sample size was determined. |
| Sample collection | 3f | Yes | | | Item 3 and 4 | Item 8 | Describe the type of samples, the procedures and timing of biological sample collection with reference to clinical factors and the methods to control metabolome changes (e.g., arterial versus venous blood, circadian oscillations, pre- and post-prandial status, the time between sampling and storage). |
| Sample storage | 3g | Yes | | | Item 5 | Item 8 | Describe the methods to control chemical and enzymatic degradation and/or interconversion. |
| Sample preparation | 3h | Yes | | | Item 5 | Item 9 | Describe the methods to control analytical errors and between-batch variations. |
| Data acquisition | 3i | Yes | | | | Item 10–13 | Describe the experimental conditions, the analytical validation methods, and the number of batches of analysis. |
| Data preprocessing and treatment | 3j | Yes | | | | Item 15 | Report parameters for peak detection (i.e., algorithms and acceptance criteria for valid peaks), deconvolution, alignment, and correction. Describe the methods to filter data noise, impute missing values, and correct batch effects (e.g., data-driven, internal standards-based, or QC-based normalization). |

Table 3. Cont.

| Section/Topic | Item | Essential Topic | STROBE [151] | TRIPOD [146] | QUADOMICS [149] | Our In-House Assessment [143] | Checklist Item ¹ |
|--|------|-----------------|--------------|----------------|-----------------|-------------------------------|--|
| Predictors | 3k | Yes | 8 | 7a–7b | | | Define all variables and explain their measurements used in constructing the multivariable prediction model. |
| Bias | 3l | Yes | 9 | | | | Characterize any attempts to show and solve potential sources of bias. |
| Missing data | 3m | Yes | 12 | 9 | | | Describe how missing data (e.g., samples) were handled. |
| Metabolite identification ² | 3n | Yes | | | | Item 14 | Describe the methods for metabolite identification and the level of confidence of identified compounds. Report whether a match with authentic standards has been conducted. |
| Multi-omics data integration | 3o | Optional | | | | | Report the multi-omics data integration, if available. Description of integration strategies, such as post-analysis data integration or simultaneous integration from different omics data types, is recommended. |
| Statistical analysis and modeling | 3p | Yes | 12 | 10a–10e and 12 | Item 16 | Item 16 | Describe statistical methods, the methods to detect outliers, validation methods, and performance measures (e.g., AUC, accuracy, sensitivity, and specificity). Describe any updates of the model after the validation, if done. For the validation purpose, determine any inconsistencies from the development data. |
| Outcome | 3q | Yes | 7 | 6a–6b | | | Explain the assessment of outcome that is predicted by the multivariable prediction model. Report any efforts to the blind assessment of the predicted outcome. |
| 4. Results | | | | | | | |
| Participants | 4a | Yes | 13 | 13a–13c | | | Report numbers of study participants at each stage (e.g., numbers potentially eligible, screened for eligibility, validated eligible, included in the study, finishing follow-up, and analyzed) and the number of participants with missing information for predictors and outcome. Give reasons for non-participation or exclusion at each stage. |
| Descriptive data | 4b | Yes | 14 | 13a–13c | | | Designate characteristics of study participants (e.g., baseline demographic, clinical, and socioeconomic status) and information on potential confounding factors. |
| Data exploratory analysis | 4c | Yes | | | | | Report exploratory data analysis (e.g., using unsupervised learning approaches). |
| Model development | 4d | Yes | 15–16 | 14a–14b | | | Specify the number of participants and outcomes in each analysis. Report the unadjusted and adjusted potential confounders that may influence associations between each predictor and outcome. |
| Model interpretation | 4e | Optional | 15–16 | 15a–15b | Item 14 and 15 | | Present the full prediction model to enable reproducibility for individuals. Explain how to interpret the prediction model in a human-friendly approach (e.g., LIME, iBreakDown). In case uninterpretable, indeterminate data should be reported. |
| Model performance | 4f | Yes | 15–16 | 16 | | | Report measures of performance of the prediction model. Describe the effects of unbalancing (e.g., cases versus controls) to the performance of the model. |
| Reference standards comparison | 4g | Optional | | | Item 6–13 | Item 7 | Compare the constructed models with currently approved approaches (e.g., CA 19.9 for diagnosing pancreatic cancer). |
| Model updating | 4h | Optional | 17 | 17 | | | Report the results of any updates of the model. An updated model using quantitative information of the tentative biomarkers are strongly recommended. |

Table 3. Cont.

| Section/Topic | Item | Essential Topic | STROBE [151] | TRIPOD [146] | QUADOMICs [149] | Our In-House Assessment [143] | Checklist Item ¹ |
|---------------------------------|------|-----------------|--------------|--------------|-----------------|-------------------------------|--|
| 5. Discussion | | | | | | | |
| Key results | 5a | Yes | 18 | | | | Summarize critical results in harmony with study objectives. |
| Interpretation | 5b | Yes | 20 | 19a–19b | | | Give an overall evaluation of results based on study objectives, results from previous similar studies, and other related evidence. For validation purpose, discuss the results in regard to the performance of development data, and any other validation data from the public databases. |
| Limitations | 5c | Yes | 19 | 18 | | | Discuss the limitations of the study, considering sources of potential confounders, biases, and statistical uncertainty. |
| Implications | 5d | Yes | 21 | 20 | | | Discuss the potential application of the model into clinical settings and suggestions for future research and practice. |
| 6. Other information | | | | | | | |
| Supplementary materials | 6a | Yes | | 21 | | | Provide available supplementary files, such as the full study protocol and generated datasets. |
| Funding | 6b | Yes | 22 | 22 | | | Give the sources of funding and the influence of each funder on the outcome and, if applicable, for the original study from which the current paper arises. |
| Conflicts of interest | 6c | Yes | | | | | Clearly declare potential conflicts of interest. |
| Repositories for generated data | 6d | Yes | | | | | Report the public repository for generated data concerning FAIR (Findability, Accessibility, Interoperability, and Reusability) principles ³ . |
| Executive commands | 6e | Optional | | | | | Report any programming code used (e.g., R code). |

Note: ¹ The description was adopted with modifications from STROBE, TRIPOD, and QUADOMICs guidelines along with our in-house assessment. ² Metabolite identification step can be performed prior to or after the Statistical analysis and modeling step. ³ In reference to [152].

7. Multi-Omics Integration for Precision Medicine

The exponential increase in amount of data in association with the gradual rise in sample numbers is the de facto characteristic of omics sciences. Single omics analysis usually reveals correlations of reactive processes rather than causations [153]. Years of constant endeavors have proved that focusing on technological “silos” is incapable of finding solutions to complex molecular pathophysiological issues [154]. It is undeniable that metabolomics itself brings substantial advantages to the multi-omics system, particularly in the diseases that have dysregulated metabolism, such as diabetes and cancer. For instance, in the predictive model of stress events in prediabetes patients, the classification performance of metabolomics was only overshadowed by that of multi-omics with the area under the receiver operating characteristic curve, reaching 80.1% [155]. The integration of multi-omics data, however, has the power to explain potential causative changes, strengthen the contribution of omics science to our understanding of biomedicine, and provide a complete picture of diseases [156]. In the context of data explosion at every omics level and accumulated knowledge of complex interactions between those molecular layers, a better paradigm integrating multi-omics is a potential solution to comprehend the underlying biological processes as well as to give better predictions on the outcome of interest (e.g., diagnosis, prognosis) [157–159]. For this exact reason, the idea of a single platform for a unifying proteomics-metabolomics analysis has been proposed [160]. We hereby provide some excellent examples to demonstrate the potential of multi-omics approaches. The combination of multi-omics technologies is undeniable in investigating cerebrospinal fluid (CSF) and expanding the coverage of CSF metabolites since biomolecules are interactive to exert their functions at multiple levels in a biological specimen [161]. The combination of multi-omics with the prospective longitudinal characteristic helps reinforce the causality. An integrated network analysis of genomics, metabolomics, and Alzheimer’s disease (AD) risk factors in 1111 subjects of Wisconsin Registry for Alzheimer’s Prevention study revealed that several genes might be indirectly associated with AD risk through particular metabolites [162]. In type 2 diabetes mellitus, a rich longitudinal investigation integrating transcriptomics, proteomics, metabolomics, and microbiomics sheds light on several molecular pathways, host-microbe interactions, and responses to infections that differed between glucose-dysregulated and healthy subjects, and highlighted the diverse intra- and inter-individual variabilities [155]. Another example, the transcriptomics-metabolomics integrated longitudinal research on molecular profiles of host-microbe interactions provided the most comprehensive summary, until now, of dysbiosis state in the gut microbiome in inflammatory bowel disease [163]. Fusing genome-wide association studies information with metabolomics data permits the simultaneous assessment of the genetic (or intrinsic) and environmental (or extrinsic) effects on disease phenotypes, which is better than using solely metabolomics [164]. Recently, an in-depth, prospective, longitudinal, phenotyping study gathering information on medical history, clinical and behavioral status, genome, transcriptome, proteome, immunome, metabolome, microbiome, exposome, and imaging from 109 subjects has been developed to comprehensively determine health status and establish risk profiles of metabolic, cardiovascular, fitness-frailty, cognitive, infectious, psychological, oncological, and inflammatory diseases [165].

