Vijay Soni
Travis E. Hartman *Editors*

Metabolomics

Recent Advances and Future Applications

Springer

Network Development and Comparison in Lipidomics and Metabolomics

Thao Nguyen-Tran, Qassim Alkassir, Steffany A. L. Bennett, and Miroslava Cuperlovic-Culf

Abbreviations

AI	Artificial intelligence
GNN	Graph Neural Network
KNC	Known node correspondence methods
LASSO	Least absolute shrinkage and selection operator
MCMC	Markov-Chain-Monte-Carlo
ML	Machine learning
NLP	Natural Language Processing
PLS	Partial least squares
RF	Random forest classification method
RNA-seq	RNA sequencing

T. Nguyen-Tran · S. A. L. Bennett
Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ottawa, ON, Canada

Neural Regeneration Laboratory, Ottawa Institute of Systems Biology, Brain and Mind Research Institute, University of Ottawa, Ottawa, ON, Canada

Department of Chemistry and Biomolecular Sciences, Centre for Catalysis Research and Innovation, University of Ottawa, Ottawa, ON, Canada

Q. Alkassir
Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ottawa, ON, Canada

Neural Regeneration Laboratory, Ottawa Institute of Systems Biology, Brain and Mind Research Institute, University of Ottawa, Ottawa, ON, Canada

M. Cuperlovic-Culf (✉)
Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ottawa, ON, Canada

Neural Regeneration Laboratory, Ottawa Institute of Systems Biology, Brain and Mind Research Institute, University of Ottawa, Ottawa, ON, Canada

Digital Technologies Research Centre, National Research Council of Canada, Ottawa, ON, Canada
e-mail: Miroslava.Cuperlovic-Culf@nrc-cnrc.gc.ca

V. Soni, T. E. Hartman (eds.), *Metabolomics*,
https://doi.org/10.1007/978-3-031-39094-4_2

scRNAseq	Single-cell RNA sequencing
UNC	Unknown node correspondence methods
WGCNA	Weighted network approach

1 Introduction

Biological networks can be described as a set of biological molecules represented as nodes (also called vertices) connected, via a measure of the strength of a biomolecular interaction, by edges. Metabolic and lipidomic networks connect metabolites and lipids as nodes through edges representing the chemical or metabolic reactions that generate product from substrate (or reactant). The association (edges) between molecules (nodes) can stem from different types of relationships or interactions that provide information about the chemical or metabolic reactions that link two nodes, their correlation or co-behavior in a specific condition, or their chemical properties that define node relationships. Metabolic reaction networks, connecting metabolites as nodes through chemical reactions as edges, describe systems that are responsible for maintaining homeostasis and regulating cellular functions. The construction of complete chemical reaction networks of metabolism are informed by the following:

(i) The precursor and product of an enzymatic reaction
(ii) Reaction stoichiometry and enzyme kinetics
(iii) Reaction directionality
(iv) Subcellular localization of reaction

Pathway is a set of context-dependent interactions with clear beginning and end and often delineated directionality.

Network is any structure of nodes connected with edges. Metabolic network aims to provide context-free representation of the complete process often by combining pathways.

Inclusion of all these properties would allow both network analysis and simulation. Alternatively, network development that only includes undirected, correlation information provide networks of relationships without causality or modeling utility.

Another layer of metabolic network complexity is that the subcellular membrane-bound compartments allow for the separation of different environments within the cell while at the same time bringing enzymes and their corresponding substrates in close proximity. Consequently, compartmentalization provides the optimal condition for enzymatic reaction. In constructing metabolic networks, information associated to subcellular compartmentalization of enzymatic reaction can be incorporated when either data or knowledge is available. Compartments provide optimal conditions for function of enzymes and additionally allow equivalent chemicals to be utilized for different purposes. Enzymes from the same family can reside across different compartments with members possibly functioning under slightly different conditions.

Networks or graphs, consist of nodes, i.e. vertices that correspond to objects, for example metabolites or lipids, and edges that show connections between objects. Edges can have *weights* indicating strengths of connections. In *undirected* graphs, edges have no direction, and only show relationship not causality. In *directed* graphs, edges have direction indicating one-way relationships showing that edge can only transverses in a single direction. In a *bipartite* graph, vertices are separated into two sets, where nodes from one set can be unidirectionally connected to any node in the other set, but there are no edges within nodes of the same set. Graphs can be also *multi-edge* containing multiple edges between same two nodes, for example, edge for data and edge for knowledge. *Hypergraph* consists of nodes and hyperedges, where an edge can join any number of vertices. Graph is called *connected* if there is a connection between any two points and *complete* if every pair of vertices is connected by a unique edge.

Through the web of metabolic reactions, biological systems are in constant flux with metabolites in a dynamic interaction with other biological molecules, experiencing continual chemical change. Metabolites concentrations, their destiny in a system, are thus determined by other members of the network, and greater understanding of either individual metabolites' behavior or the biological systems can only come from the analysis of the network of associates.

2 Network Development Methods

High-throughput bioanalytical methods are providing increasingly detailed molecular coverage in a variety of sample types delivering range of datasets that can be explored through network investigation. At the same time, our knowledge about metabolite and lipid functions and processes is increasing, and there is a growing appreciation of the importance of their relationships within pathways and networks. Metabolic networks can be derived from data – data-driven networks, from biological information in knowledge-driven networks or in a hybrid approach combining knowledge and data. Each of these approaches comes with its own set of advantages and disadvantages and the road taken has to be optimized based on the application of interest.

Knowledge-based network development includes combining enzymatic reaction information from databases or known pathways as well as literature derivation of, for example, possible enzyme-metabolite relationships. Derivation of this information can be done directly from existing databases (some examples shown in Table 1) or from literature search, manually or with the help of Natural Language Processing methods (NLP). Clearly, in this approach, the network depends on the level of prior biological knowledge provided in selected information resource. These types of networks provide a map for observing interactions from the data in a context of predefined possibilities. Knowledge-based networks can be small, for example, only observing individual pathways and showing relationships between only small

Table 1 Examples of resources for knowledge-based network and pathway information

Pathway and network information databases	Brief information	Number of metabolic pathways	Reference and site
Kyoto Encyclopedia of Genes and Genomes (KEGG)	Major database of metabolic pathways	~6500 (Nov 10, 2022)	[18] https://www.kegg.jp/
LIPID MAPS Reaction Explorer	Lipid reactions database is linked to the LIPID MAPS® Structure Database (LMSD)	~47,800 unique lipid structures	https://www.lipidmaps.org/
SwissLipids	Database and searchable site for lipid metabolic reactions	779,688 lipid species	[1] https://www.swisslipids.org
Small Molecule Pathway Database (SMDB)	Database containing detailed information about small molecule metabolites found in the human body	~29,011	http://www.smpdb.ca/
Rhea	Expert-curated knowledgebase of chemical and transport reactions of biological interest	15,453 reactions with 12,972 unique compounds	Rhea, the reaction knowledgebase in 2022 [4] https://www.rhea-db.org/
Reactome	Relational database of signaling and metabolic molecules and their relations organized into biological pathways and processes	Overall metabolic map	https://reactome.org/ [13]
MetaCyc (bioCyc – add separately)	Curated database of experimentally elucidated metabolic pathways from all domains of life	~20,000 across different species	https://metacyc.org/
Biocarta	Database of maps of biochemical pathways including metabolic and signaling transduction pathways	Metabolic maps	https://maayanlab.cloud/Harmonizome/dataset/Biocarta+Pathways
BioPAN	Web-based application for the visualization of lipidomics data on a mammalian lipidome metabolic pathways	Data drive visual representation of networks	https://lipidmaps.org/biopan/ [14]
LINEX	Lipid Network Explorer is a Webapp used to analyze lipid metabolic networks and provides data-specific network from user data	Data drive visual representation of networks	https://exbio.wzw.tum.de/linex/ [21]