The practicality of clinical metabolomics, as a core member of the omics-informed precision medicine system, is suggested in Figure 3. When all pitfalls are overcome, we may be able to establish a model from quantitative multi-omics data to determine the normal range of the population and is capable of detecting unspecific abnormal multi-omics profiles as seen in Figure 3A. It is essential to emphasize that the genetic background should be considered along with other socioeconomic factors, and the interactions among factors should be taken into account when constructing an omics-based model [166,167]. Combining with the advancement of omics-based disease-specific prediction, persons with a suspect medical condition will be determined for a proper intervention (Figure 3B). With the accumulation of these smart engines over time, together with benefits from data sharing of every aspect, we will be able to transform the medical system, of which omics profile is the major pillar (Figure 3C). The future of precision medicine is personomics, which is the combination of multi-omics, environmental factors, social interactions, and lifestyles (behaviors) based on unique personal references.

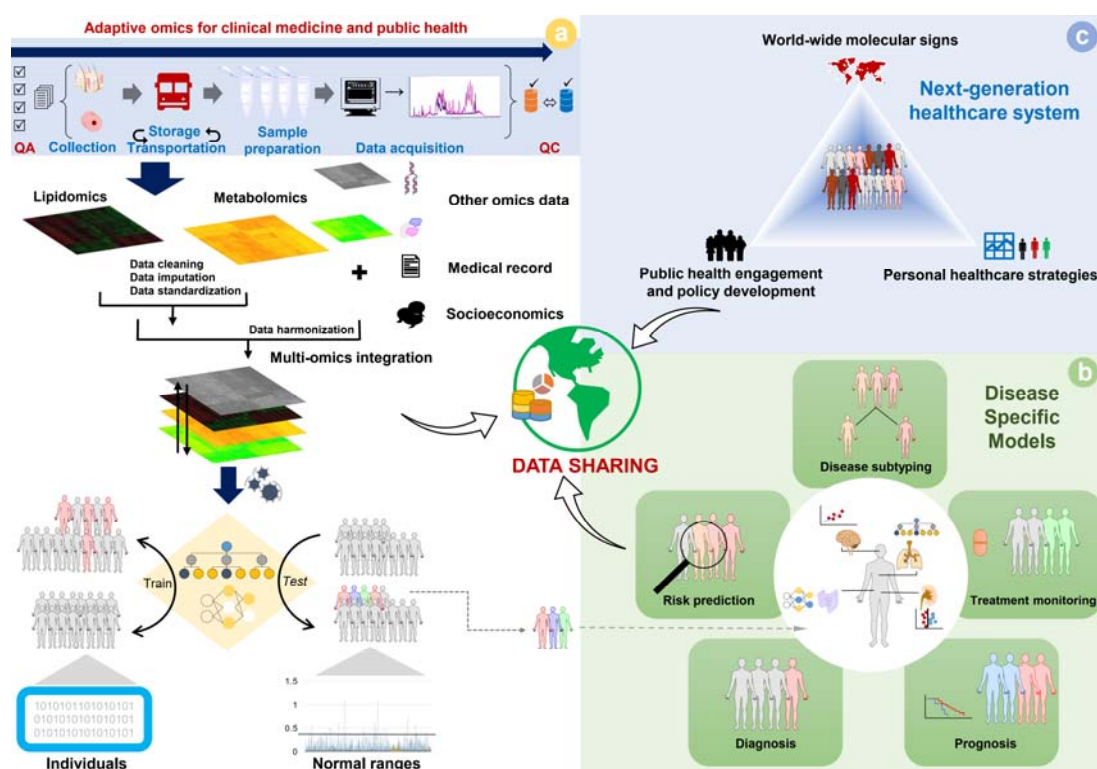


Figure 3. Metabolic phenotyping as a core member to establish multi-omics biosignatures for the development of the next-generation healthcare system: (a) adaptive workflow for a population-based study to determine the normal range of lipidome/metabolome, and detect nonspecifically abnormal conditions based on quantitative multi-omics profiles in reference to intrinsic, extrinsic, and socioeconomic factors; (b) corresponding models constructed for disease-specific purposes; (c) the next-generation healthcare system based on advances in smart engines with omics data as the primary pillar, and promoted by data sharing and machine learning.

Once again, it is important to emphasize the extraordinary role of machine learning in advancing multi-omics studies and reshaping our healthcare system. High quality multi-omics studies should formulate proper study hypotheses and select samples considering biological matrix characteristics. Regarding data processing and analysis, they are expected to handle and store samples avoiding adverse effects caused by intrinsic and extrinsic factors, carefully collect quantitative multi-omics data and associated meta-data, and employ better integration and interpretation software. For data replication, authors should adopt minimum standards for multi-omics methods and meta-data, and develop new tools for depositing intact multi-omics data rather than splitting it into multiple single datasets [168]. Regarding this matter, considerable efforts have been made toward an algorithm and comprehensive tools for analyzing omics data at the multi-omics scale to archive the most information and eliminate the most noise [169–171]. Herein, three major opinions have been raised with distinct properties: the genome first, the phenome first, and the environment first approaches [153]. Future tools and software should consider the interactions between different omics layers to extend the potential of this approach entirely.

8. Potential Bioethical Issues Associated with the Sharing of Clinical Metabolomics and Lipidomics Data

As in genomics, the most mature of the omics fields, there are ethical concerns that are worth paying attention to when we handle metabolomics data, including, but not limited to, privacy and identifiability, harmonization across sites regarding informed consent and institutional review board, incidental findings, data ownership, and funding priority [172]. Toward clinical application, the

completeness of personal information, including medical records, medication lists, and socioeconomic status, would be the major obstacle. Sharing data from research involving human samples should comply with the ethical, legal, and social implications, and patient privacy must always be protected as the first priority [173]. Nevertheless, unlike genomics, metabolomics data have a cross-sectional characteristic and are not unique to a particular individual; therefore, they do not contain factors listed in the “Safe Harbor” method for de-identification under the Health Insurance Portability and Accountability Act privacy rule. There are currently no known methods of recognizing a patient based on their metabolic panel, except potentially for rare diseases [173]. In this way, risks of identification can be avoided, which is the everlasting concern for genomics research participants who are afraid that their genomic information may be used for adversary attacks.

9. Perspectives and Conclusions

It is undebatable that the availability of research data and standardized report publications will substantially eradicate barriers against scientific progress. Throughout the paper, we have made considerable attempts to explore the remaining challenges, evaluate current opportunities, and suggest several future directions that should be considered by the metabolomics and lipidomics community to facilitate data acquisition, analysis, and sharing at the clinical and epidemiological scale. Physicians should be prepared to use “cognitive computers” applying artificial intelligence and machine learning, which in the near future, will be implemented in the hospital settings to assist in disease diagnosis and prediction of patient outcomes. In the big picture, metabolomics and lipidomics research has to be allied to a network of multi-layer omics combining with the power of machine learning, artificial intelligence, and standardized QA/QC protocol. We are also pioneers in the suggestion of a complete framework for quantitative metabolomics research strictly based on standardizations and reporting guidelines from study design to interpretation to the end-users. On the whole, the metabolomics community is encouraged to be involved in the joint multi-omics platforms, e.g., The Cancer Genome Atlas and International Cancer Genome Consortium, to facilitate the advancement of molecular research and precision medicine. International and multidisciplinary efforts are required to achieve this goal. Finally, although we did not intensively go through related bioethical issues in this review, they should always be taken into account to complement the next-generation healthcare system.