subset of molecules based on some property of interest. Alternatively, these networks can represent a large-scale efforts in building genome-scale, i.e. complete, metabolic processes networks for subsequent system modeling [11]. Some examples of metabolic knowledge resources as well as knowledge-based networks are listed in Table 1. An advantage of knowledge-based networks is that they do not rely on the quality, quantity, or accuracy of the data in user's possession and thus will not be biased by small, sparse, or erroneous datasets. However, as knowledge-based networks depend on the current biological knowledge, it is natural to assume that they still have gaps or missing links in information that have not been discovered yet or is not readily available in resources utilized for creation of a network. Also, with different level of knowledge available for distinct biological systems number of known interactions varies between species as well as metabolic processes. As an example, Recon3D represents the most comprehensive human metabolic network model to date. The version of this genome-scale model made available on MetabolicAtlas has 13,070 reactions and 8369 metabolites (current version available in metabolicatlas.org [32]. When building metabolic networks for other species, a number of methods are made available including PathoLogic [20], which is used to build pathways from predicted enzymes that can be further validated using Semi-Automated Validation infrastructure (SAVI) software applying range of curation decisions [34]. With lipidomics methods becoming increasingly powerful and providing concentrations for hundreds of individual lipid molecules, there is an increasing effort to also deliver corresponding knowledge-based networks for lipidome. BioPAN [14] and LINEX [21] provide users with the pathway mapping for lipidomic dataset, where in both cases knowledge-based lipid networks are used to provide insights about functional lipid associations.

An alternative knowledge-based approach relies on the knowledge of properties of the molecular set rather than knowledge of biological interactions. In this case, chemical ontologies or molecular characteristics can be used to educate building of relationships among compounds. In the example of ChEBI approach [15], users' entry is represented in a network based on chemical properties information. This is a very useful approach for exploring known properties for molecules of interest; however, in this case network is developed for representation of properties, not for further utilization and can be a useful way to analyze chemical properties of a selected set of compounds. However this approach does not provide avenue for analysis of network interactions, only organization of known information. An alternative way to explore statistical enrichment of molecular groups based on chemical ontologies as well as structural similarities is provided by ChemRICH [6]. In this method, the goal is to step away from reliance on often limiting pathway information in obtaining representation of the set and instead use structural similarities and chemical ontologies to map molecules (metabolites or lipids). This approach follows the notion that chemically similar compounds remain in biochemical proximity [7], thus possibly providing a way for assigning unknown molecules based on their properties and network clusters. Chemical ontology or properties networks can

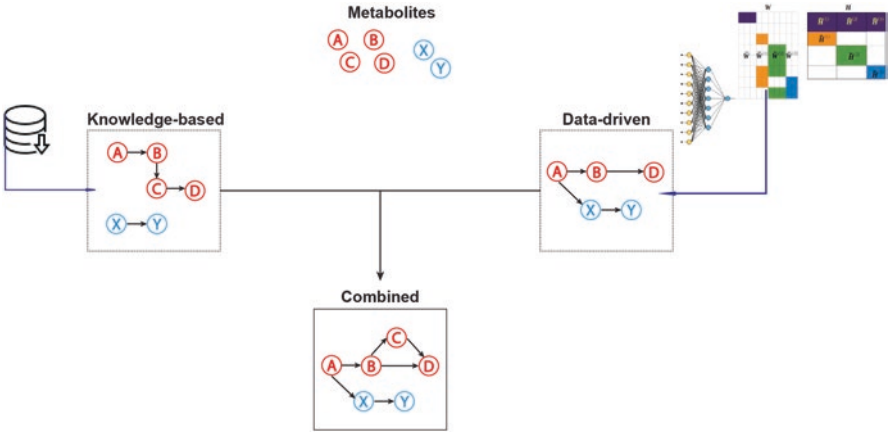

Fig. 1 Most methods for network development are combining knowledge and data to give the most accurate network for features of interest. Based on information in databases and literature in knowledge-based approach, the user creates connections between features. In data-driven approach, the user relies on data and mathematical procedure to establish relationships. Combining the results of two approaches is done in order to reduce errors of the two methods and establish closest network to reality