Author Contributions: S.W.K. and N.P.L. conceived the project. S.W.K. and S.K.P. supervised the work. N.P.L. and T.D.N. searched for references and prepared the original draft. N.P.L., T.D.N., Y.P.K., N.H.A., H.M.K., S.K.P., and S.W.K. reviewed critically and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2018R1A5A2024425) and the Bio-Synergy Research Project of the Ministry of Science, ICT and Future Planning through the National Research Foundation (NRF-2012M3A9C4048796).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Patti, G.J.; Yanes, O.; Siuzdak, G. Metabolomics: The apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263. [[CrossRef](#)] [[PubMed](#)]
2. Holmes, E.; Wilson, I.D.; Nicholson, J.K. Metabolic phenotyping in health and disease. *Cell* **2008**, *134*, 714–717. [[CrossRef](#)] [[PubMed](#)]
3. Bizzarri, M.; Brash, D.E.; Briscoe, J.; Grieneisen, V.A.; Stern, C.D.; Levin, M. A call for a better understanding of causation in cell biology. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 261–262. [[CrossRef](#)] [[PubMed](#)]
4. Chen, R.; Snyder, M. Promise of personalized omics to precision medicine. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2013**, *5*, 73–82. [[CrossRef](#)]
5. Grapov, D.; Fahrman, J.; Wanichthanarak, K.; Khoomrung, S. Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. *OMICS* **2018**, *22*, 630–636. [[CrossRef](#)]

6. Sampson, J.N.; Boca, S.M.; Shu, X.O.; Stolzenberg-Solomon, R.Z.; Matthews, C.E.; Hsing, A.W.; Tan, Y.T.; Ji, B.-T.; Chow, W.-H.; Cai, Q.; et al. Metabolomics in epidemiology: Sources of variability in metabolite measurements and implications. *Cancer Epidemiol. Biomark. Prev.* **2013**, *22*, 631. [[CrossRef](#)]
7. Ala-Korpela, M.; Davey Smith, G. Metabolic profiling—multitude of technologies with great research potential, but (when) will translation emerge? *Int. J. Epidemiol.* **2016**, *45*, 1311–1318. [[CrossRef](#)]
8. Valdes, A.; Tzoulaki, I.; Ioannidis, J.P.A.; Elliott, P.; Ebbels, T.M.D. Design and analysis of metabolomics studies in epidemiologic research: A primer on -omic technologies. *Am. J. Epidemiol.* **2014**, *180*, 129–139. [[CrossRef](#)]
9. Fearnley, L.G.; Inouye, M. Metabolomics in epidemiology: From metabolite concentrations to integrative reaction networks. *Int. J. Epidemiol.* **2016**, *45*, 1319–1328. [[CrossRef](#)]
10. Mundra, P.A.; Shaw, J.E.; Meikle, P.J. Lipidomic analyses in epidemiology. *Int. J. Epidemiol.* **2016**, *45*, 1329–1338. [[CrossRef](#)]
11. van Roekel, E.H.; Lofftfield, E.; Kelly, R.S.; Zeleznik, O.A.; Zanetti, K.A. Metabolomics in epidemiologic research: Challenges and opportunities for early-career epidemiologists. *Metabolomics* **2019**, *15*, 9. [[CrossRef](#)] [[PubMed](#)]
12. Papandreou, C.; Hernández-Alonso, P.; Bulló, M.; Ruiz-Canela, M.; Yu, E.; Guasch-Ferré, M.; Toledo, E.; Dennis, C.; Deik, A.; Clish, C.; et al. Plasma metabolites associated with coffee consumption: A metabolomic approach within the predimed study. *Nutrients* **2019**, *11*, 1032. [[CrossRef](#)] [[PubMed](#)]
13. Pesch, B.; Brüning, T.; Johnen, G.; Casjens, S.; Bonberg, N.; Taeger, D.; Müller, A.; Weber, D.G.; Behrens, T. Biomarker research with prospective study designs for the early detection of cancer. *Biochim. Biophys. Acta Proteins Proteom.* **2014**, *1844*, 874–883. [[CrossRef](#)] [[PubMed](#)]
14. Fest, J.; Vijfhuizen, L.S.; Goeman, J.J.; Veth, O.; Joensuu, A.; Perola, M.; Mannisto, S.; Ness-Jensen, E.; Hveem, K.; Haller, T.; et al. Search for early pancreatic cancer blood biomarkers in five European prospective population biobanks using metabolomics. *Endocrinology* **2019**, *160*, 1731–1742. [[CrossRef](#)] [[PubMed](#)]
15. Carter, R.A.; Pan, K.; Harville, E.W.; McRitchie, S.; Sumner, S. Metabolomics to reveal biomarkers and pathways of preterm birth: A systematic review and epidemiologic perspective. *Metabolomics* **2019**, *15*, 124. [[CrossRef](#)] [[PubMed](#)]
16. Playdon, M.C.; Joshi, A.D.; Tabung, F.K.; Cheng, S.; Henglin, M.; Kim, A.; Lin, T.; van Roekel, E.H.; Huang, J.; Krumsiek, J.; et al. Metabolomics analytics workflow for epidemiological research: Perspectives from the Consortium of Metabolomics Studies (COMETS). *Metabolites* **2019**, *9*, 145. [[CrossRef](#)]
17. Chu, S.H.; Huang, M.; Kelly, R.S.; Benedetti, E.; Siddiqui, J.K.; Zeleznik, O.A.; Pereira, A.; Herrington, D.; Wheelock, C.E.; Krumsiek, J.; et al. Integration of metabolomic and other omics data in population-based study designs: An epidemiological perspective. *Metabolites* **2019**, *9*, 117. [[CrossRef](#)]
18. Cuperlovic-Culf, M. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* **2018**, *8*, 4. [[CrossRef](#)]
19. Ciocan-Cartita, A.C.; Jurj, A.; Buse, M.; Gulei, D.; Braicu, C.; Raduly, L.; Cojocneanu, R.; Pruteanu, L.L.; Iuga, A.C.; Coza, O.; et al. The relevance of mass spectrometry analysis for personalized medicine through its successful application in cancer “omics”. *Int. J. Mol. Sci.* **2019**, *20*, 2576. [[CrossRef](#)]
20. Zhao, H.; Shen, J.; Moore, S.C.; Ye, Y.; Wu, X.; Esteva, F.J.; Tripathy, D.; Chow, W.H. Breast cancer risk in relation to plasma metabolites among Hispanic and African American women. *Breast Cancer Res. Treat.* **2019**, *176*, 687–696. [[CrossRef](#)]
21. Wang, R.; Zhao, H.; Zhang, X.; Zhao, X.; Song, Z.; Ouyang, J. Metabolic discrimination of breast cancer subtypes at the single-cell level by multiple microextraction coupled with mass spectrometry. *Anal. Chem.* **2019**, *91*, 3667–3674. [[CrossRef](#)] [[PubMed](#)]
22. Ismail, I.T.; Showalter, M.R.; Fiehn, O. Inborn errors of metabolism in the era of untargeted metabolomics and lipidomics. *Metabolites* **2019**, *9*, 242. [[CrossRef](#)] [[PubMed](#)]
23. Anh, N.H.; Long, N.P.; Kim, S.J.; Min, J.E.; Yoon, S.J.; Kim, H.M.; Yang, E.; Hwang, E.S.; Park, J.H.; Hong, S.S.; et al. Steroidomics for the prevention, assessment, and management of cancers: A systematic review and functional analysis. *Metabolites* **2019**, *9*, 199. [[CrossRef](#)] [[PubMed](#)]
24. Kim, D.J.; Cho, E.J.; Yu, K.S.; Jang, I.J.; Yoon, J.H.; Park, T.; Cho, J.Y. Comprehensive metabolomic search for biomarkers to differentiate early stage hepatocellular carcinoma from cirrhosis. *Cancers* **2019**, *11*, 1497. [[CrossRef](#)] [[PubMed](#)]

25. Guasch-Ferre, M.; Hruby, A.; Toledo, E.; Clish, C.B.; Martinez-Gonzalez, M.A.; Salas-Salvado, J.; Hu, F.B. Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis. *Diabetes Care* **2016**, *39*, 833–846. [[CrossRef](#)]
26. Arneth, B.; Arneth, R.; Shams, M. Metabolomics of type 1 and type 2 diabetes. *Int. J. Mol. Sci.* **2019**, *20*, 2467. [[CrossRef](#)]
27. Arlt, W.; Biehl, M.; Taylor, A.E.; Hahner, S.; Libe, R.; Hughes, B.A.; Schneider, P.; Smith, D.J.; Stiekema, H.; Krone, N.; et al. Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J. Clin. Endocrinol. Metab.* **2011**, *96*, 3775–3784. [[CrossRef](#)]
28. Wang, T.J.; Larson, M.G.; Vasan, R.S.; Cheng, S.; Rhee, E.P.; McCabe, E.; Lewis, G.D.; Fox, C.S.; Jacques, P.F.; Fernandez, C.; et al. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **2011**, *17*, 448. [[CrossRef](#)]
29. Kerkhofs, T.M.A.; Kerstens, M.N.; Kema, I.P.; Willems, T.P.; Haak, H.R. Diagnostic value of urinary steroid profiling in the evaluation of adrenal tumors. *Horm. Cancer* **2015**, *6*, 168–175. [[CrossRef](#)]
30. Dunn, W.B.; Lin, W.; Broadhurst, D.; Begley, P.; Brown, M.; Zelena, E.; Vaughan, A.A.; Halsall, A.; Harding, N.; Knowles, J.D.; et al. Molecular phenotyping of a UK population: Defining the human serum metabolome. *Metabolomics* **2015**, *11*, 9–26. [[CrossRef](#)]
31. Ziegler, R.G.; Fuhrman, B.J.; Moore, S.C.; Matthews, C.E. Epidemiologic studies of estrogen metabolism and breast cancer. *Steroids* **2015**, *99*, 67–75. [[CrossRef](#)] [[PubMed](#)]
32. Rewers, M.; Hyoty, H.; Lernmark, A.; Hagopian, W.; She, J.X.; Schatz, D.; Ziegler, A.G.; Toppari, J.; Akolkar, B.; Krischer, J. The Environmental Determinants of Diabetes in the Young (TEDDY) study: 2018 Update. *Curr. Diabetes Rep.* **2018**, *18*, 136. [[CrossRef](#)] [[PubMed](#)]
33. Hoyles, L.; Fernández-Real, J.-M.; Federici, M.; Serino, M.; Abbott, J.; Charpentier, J.; Heymes, C.; Luque, J.L.; Anthony, E.; Barton, R.H.; et al. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nat. Med.* **2018**, *24*, 1070–1080. [[CrossRef](#)] [[PubMed](#)]
34. Harada, S.; Hirayama, A.; Chan, Q.; Kurihara, A.; Fukai, K.; Iida, M.; Kato, S.; Sugiyama, D.; Kuwabara, K.; Takeuchi, A.; et al. Reliability of plasma polar metabolite concentrations in a large-scale cohort study using capillary electrophoresis-mass spectrometry. *PLoS ONE* **2018**, *13*, e0191230. [[CrossRef](#)] [[PubMed](#)]
35. Deelen, J.; Kettunen, J.; Fischer, K.; van der Spek, A.; Trompet, S.; Kastenmuller, G.; Boyd, A.; Zierer, J.; van den Akker, E.B.; Ala-Korpela, M.; et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat. Commun.* **2019**, *10*, 3346. [[CrossRef](#)]
36. Trivedi, D.K.; Sinclair, E.; Xu, Y.; Sarkar, D.; Walton-Doyle, C.; Liscio, C.; Banks, P.; Milne, J.; Silverdale, M.; Kunath, T.; et al. Discovery of volatile biomarkers of Parkinson’s disease from sebum. *ACS Cent. Sci.* **2019**, *5*, 599–606. [[CrossRef](#)]
37. Tzoulaki, I.; Castagné, R.; Boulangé, C.L.; Karaman, I.; Chekmeneva, E.; Evangelou, E.; Ebbels, T.M.D.; Kaluarachchi, M.R.; Chadeau-Hyam, M.; Mosen, D.; et al. Serum metabolic signatures of coronary and carotid atherosclerosis and subsequent cardiovascular disease. *Eur. Heart J.* **2019**, *40*, 2883–2896. [[CrossRef](#)]
38. Cirulli, E.T.; Guo, L.; Leon Swisher, C.; Shah, N.; Huang, L.; Napier, L.A.; Kirkness, E.F.; Spector, T.D.; Caskey, C.T.; Thorens, B.; et al. Profound perturbation of the metabolome in obesity is associated with health risk. *Cell Metab.* **2019**, *29*, 488–500. [[CrossRef](#)]
39. Chen, Z.Z.; Liu, J.; Morningstar, J.; Heckman-Stoddard, B.M.; Lee, C.G.; Dagogo-Jack, S.; Ferguson, J.F.; Hamman, R.F.; Knowler, W.C.; Mather, K.J.; et al. Metabolite Profiles of Incident Diabetes and Heterogeneity of Treatment Effect in the Diabetes Prevention Program. *Diabetes* **2019**, *68*, 2337–2349. [[CrossRef](#)]
40. Gomez, C.; Gonzalez-Riano, C.; Barbas, C.; Kolmert, J.; Hyung Ryu, M.; Carlsten, C.; Dahlen, S.E.; Wheelock, C.E. Quantitative metabolic profiling of urinary eicosanoids for clinical phenotyping. *J. Lipid Res.* **2019**, *60*, 1164–1173. [[CrossRef](#)]
41. Yu, Z.; Zhai, G.; Singmann, P.; He, Y.; Xu, T.; Prehn, C.; Romisch-Margl, W.; Lattka, E.; Gieger, C.; Soranzo, N.; et al. Human serum metabolic profiles are age dependent. *Aging Cell* **2012**, *11*, 960–967. [[CrossRef](#)]
42. Meikle, P.J.; Wong, G.; Barlow, C.K.; Weir, J.M.; Greeve, M.A.; MacIntosh, G.L.; Almasry, L.; Comuzzie, A.G.; Mahaney, M.C.; Kowalczyk, A.; et al. Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. *PLoS ONE* **2013**, *8*, e74341. [[CrossRef](#)] [[PubMed](#)]
43. Ganna, A.; Salihovic, S.; Sundstrom, J.; Broeckling, C.D.; Hedman, A.K.; Magnusson, P.K.; Pedersen, N.L.; Larsson, A.; Siegbahn, A.; Zilmer, M.; et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet.* **2014**, *10*, e1004801. [[CrossRef](#)] [[PubMed](#)]