be utilized to determine related compounds through their shared class belongings, where graph distance between molecular node and a class node can be used to quantify relatedness between pairs of compounds (Fig. 1).

As a rule, *data-driven methods* for network derivation depend on the availability of datasets of sufficiently high quality and quantity and rely on a variety of mathematical tools to build network directly from the data. Network edge determination, in this case, searches for dependences or similarities between node behaviors in samples or similarities in node properties based on a measure of choice. Applications for these highly versatile approaches range from spectral assignments [35] to derivation of metabolic or signaling processes functions or dependencies between features or samples (reviewed recently in [3]. A number of methods have originally been developed for general graph theory and a number of them are applied to other omics datasets, but they are in general also appropriate for metabolomics or lipidomics data as well. Data-driven methods can be further combined with knowledge-based networks in *hybrid approaches.* In hybrid methods the attempt is to take advantage of the available knowledge to either initiate network development from the data by using known interactions as a base for growing more extensive networks from data or to threshold fully data-derived networks at the end of the process. Several different approaches have been developed and tested for data-driven applications, without or with combination to knowledge-based networks. These methodologies can be broadly divided into correlation and classification-based network development methods.

3 Correlation-Based Methods for Molecular Network Derivation

Metabolites and lipids that are linked through enzymatic or signaling pathways often show co-dependencies in the values that are represented through correlated changes in their concentrations across samples. Calculation of these pairwise correlations based on the data from metabolomics or lipidomics measurements can provide data-derived adjacency matrix, where it is hypothesized that two metabolites are linked if their correlation value is statistically significant and larger than a user-defined threshold. Correlation between molecules can then be viewed as the network edge value and a way to construct molecular network. General steps in constructing network from correlation analysis of the data are shown in Fig. 2.

The standard approach for correlation derivation is Pearson method, where correlation is calculated as $r = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}}$ for features $\bar{X}$ and $\bar{Y}$ measured across N samples and having mean values of $\bar{X}$ and $\bar{Y}$, respectively. Pearson's correlations are easy to interpret and calculate; however, this method does not accurately determine nonlinear dependencies. Alternative methods include Spearman – rank-based method, distance correlation [38] – calculating distance covariance, or mutual information [37] – machine learning-based correlation analysis method. Correlation network design with any of these approaches does not guarantee the capture of biologically relevant mechanisms nor does it ensure selection of only direct relationships. A number of additional approaches have been developed in order to help narrow correlations down to only significant ones. The

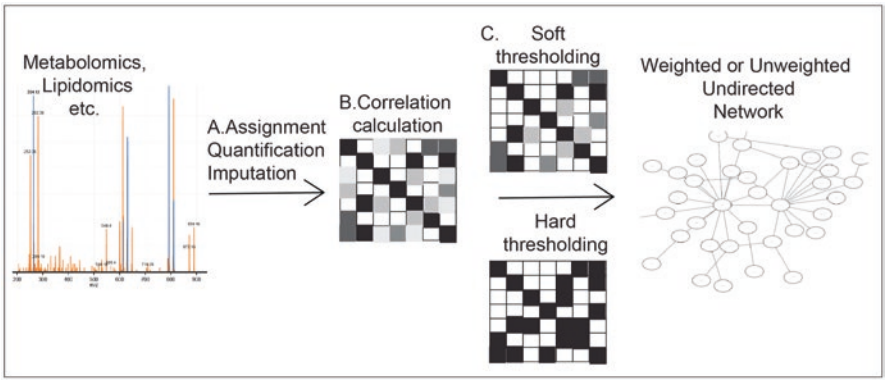

Fig. 2 Steps involved in the construction of molecular network from data using correlation-based methods including (**a**) data quantification and preprocessing; (**b**) determination of pairwise correlations; (**c**) selection of correlations that are statistically significantly different from zero through comparison of *p*-value for significance of the difference of correlation from zero with a significance level observed after appropriate multiple hypothesis testing. Finally, remaining edges can be represented in a network plot with either binary or weighted edges

simplest approach utilizes threshold parameters for selection of relevant edges through p-value of correlation level significance as well as direct correlation level thresholding. In addition, regardless of the method of choice, correlation values are sample size dependent and thus networks have to be constrained by appropriate thresholds for statistical significance (p-value) and/or correlation level for reduction of the effect of sample size as shown by [40]. Once the significant correlation values are selected in a hard-threshold approach, they can be combined in an Adjacency matrix $A = [a_{ij}]$ with entries that are either 1 or 0 decided up on using threshold values as

$$a_{ij} = \begin{cases} 1 \text{ if } r_{ij} \geq r_{\text{threshold}} \text{ and } p_{ij} \leq p_{\text{threshold}} \\ 0 \text{ otherwise} \end{cases}$$

Alternatively, actual correlation values can be kept as edge weights showing pairs' "closeness" level in the soft threshold application.

The statistical significance can be determined using T value calculation forcorrelation as $T = r_{ij}\sqrt{\frac{N-2}{1-r_{ij}^2}}$, where r_{ij} is the correlation value between nodes i and j and N is a number of samples used for correlation calculation. P-value can be determined from T using Student's t cumulative distribution function if normal distribution can be assumed for correlations. Fisher's z-transformation of correlation levels establishes normal distribution for correlation values. Fisher's z-transformed correlation is obtained as $z_{ij} = \frac{1}{2}\ln\frac{1+r_{ij}}{1-r_{ij}}$ and corresponding p-values are calculated as $p_{ij} = 2\left(1-\theta(z_{ij}\sqrt{(N-3)-(M-2)}\right)$ for a sample set with N samples and M features; θ corresponds to cumulative distribution function of standard normal distribution. It is important to notice that p-value in both approaches depend on the sample size or both size of sample and feature space. Correlation values are generally sample size dependent, where in small sample sizes, it is more likely to get spuriously large correlation values due to random sample variations, while with sample size increase individual sample variations contribute less and correlations become a better reflection of the population levels. In smaller populations p-value for correlations is larger, thereby making higher correlation values statistically insignificant and reducing the error caused by artificially large correlations of small sets. Although thresholding to the significance level helps reduce number of low significance edges, it does not guaranty that only direct relationships are kept in the network.

Selection of an optimal p-value correlation coefficient threshold depends on the distribution of the number of edges at different p-values [40] similarly to p-value threshold selection performed based on the p-value distribution in Benjamini-Hochberg FDR multiple hypotheses testing corrected set [9]. Therefore, multiple testing correction should be applied to significance thresholding of correlations, where Bonferroni correction, as the most conservative approach, has been utilized to determine nominal significance level of $a = 0.01$ for a given sample size and can