44. Stegemann, C.; Pechlaner, R.; Willeit, P.; Langley, S.R.; Mangino, M.; Mayr, U.; Menni, C.; Moayyeri, A.; Santer, P.; Rungger, G.; et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation* **2014**, *129*, 1821–1831. [CrossRef] [PubMed]
45. Lofffield, E.; Rothwell, J.A.; Sinha, R.; Keski-Rahkonen, P.; Robinot, N.; Albanes, D.; Weinstein, S.J.; Derkach, A.; Sampson, J.; Scalbert, A.; et al. Prospective investigation of serum metabolites, coffee drinking, liver cancer incidence, and liver disease mortality. *J. Natl. Cancer Inst.* **2019**, *112*. [CrossRef]
46. Floegel, A.; Kuhn, T.; Sookthai, D.; Johnson, T.; Prehn, C.; Rolle-Kampczyk, U.; Otto, W.; Weikert, C.; Illig, T.; von Bergen, M.; et al. Serum metabolites and risk of myocardial infarction and ischemic stroke: A targeted metabolomic approach in two German prospective cohorts. *Eur. J. Epidemiol.* **2018**, *33*, 55–66. [CrossRef]
47. Derkach, A.; Sampson, J.; Joseph, J.; Playdon, M.C.; Stolzenberg-Solomon, R.Z. Effects of dietary sodium on metabolites: The Dietary Approaches to Stop Hypertension (DASH)-Sodium Feeding Study. *Am. J. Clin. Nutr.* **2017**, *106*, 1131–1141. [CrossRef]
48. Ahmed, W.M.; Brinkman, P.; Weda, H.; Knobel, H.H.; Xu, Y.; Nijssen, T.M.; Goodacre, R.; Rattray, N.; Vink, T.J.; Santonico, M.; et al. Methodological considerations for large-scale breath analysis studies: Lessons from the U-BIOPRED severe asthma project. *J. Breath Res.* **2018**, *13*, 016001. [CrossRef]
49. Kirwan, J.A.; Brennan, L.; Broadhurst, D.; Fiehn, O.; Cascante, M.; Dunn, W.B.; Schmidt, M.A.; Velagapudi, V. Preanalytical processing and biobanking procedures of biological samples for metabolomics research: A White paper, community perspective (for “Precision Medicine and Pharmacometabolomics Task Group”—the Metabolomics Society Initiative). *Clin. Chem.* **2018**, *64*, 1158–1182. [CrossRef]
50. Breier, M.; Wahl, S.; Prehn, C.; Fugmann, M.; Ferrari, U.; Weise, M.; Banning, F.; Seissler, J.; Grallert, H.; Adamski, J.; et al. Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. *PLoS ONE* **2014**, *9*, e89728. [CrossRef]
51. Vaught, J.; Bledsoe, M.; Watson, P. Biobanking on multiple continents: Will international coordination follow? *Biopreserv. Biobank.* **2014**, *12*, 1–2. [CrossRef] [PubMed]
52. International Agency for Research on Cancer. IARC Biobank. Available online: <http://ibb.iarc.fr/links/index.php> (accessed on 30 November 2019).
53. Norwegian Institute of Public Health. Promoting Harmonization of Epidemiological Biobanks in Europe. Available online: <https://www.fhi.no/en/projects/fp6-phoebe-promoting-harmonisat/> (accessed on 30 November 2019).
54. Consortium, BBMRI-ERIC. Biobanking and Biomolecular Resources Research Infrastructure. Available online: <http://www.bbMRI-eric.eu/services/standardisation/> (accessed on 30 November 2019).
55. La Frano, M.R.; Carmichael, S.L.; Ma, C.; Hardley, M.; Shen, T.; Wong, R.; Rosales, L.; Borkowski, K.; Pedersen, T.L.; Shaw, G.M.; et al. Impact of post-collection freezing delay on the reliability of serum metabolomics in samples reflecting the California mid-term pregnancy biobank. *Metabolomics* **2018**, *14*, 151. [CrossRef] [PubMed]
56. Haid, M.; Muschet, C.; Wahl, S.; Römisch-Margl, W.; Prehn, C.; Möller, G.; Adamski, J. Long-term stability of human plasma metabolites during storage at $-80\text{ }^{\circ}\text{C}$. *J. Proteome Res.* **2018**, *17*, 203–211. [CrossRef] [PubMed]
57. Lippi, G.; Betsou, F.; Cadamuro, J.; Cornes, M.; Fleischhacker, M.; Fruekilde, P.; Neumaier, M.; Nybo, M.; Padoan, A.; Plebani, M.; et al. Preanalytical challenges—Time for solutions. *Clin. Chem. Lab. Med.* **2019**, *57*, 974–981. [CrossRef] [PubMed]
58. 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]
59. Hyotylainen, T.; Ahonen, L.; Poho, P.; Oresic, M. Lipidomics in biomedical research—practical considerations. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2017**, *1862*, 800–803. [CrossRef]
60. 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]
61. Kamleh, M.A.; Ebbels, T.M.; Spagou, K.; Masson, P.; Want, E.J. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Anal. Chem.* **2012**, *84*, 2670–2677. [CrossRef]
62. 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]

63. Liu, X.; Hoene, M.; Wang, X.; Yin, P.; Haring, H.U.; Xu, G.; Lehmann, R. Serum or plasma, what is the difference? Investigations to facilitate the sample material selection decision making process for metabolomics studies and beyond. *Anal. Chim. Acta* **2018**, *1037*, 293–300. [[CrossRef](#)]
64. Kohler, I.; Hankemeier, T.; van der Graaf, P.H.; Knibbe, C.A.J.; van Hasselt, J.G.C. Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. *Eur. J. Pharm. Sci.* **2017**, *109*, S15–S21. [[CrossRef](#)] [[PubMed](#)]
65. Sieber-Ruckstuhl, N.S.; Burla, B.; Spoerel, S.; Schmid, F.; Venzin, C.; Cazenave-Gassiot, A.; Bendt, A.K.; Torta, F.; Wenk, M.R.; Boretti, F.S. Changes in the canine plasma lipidome after short- and long-term excess glucocorticoid exposure. *Sci. Rep.* **2019**, *9*, 6015. [[CrossRef](#)] [[PubMed](#)]
66. Beger, R.D. Interest is high in improving quality control for clinical metabolomics: Setting the path forward for community harmonization of quality control standards. *Metabolomics* **2018**, *15*, 1. [[CrossRef](#)] [[PubMed](#)]
67. Plebani, M. The detection and prevention of errors in laboratory medicine. *Ann. Clin. Biochem.* **2009**, *47*, 101–110. [[CrossRef](#)] [[PubMed](#)]
68. Bowden, J.A.; Ulmer, C.Z.; Jones, C.M.; Koelmel, J.P.; Yost, R.A. NIST lipidomics workflow questionnaire: An assessment of community-wide methodologies and perspectives. *Metabolomics* **2018**, *14*, 53. [[CrossRef](#)] [[PubMed](#)]
69. Lei, Z.; Huhman, D.V.; Sumner, L.W. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **2011**, *286*, 25435–25442. [[CrossRef](#)] [[PubMed](#)]
70. Khadka, M.; Todor, A.; Maner-Smith, K.M.; Colucci, J.K.; Tran, V.; Gaul, D.A.; Anderson, E.J.; Natrajan, M.S.; Roupheal, N.; Mulligan, M.J.; et al. The effect of anticoagulants, temperature, and time on the human plasma metabolome and lipidome from healthy donors as determined by liquid chromatography-mass spectrometry. *Biomolecules* **2019**, *9*, 200. [[CrossRef](#)]
71. Robin, T.; Barnes, A.; Dulaurent, S.; Loftus, N.; Baumgarten, S.; Moreau, S.; Marquet, P.; El Balkhi, S.; Saint-Marcoux, F. Fully automated sample preparation procedure to measure drugs of abuse in plasma by liquid chromatography tandem mass spectrometry. *Anal. Bioanal. Chem.* **2018**, *410*, 5071–5083. [[CrossRef](#)]
72. Dudzik, D.; Barbas-Bernardos, C.; García, A.; Barbas, C. Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. *J. Pharm. Biomed. Anal.* **2018**, *147*, 149–173. [[CrossRef](#)]
73. Blaženović, I.; Kind, T.; Sa, M.R.; Ji, J.; Vaniya, A.; Wanciewicz, B.; Roberts, B.S.; Torbašinović, H.; Lee, T.; Mehta, S.S.; et al. Structure annotation of all mass spectra in untargeted metabolomics. *Anal. Chem.* **2019**, *91*, 2155–2162. [[CrossRef](#)]
74. Anton, G.; Wilson, R.; Yu, Z.H.; Prehn, C.; Zukunft, S.; Adamski, J.; Heier, M.; Meisinger, C.; Romisch-Margl, W.; Wang-Sattler, R.; et al. Pre-analytical sample quality: Metabolite ratios as an intrinsic marker for prolonged room temperature exposure of serum samples. *PLoS ONE* **2015**, *10*, e0121495. [[CrossRef](#)] [[PubMed](#)]
75. Yin, P.; Peter, A.; Franken, H.; Zhao, X.; Neukamm, S.S.; Rosenbaum, L.; Lucio, M.; Zell, A.; Haring, H.U.; Xu, G.; et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin. Chem.* **2013**, *59*, 833–845. [[CrossRef](#)] [[PubMed](#)]
76. Mooney, S.J.; Pejaver, V. Big data in public health: Terminology, machine learning, and privacy. *Annu. Rev. Public Health* **2018**, *39*, 95–112. [[CrossRef](#)] [[PubMed](#)]
77. De Livera, A.M.; Sysi-Aho, M.; Jacob, L.; Gagnon-Bartsch, J.A.; Castillo, S.; Simpson, J.A.; Speed, T.P. Statistical methods for handling unwanted variation in metabolomics data. *Anal. Chem.* **2015**, *87*, 3606–3615. [[CrossRef](#)] [[PubMed](#)]
78. Sonesson, C.; Gerster, S.; Delorenzi, M. Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PLoS ONE* **2014**, *9*, e100335. [[CrossRef](#)] [[PubMed](#)]
79. Qin, L.X.; Huang, H.C.; Begg, C.B. Cautionary note on using cross-validation for molecular classification. *J. Clin. Oncol.* **2016**, *34*, 3931–3938. [[CrossRef](#)] [[PubMed](#)]
80. Fernandez-Albert, F.; Llorach, R.; Garcia-Aloy, M.; Ziyatdinov, A.; Andres-Lacueva, C.; Perera, A. Intensity drift removal in LC/MS metabolomics by common variance compensation. *Bioinformatics* **2014**, *30*, 2899–2905. [[CrossRef](#)] [[PubMed](#)]
81. Wehrens, R.; Hageman, J.A.; van Eeuwijk, F.; Kooke, R.; Flood, P.J.; Wijnker, E.; Keurentjes, J.J.; Lommen, A.; van Eekelen, H.D.; Hall, R.D.; et al. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* **2016**, *12*, 88. [[CrossRef](#)]
82. Luan, H.; Ji, F.; Chen, Y.; Cai, Z. statTarget: A streamlined tool for signal drift correction and interpretations of quantitative mass spectrometry-based omics data. *Anal. Chim. Acta* **2018**, *1036*, 66–72. [[CrossRef](#)]