then be used to determine corrected *p*-value correlation threshold. In this way, *p*-value thresholding follows statistical rules and is determined appropriately for the sample size. Setting of correlation value threshold is less clearly defined. Shen et al. [36] have proposed a theoretically derived threshold for distance correlation that depends exclusively on sample size and is determined as $\tau_N = 2F_\beta^{-1}\left(1-\frac{0.02}{N}\right)-1$, where F_β^{-1} is the inverse cumulative distribution function with symmetric Beta distribution with shape parameter equal to $\frac{1}{2}\left(\frac{1}{2}N(N-3)-1\right)$. This approach provides an interesting, theory-based method for general threshold dependence on sample size in a random dataset giving theoretical lower bound for the threshold; however, it does not include any specific properties of data. Toubiana and Maruenda [40] proposed an iterative approach where topologies of the correlation networks constructed at different levels of threshold are compared and the point of significant change is selected as an analysis threshold. In principle, this approach can be applied to any correlation analysis methodology. In the vicinity of the optimal threshold level for correlation value, the number of network edges is expected to remain stable with gradual increase in *p*-value stringency, going from 0.05 to 0.01 as a cut-off point for statistical significance. Following this assumption, the analysis of the significance of changes in the edge number, using statistical methods such as modified Cox method can be implemented to determine optimal thresholds for both correlation and *p*-value [40].

Edges between nodes in correlation matrix should ideally correspond to metabolic fluxes, that is, reactions in the metabolic network. However, a number of factors influence the correlation results including short-term changes in enzymatic activity due to inhibitors or activations, random fluctuations due to noise in the data or reactions, metabolic processes compartmentalization in cells and organs, involvement of metabolites in multiple pathways, or incomplete experimental coverage. Additionally, through the network, interaction between metabolites could result in indirect correlations leading to highly dense networks. Pairwise correlations that are caused by the presence of mediators can be high and thus remain after correlation thresholding. Issue of indirect correlations is addressed in the Gaussian graphical model approach used to determine partial correlations, that is, remove indirect dependencies. Briefly, Krumsiek et al. [23] showed that when dataset includes many more samples then features correlation matrix has a full rank andit is possible to calculate an inverse of correlation matrix. Following our nomenclature above, partial correlation coefficients can be calculated as $\rho_{ij} = -\vartheta_{ij} / \sqrt{\vartheta_{ii}\vartheta_{jj}}$, where $(\vartheta_{ij}) = R^{-1}$ and R is the correlation adjacency matrix. Partial correlation values correspond to pairwise correlation of metabolites i, j after correction for the correlation through all the other metabolites. Partial correlation in this way accounts for the presence of confounders and covariates, that is, correlation between features through the network. This is a very powerful approach when significant number of samples is made available (Fig. 3 shows graphical explanation of the approach).

In a case when the number of features is larger than the number of samples or if any of the features are a linear combination of other features in the set, the resulting

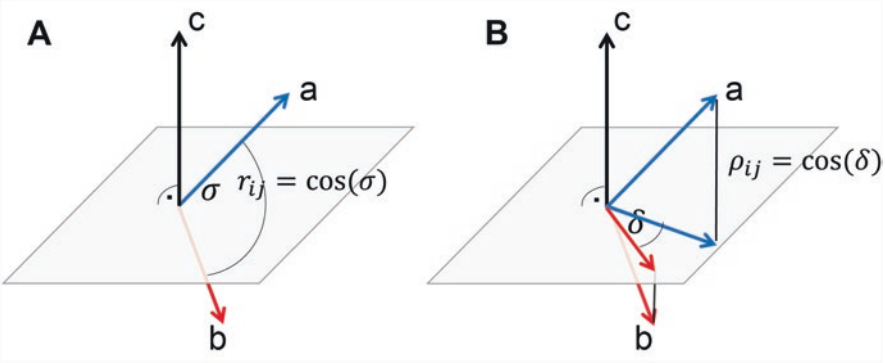

Fig. 3 Geometrical representation of the partial correlation through removal of orthogonal interactions. (**a**) Correlation between two features, *a* and *b* with values across measurements represented vectors in the figure. *a* and *b* are not orthogonal to vector representing feature c and are thus correlated with feature *c*. (**b**) Projections of vectors *a* and *b* onto a plane orthogonal to vector of feature *c* provide value for correlation between and *b* with contributions of *c* removed represented as a cosine of angle between projects