83. Willforss, J.; Chawade, A.; Levander, F. NormalyzerDE: Online tool for improved normalization of omics expression data and high-sensitivity differential expression analysis. *J. Proteome Res.* **2019**, *18*, 732–740. [[CrossRef](#)]
84. Li, B.; Tang, J.; Yang, Q.; Li, S.; Cui, X.; Li, Y.; Chen, Y.; Xue, W.; Li, X.; Zhu, F. NOREVA: Normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* **2017**, *45*, W162–W170. [[CrossRef](#)] [[PubMed](#)]
85. Goh, W.W.B.; Wang, W.; Wong, L. Why batch effects matter in omics data, and how to avoid them. *Trends Biotechnol.* **2017**, *35*, 498–507. [[CrossRef](#)] [[PubMed](#)]
86. Bell, A.W.; Deutsch, E.W.; Au, C.E.; Kearney, R.E.; Beavis, R.; Sechi, S.; Nilsson, T.; Bergeron, J.J. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Methods* **2009**, *6*, 423–430. [[CrossRef](#)] [[PubMed](#)]
87. Sumner, L.W.; Amberg, A.; Barrett, D.; Beale, M.H.; Beger, R.; Daykin, C.A.; Fan, T.W.; 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](#)]
88. Schrimpe-Rutledge, A.C.; Codreanu, S.G.; Sherrod, S.D.; McLean, J.A. Untargeted metabolomics strategies-challenges and emerging directions. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1897–1905. [[CrossRef](#)]
89. 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](#)]
90. Guijas, C.; Montenegro-Burke, J.R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A.E.; et al. METLIN: A technology platform for identifying knowns and unknowns. *Anal. Chem.* **2018**, *90*, 3156–3164. [[CrossRef](#)]
91. Vinaixa, M.; Schymanski, E.L.; Neumann, S.; Navarro, M.; Salek, R.M.; Yanes, O. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *Trends Anal. Chem.* **2016**, *78*, 23–35. [[CrossRef](#)]
92. 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**. [[CrossRef](#)]
93. Wang, L.; Xing, X.; Chen, L.; Yang, L.; Su, X.; Rabitz, H.; Lu, W.; Rabinowitz, J.D. Peak annotation and verification engine for untargeted LC-MS metabolomics. *Anal. Chem.* **2019**, *91*, 1838–1846. [[CrossRef](#)]
94. 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](#)] [[PubMed](#)]
95. Domingo-Almenara, X.; Montenegro-Burke, J.R.; Ivanisevic, J.; Thomas, A.; Sidibé, J.; Teav, T.; Guijas, C.; Aisporna, A.E.; Rinehart, D.; Hoang, L.; et al. XCMS-MRM and METLIN-MRM: A cloud library and public resource for targeted analysis of small molecules. *Nat. Methods* **2018**, *15*, 681–684. [[CrossRef](#)] [[PubMed](#)]
96. Ni, Y.; Su, M.; Qiu, Y.; Jia, W.; Du, X. ADAP-GC 3.0: Improved peak detection and deconvolution of co-eluting metabolites from GC/TOF-MS data for metabolomics studies. *Anal. Chem.* **2016**, *88*, 8802–8811. [[CrossRef](#)] [[PubMed](#)]
97. Korf, A.; Jeck, V.; Schmid, R.; Helmer, P.O.; Hayen, H. Lipid species annotation at double bond position level with custom databases by extension of the mzmine 2 open-source software package. *Anal. Chem.* **2019**, *91*, 5098–5105. [[CrossRef](#)] [[PubMed](#)]
98. Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D. MZmine 2 data-preprocessing to enhance molecular networking reliability. *Anal. Chem.* **2017**, *89*, 7836–7840. [[CrossRef](#)] [[PubMed](#)]
99. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [[CrossRef](#)]
100. 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. [[CrossRef](#)]
101. Lai, Z.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K.; et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* **2017**, *15*, 53. [[CrossRef](#)]
102. Beaulieu-Jones, B.K.; Greene, C.S. Reproducibility of computational workflows is automated using continuous analysis. *Nat. Biotechnol.* **2017**, *35*, 342–346. [[CrossRef](#)]
103. Chong, J.; Yamamoto, M.; Xia, J. MetaboAnalystR 2.0: From raw spectra to biological insights. *Metabolites* **2019**, *9*, 57. [[CrossRef](#)]

104. Caron, C.; Duperier, C.; Jacob, D.; Thévenot, E.A.; Giacomoni, F.; Le Corguillé, G.; Martin, J.-F.; Tremblay-Franco, M.; Landi, M.; Pétéra, M.; et al. Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics* **2014**, *31*, 1493–1495. [[CrossRef](#)]
105. Ruttkies, C.; Schober, D.; Peters, K.; Neumann, S.; Gonzalez-Beltran, A.; Izzo, M.; Rocca-Serra, P.; Sansone, S.-A.; Johnson, D.; Reed, M.A.C.; et al. PhenoMeNal: Processing and analysis of metabolomics data in the cloud. *GigaScience* **2018**, *8*. [[CrossRef](#)]
106. Weber, R.J.M.; Lawson, T.N.; Salek, R.M.; Ebbels, T.M.D.; Glen, R.C.; Goodacre, R.; Griffin, J.L.; Haug, K.; Koulman, A.; Moreno, P.; et al. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. *Metabolomics* **2016**, *13*, 12. [[CrossRef](#)] [[PubMed](#)]
107. Armitage, E.G.; Godzien, J.; Alonso-Herranz, V.; Lopez-Gonzalvez, A.; Barbas, C. Missing value imputation strategies for metabolomics data. *Electrophoresis* **2015**, *36*, 3050–3060. [[CrossRef](#)] [[PubMed](#)]
108. Taylor, S.L.; Ruhaak, L.R.; Kelly, K.; Weiss, R.H.; Kim, K. Effects of imputation on correlation: Implications for analysis of mass spectrometry data from multiple biological matrices. *Brief. Bioinform.* **2017**, *18*, 312–320. [[CrossRef](#)]
109. Hrydziusko, O.; Viant, M.R. Missing values in mass spectrometry based metabolomics: An undervalued step in the data processing pipeline. *Metabolomics* **2012**, *8*, 161–174. [[CrossRef](#)]
110. Jin, Z.; Yu, T.; Kang, J. Missing value imputation for LC-MS metabolomics data by incorporating metabolic network and adduct ion relations. *Bioinformatics* **2017**, *34*, 1555–1561. [[CrossRef](#)]
111. Voillet, V.; Besse, P.; Liaubet, L.; San Cristobal, M.; Gonzalez, I. Handling missing rows in multi-omics data integration: Multiple imputation in multiple factor analysis framework. *BMC Bioinform.* **2016**, *17*, 402. [[CrossRef](#)]
112. Gerl, M.J.; Klose, C.; Surma, M.A.; Fernandez, C.; Melander, O.; Mannisto, S.; Borodulin, K.; Havulinna, A.S.; Salomaa, V.; Ikonen, E.; et al. Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. *PLoS Biol.* **2019**, *17*, e3000443. [[CrossRef](#)]
113. Vvedenskaya, O.; Wang, Y.; Ackerman, J.M.; Knittelfelder, O.; Shevchenko, A. Analytical challenges in human plasma lipidomics: A winding path towards the truth. *TrAC Trends Anal. Chem.* **2019**, *120*, 115277. [[CrossRef](#)]
114. Burla, B.; Arita, M.; Arita, M.; Bendt, A.K.; Cazenave-Gassiot, A.; Dennis, E.A.; Ekroos, K.; Han, X.; Ikeda, K.; Liebisch, G.; et al. MS-based lipidomics of human blood plasma: A community-initiated position paper to develop accepted guidelines. *J. Lipid Res.* **2018**, *59*, 2001–2017. [[CrossRef](#)] [[PubMed](#)]
115. Bowden, J.A.; Heckert, A.; Ulmer, C.Z.; Jones, C.M.; Koelmel, J.P.; Abdullah, L.; Ahonen, L.; Alnouti, Y.; Armando, A.M.; Asara, J.M.; et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-Metabolites in Frozen Human Plasma. *J. Lipid Res.* **2017**, *58*, 2275–2288. [[CrossRef](#)] [[PubMed](#)]
116. Liebisch, G.; Ahrends, R.; Arita, M.; Arita, M.; Bowden, J.A.; Ejsing, C.S.; Griffiths, W.J.; Holcapek, M.; Köfeler, H.; Mitchell, T.W.; et al. Lipidomics needs more standardization. *Nat. Metab.* **2019**, *1*, 745–747. [[CrossRef](#)]
117. Rustam, Y.H.; Reid, G.E. Analytical challenges and recent advances in mass spectrometry based lipidomics. *Anal. Chem.* **2018**, *90*, 374–397. [[CrossRef](#)] [[PubMed](#)]
118. Lv, W.; Shi, X.; Wang, S.; Xu, G. Multidimensional liquid chromatography-mass spectrometry for metabolomic and lipidomic analyses. *TrAC Trends Anal. Chem.* **2019**, *120*, 115302. [[CrossRef](#)]
119. Barupal, D.K.; Fan, S.; Wancewicz, B.; Cajka, T.; Sa, M.; Showalter, M.R.; Baillie, R.; Tenenbaum, J.D.; Louie, G.; Kaddurah-Daouk, R.; et al. Generation and quality control of lipidomics data for the alzheimer’s disease neuroimaging initiative cohort. *Sci. Data* **2018**, *5*, 180263. [[CrossRef](#)]
120. Fan, S.; Kind, T.; Cajka, T.; Hazen, S.L.; Tang, W.H.W.; Kaddurah-Daouk, R.; Irvin, M.R.; Arnett, D.K.; Barupal, D.K.; Fiehn, O. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. *Anal. Chem.* **2019**, *91*, 3590–3596. [[CrossRef](#)]
121. Quell, J.D.; Römisch-Margl, W.; Haid, M.; Krumsiek, J.; Skurk, T.; Halama, A.; Stephan, N.; Adamski, J.; Hauner, H.; Mook-Kanamori, D.; et al. Characterization of bulk phosphatidylcholine compositions in human plasma using side-chain resolving lipidomics. *Metabolites* **2019**, *9*, 109. [[CrossRef](#)]
122. Gathungu, R.M.; Larrea, P.; Sniatynski, M.J.; Marur, V.R.; Bowden, J.A.; Koelmel, J.P.; Starke-Reed, P.; Hubbard, V.S.; Kristal, B.S. Optimization of electrospray ionization source parameters for lipidomics to reduce misannotation of in-source fragments as precursor ions. *Anal. Chem.* **2018**, *90*, 13523–13532. [[CrossRef](#)]

123. Leaptrot, K.L.; May, J.C.; Dodds, J.N.; McLean, J.A. Ion mobility conformational lipid atlas for high confidence lipidomics. *Nat. Commun.* **2019**, *10*, 985. [[CrossRef](#)]
124. Lipidomic Standards Initiative. The Lipidomics Standard Initiative. Available online: <https://lipidomics-standards-initiative.org/> (accessed on 30 November 2019).
125. Liebisch, G.; Vizcaino, J.A.; Kofeler, H.; Trotschmuller, M.; Griffiths, W.J.; Schmitz, G.; Spener, F.; Wakelam, M.J. Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.* **2013**, *54*, 1523–1530. [[CrossRef](#)] [[PubMed](#)]
126. Pauling, J.K.; Hermansson, M.; Hartler, J.; Christiansen, K.; Gallego, S.F.; Peng, B.; Ahrends, R.; Ejsing, C.S. Proposal for a common nomenclature for fragment ions in mass spectra of lipids. *PLoS ONE* **2017**, *12*, e0188394. [[CrossRef](#)]
127. Koelmel, J.P.; Kroeger, N.M.; Ulmer, C.Z.; Bowden, J.A.; Patterson, R.E.; Cochran, J.A.; Beecher, C.W.W.; Garrett, T.J.; Yost, R.A. LipidMatch: An automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinform.* **2017**, *18*, 331. [[CrossRef](#)] [[PubMed](#)]
128. Kind, T.; Liu, K.H.; Lee, D.Y.; DeFelice, B.; Meissen, J.K.; Fiehn, O. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* **2013**, *10*, 755–758. [[CrossRef](#)] [[PubMed](#)]
129. Fahy, E.; Subramaniam, S.; Murphy, R.C.; Nishijima, M.; Raetz, C.R.; Shimizu, T.; Spener, F.; van Meer, G.; Wakelam, M.J.; Dennis, E.A. Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **2009**, *50*, S9–S14. [[CrossRef](#)] [[PubMed](#)]
130. Clair, G.; Reehl, S.; Stratton, K.G.; Monroe, M.E.; Tfaily, M.M.; Ansong, C.; Kyle, J.E. Lipid Mini-On: Mining and ontology tool for enrichment analysis of lipidomic data. *Bioinformatics* **2019**, *35*, 4507–4508. [[CrossRef](#)] [[PubMed](#)]
131. Molenaar, M.R.; Jeucken, A.; Wassenaar, T.A.; van de Lest, C.H.A.; Brouwers, J.F.; Helms, J.B. LION/web: A web-based ontology enrichment tool for lipidomic data analysis. *Gigascience* **2019**, *8*. [[CrossRef](#)]
132. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *New Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)]
133. Trainor, P.J.; DeFilippis, A.P.; Rai, S.N. Evaluation of classifier performance for multiclass phenotype discrimination in untargeted metabolomics. *Metabolites* **2017**, *7*, 30. [[CrossRef](#)]
134. Long, N.P.; Jung, K.H.; Anh, N.H.; Yan, H.H.; Nghi, T.D.; Park, S.; Yoon, S.J.; Min, J.E.; Kim, H.M.; Lim, J.H.; et al. An integrative data mining and omics-based translational model for the identification and validation of oncogenic biomarkers of pancreatic cancer. *Cancers* **2019**, *11*, 155. [[CrossRef](#)]
135. Gosiewska, A.; Biecek, P. iBreakDown: Uncertainty of Model Explanations for Non-Additive Predictive Models. Available online: <https://arxiv.org/abs/1903.11420> (accessed on 30 November 2019).
136. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. Available online: <https://arxiv.org/abs/1602.04938> (accessed on 30 November 2019).
137. Long, N.P.; Park, S.; Anh, N.H.; Min, J.E.; Yoon, S.J.; Kim, H.M.; Nghi, T.D.; Lim, D.K.; Park, J.H.; Lim, J.; et al. Efficacy of integrating a novel 16-gene biomarker panel and intelligence classifiers for differential diagnosis of rheumatoid arthritis and osteoarthritis. *J. Clin. Med.* **2019**, *8*, 50. [[CrossRef](#)] [[PubMed](#)]
138. Xu, C.; Jackson, S.A. Machine learning and complex biological data. *Genome Biol.* **2019**, *20*, 76. [[CrossRef](#)] [[PubMed](#)]
139. Beirnaert, C.; Peeters, L.; Meysman, P.; Bittremieux, W.; Foubert, K.; Custers, D.; Van der Auwera, A.; Cuykx, M.; Pieters, L.; Covaci, A.; et al. Using expert driven machine learning to enhance dynamic metabolomics data analysis. *Metabolites* **2019**, *9*, 54. [[CrossRef](#)] [[PubMed](#)]
140. Silva, C.; Perestrelo, R.; Silva, P.; Tomás, H.; Câmara, J.S. Breast cancer metabolomics: From analytical platforms to multivariate data analysis. A review. *Metabolites* **2019**, *9*, 102. [[CrossRef](#)]
141. Broadhurst, D.I.; Kell, D.B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006**, *2*, 171–196. [[CrossRef](#)]
142. Long, N.P.; Yoon, S.J.; Anh, N.H.; Nghi, T.D.; Lim, D.K.; Hong, Y.J.; Hong, S.-S.; Kwon, S.W. A systematic review on metabolomics-based diagnostic biomarker discovery and validation in pancreatic cancer. *Metabolomics* **2018**, *14*, 109. [[CrossRef](#)]
143. Patti, G.J.; Tautenhahn, R.; Siuzdak, G. Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nat. Protoc.* **2012**, *7*, 508–516. [[CrossRef](#)]
144. 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* **2017**, *14*, 7. [[CrossRef](#)]