covariance matrix is ill-conditioned, that is, singular and matrix inversion and thus partial correlation calculation through inversion is not possible. GeneNet [30, 33] is initially built for analysis of genomics data; however, it has been now successfully applied to metabolomics as well [12]. In order to allow determination of partial correlations in smaller sample sizes, GeneNet utilizes novel algorithm for shrinking correlation (covariance) matrix making it nonsingular and allowing inversion and derivation of partial correlation matrix possible for all sample sizes. The methodology used for correction of covariance matrix in this approach is analytical shrinkage estimation of covariance and partial correlation matrices on model selection using local FDR multiple testing [33]. In GeneNet, authors decided to shrink correlation matrix toward identity matrix while leaving empirical variances unchanged. The goal of GeneNet is to provide a graph, where edges show direct dependencies between nodes. Alternative methods for covariance matrix shrinkage have been proposed in order to provide improvement in network reconstruction performance [10].

Benedetti et al. [8] proposed an algorithm that optimizes correlation level cutoff selection through maximization of the overlap between the inferred network and available biological, prior, knowledge. With this approach, the focus is on finding network threshold that has the highest overlap with the known biological network, rather than utilizing predefined *p*-value threshold. Several methods have been compared in the analysis including Pearson correlation network, inversion of covariance matrix, and GeneNet, where the approach of Benedetti et al. showed the best performance overall. Interestingly, even in the optimization using a very limited knowledge-based network, authors were able to obtain significantly better network overall.

An alternative approach replaces need for user-defined cut-off point with the user-specified power value for correlations in weighted network approach (WGCNA)

[25]. WGCNA produces a fully connected network with edge weights and can be utilized to determine clusters of co-regulated molecules. In the original version, WGCNA is based on Pearson's correlation matrix; however, recently other correlation approaches have been tested with this method [42]. In WGCNA, the initial step is the calculation of correlation matrix that is transformed by rising all values to a "soft threshold power," that is, value that is used to power the correlation of feature thereby emphasizing strong correlations. Soft power and threshold for the transformed correlation matrix are optimized in this approach for maximal scale-free properties of the selected network. The scale-free network property is optimized by selecting threshold for soft power transformed correlation values that leads to a best linear fit for $\log_{10}(H(d))$ vs $\log_{10}(d)$, where d corresponds to a degree and $H(d)$ is distribution of a degree d across the network. Soft power and threshold are determined for each sample set separately by maximizing R^2 value for the scale-free plot. Specific criteria in selecting correlation exponent in soft thresholding with weights are as follows (following recommendations of [41]: (a) power leads to a network satisfying scale-free topology at least approximately; (b) the mean connectivity should be high so that the network contains enough information (e.g., for module detection); (c) the slope of the regression line between $\log(p(k))$ and $log(k)$ should be negative (typically smaller than -2). The main result of a WGCNA method is network that is used for the determination of clusters, that is, node modules. These modules often represent specific processes, and highly connected modules have been shown to have, for example, major regulatory role. Although this is an interesting approach for selection of major nodes through correlation analysis as well as clustering, WGCNA does not focus on the determine of single-step enzymatic reactions although they are of major interest in metabolomics and lipidomics analysis. Additionally, WGCNA method's assumption of a scale-free network topology is not always appropriate, particularly in metabolic networks.

It should be underlined that all correlation approaches thus far can only be viewed as exploratory methods developed to identify functionally related groups of metabolites or lipids and are not guaranteed to provide only direct mechanistic interactions. Even in the case of partial correlation analysis, it cannot ensure presentation of only direct interaction particularly in a case of partial metabolite or lipid coverage with a range of latent variables. Results of these networks need to be further validated and interpreted using biological knowledge and focused experimental analysis; however, they provide very valuable information to guide future experiments. In spite of their approximate nature, they provide valuable information about network changes across experimental conditions or phenotypes.