145. Hicks, S.C.; Peng, R.D. Elements and Principles for Characterizing Variation between Data Analyses. Available online: <https://arxiv.org/abs/1903.07639v2> (accessed on 30 November 2019).
146. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G.M. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann. Intern. Med.* **2015**, *162*, 55–63. [[CrossRef](#)]
147. Bossuyt, P.M.; Reitsma, J.B.; Bruns, D.E.; Gatsonis, C.A.; Glasziou, P.P.; Irwig, L.; Lijmer, J.G.; Moher, D.; Rennie, D.; de Vet, H.C.; et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **2015**, *351*, h5527. [[CrossRef](#)]
148. Moore, H.M.; Kelly, A.B.; Jewell, S.D.; McShane, L.M.; Clark, D.P.; Greenspan, R.; Hayes, D.F.; Hainaut, P.; Kim, P.; Mansfield, E.A.; et al. Biospecimen reporting for improved study quality (BRISQ). *Cancer Cytopathol.* **2011**, *119*, 92–101. [[CrossRef](#)] [[PubMed](#)]
149. Lumberras, B.; Porta, M.; Marquez, S.; Pollan, M.; Parker, L.A.; Hernandez-Aguado, I. QUADOMICS: An adaptation of the Quality Assessment of Diagnostic Accuracy Assessment (QUADAS) for the evaluation of the methodological quality of studies on the diagnostic accuracy of ‘-omics’-based technologies. *Clin. Biochem.* **2008**, *41*, 1316–1325. [[CrossRef](#)] [[PubMed](#)]
150. Considine, E.; Salek, R. A tool to encourage minimum reporting guideline uptake for data analysis in metabolomics. *Metabolites* **2019**, *9*, 43. [[CrossRef](#)] [[PubMed](#)]
151. von Elm, E.; Altman, D.G.; Egger, M.; Pocock, S.J.; Gøtzsche, P.C.; Vandenbroucke, J.P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Int. J. Surg.* **2014**, *12*, 1495–1499. [[CrossRef](#)]
152. 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](#)]
153. Hasin, Y.; Seldin, M.; Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **2017**, *18*, 83. [[CrossRef](#)]
154. Veenstra, T.D. Metabolomics: The final frontier? *Genome Med.* **2012**, *4*, 40. [[CrossRef](#)]
155. Zhou, W.; Sailani, M.R.; Contrepois, K.; Zhou, Y.; Ahadi, S.; Leopold, S.R.; Zhang, M.J.; Rao, V.; Avina, M.; Mishra, T.; et al. Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* **2019**, *569*, 663–671. [[CrossRef](#)]
156. Bakker, O.B.; Aguirre-Gamboa, R.; Sanna, S.; Oosting, M.; Smeekens, S.P.; Jaeger, M.; Zorro, M.; Vösa, U.; Withoff, S.; Netea-Maier, R.T.; et al. Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses. *Nat. Immunol.* **2018**, *19*, 776–786. [[CrossRef](#)]
157. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85. [[CrossRef](#)]
158. Shapiro, J.A. Revisiting the central dogma in the 21st century. *Ann. N. Y. Acad. Sci.* **2009**, *1178*, 6–28. [[CrossRef](#)] [[PubMed](#)]
159. Olivier, M.; Asmis, R.; Hawkins, G.A.; Howard, T.D.; Cox, L.A. The need for multi-omics biomarker signatures in precision medicine. *Int. J. Mol. Sci.* **2019**, *20*, 4781. [[CrossRef](#)] [[PubMed](#)]
160. Blum, B.C.; Mousavi, F.; Emili, A. Single-platform ‘multi-omic’ profiling: Unified mass spectrometry and computational workflows for integrative proteomics-metabolomics analysis. *Mol. Omics* **2018**, *14*, 307–319. [[CrossRef](#)] [[PubMed](#)]
161. Schwarz, E.; Torrey, E.F.; Guest, P.C.; Bahn, S. Biomarker discovery in human cerebrospinal fluid: The need for integrative metabolome and proteome databases. *Genome Med.* **2012**, *4*, 39. [[CrossRef](#)] [[PubMed](#)]
162. Darst, B.F.; Lu, Q.; Johnson, S.C.; Engelman, C.D. Integrated analysis of genomics, longitudinal metabolomics, and Alzheimer’s risk factors among 1,111 cohort participants. *Genet. Epidemiol.* **2019**, *43*, 657–674. [[CrossRef](#)] [[PubMed](#)]
163. Lloyd-Price, J.; Arze, C.; Ananthakrishnan, A.N.; Schirmer, M.; Avila-Pacheco, J.; Poon, T.W.; Andrews, E.; Ajami, N.J.; Bonham, K.S.; Brislawn, C.J.; et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **2019**, *569*, 655–662. [[CrossRef](#)]
164. Adamski, J. Genome-wide association studies with metabolomics. *Genome Med.* **2012**, *4*, 34. [[CrossRef](#)]
165. Schüssler-Fiorenza Rose, S.M.; Contrepois, K.; Moneghetti, K.J.; Zhou, W.; Mishra, T.; Mataraso, S.; Dagan-Rosenfeld, O.; Ganz, A.B.; Dunn, J.; Hornburg, D.; et al. A longitudinal big data approach for precision health. *Nat. Med.* **2019**, *25*, 792–804. [[CrossRef](#)]

166. Yousri, N.A.; Fakhro, K.A.; Robay, A.; Rodriguez-Flores, J.L.; Mohny, R.P.; Zeriri, H.; Odeh, T.; Kader, S.A.; Aldous, E.K.; Thareja, G.; et al. Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population. *Nat. Commun.* **2018**, *9*, 333. [[CrossRef](#)]
167. Yet, I.; Menni, C.; Shin, S.Y.; Mangino, M.; Soranzo, N.; Adamski, J.; Suhre, K.; Spector, T.D.; Kastenmuller, G.; Bell, J.T. Genetic influences on metabolite levels: A comparison across metabolomic platforms. *PLoS ONE* **2016**, *11*, e0153672. [[CrossRef](#)]
168. Pinu, F.R.; Beale, D.J.; Paten, A.M.; Kouremenos, K.; Swarup, S.; Schirra, H.J.; Wishart, D. Systems biology and multi-omics integration: Viewpoints from the metabolomics research community. *Metabolites* **2019**, *9*, 76. [[CrossRef](#)] [[PubMed](#)]
169. Ebrahim, A.; Brunk, E.; Tan, J.; O'Brien, E.J.; Kim, D.; Szubin, R.; Lerman, J.A.; Lechner, A.; Sastry, A.; Bordbar, A.; et al. Multi-omic data integration enables discovery of hidden biological regularities. *Nat. Commun.* **2016**, *7*, 13091. [[CrossRef](#)] [[PubMed](#)]
170. Huang, S.; Chaudhary, K.; Garmire, L.X. More is better: Recent progress in multi-omics data integration methods. *Front. Genet.* **2017**, *8*, 84. [[CrossRef](#)] [[PubMed](#)]
171. Rappoport, N.; Shamir, R. Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. *Nucleic Acids Res.* **2018**, *46*, 10546–10562. [[CrossRef](#)] [[PubMed](#)]
172. McGuire, A.L.; Basford, M.; Dressler, L.G.; Fullerton, S.M.; Koenig, B.A.; Li, R.; McCarty, C.A.; Ramos, E.; Smith, M.E.; Somkin, C.P.; et al. Ethical and practical challenges of sharing data from genome-wide association studies: The eMERGE Consortium experience. *Genome Res.* **2011**, *21*, 1001–1007. [[CrossRef](#)] [[PubMed](#)]
173. Spicer, R.A.; Steinbeck, C. A lost opportunity for science: Journals promote data sharing in metabolomics but do not enforce it. *Metabolomics* **2018**, *14*, 16. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).