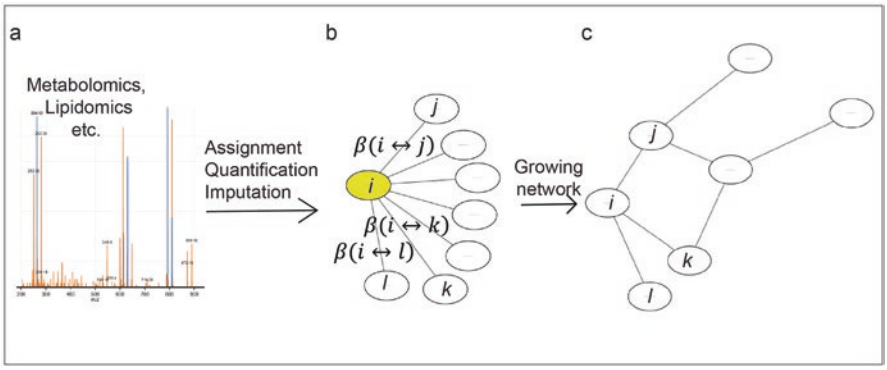

Fig. 4 Steps involved in the construction of molecular network from data using general classification-based methods including (**a**) data quantification and preprocessing; (**b**) determination of regression relationship between features; (**c**) selection of significant edges based on regression analysis and continuing to grow network for all features

4 Classification-Based Methods

Network derivation using classification approaches is once again developed extensively for genomics with metabolomics and lipidomics only starting to benefit from these methods. In this type of approaches, regression of each feature $i \in \{1, \ldots, N\}$ is estimated against all the remaining N-1 features. Edge between pair of features is calculated as $r_{ij} = sign\left(\beta_i^{(j)}\right)\sqrt{\beta_i^{(j)}\beta_j^{(i)}}$, where $\beta_i^{(j)}$ is the regression coefficient of predictor variable x_i for the response x_j (Fig. 4). This approach can be used for the determination of regression-based edges using a variety of methods, outlined in great detail in [33].

Random forest (RF), a classification method, has been proposed by several authors as a base for data-driven network derivation. Two interesting examples are GENIE3 [17], directly applying RF to the dataset and iRafNet [31], combining different data types under a unified RF framework. Both approaches have been developed for genomics but are directly applicable to metabolomics or lipidomics data as well. The GENIE3 model considers characteristics (e.g., concentration or expression level) of each feature as a function of values for all other features sampled randomly from the complete dataset. iRafNet, a subset of potential network partners, is selected based on the information in other provided datasets. When additional data or information is available, iRafNet generally performs better, as it includes prior knowledge, but in the case of fully unique dataset-driven dataset, two methods are equivalent. In this approach, determination of network is viewed as a collection of M subproblems trying to find regulators for M features, where determination of regulators is viewed as a classification equivalent to feature selection problem in classification.

In both approaches, the measure of feature x_i is modeled as a function of the values for other features using RF, that is, tree ensemble. Features that are strong

predictors of x_i are considered as regulators of this feature. Specifically, the importance score for feature x_k as a predictor of x_i, S_{ki} is equal to the total decrease in node impurity following the splitting of samples based on the measurements of feature x_k.

Partial least squares (PLS) regression is also presented as a powerful method for exploring relationships between biological molecules, with application for lipid network derivation presented by Kujala et al. [24]. Connectivity score in this approach is based on the fitting of n PLS models one for each lipid, where each lipid measurements are predicted with n-1 remaining lipids. The latent factors $t_j^{(l)}$ for lipid j are a linear combination of values for all other lipids with PLS determined regression coefficient $c_j^{(l)}$ such as $x_j = \sum_{\vartheta}^{l=1} \beta_l t_j^{(l)} + \varepsilon$; ϑ is number of orthogonal latent factors used for the fitting, that is, number of PLS components used in the model. The connectivity score that can be viewed broadly as the edge between pair of lipids is calculated from PLS parameters as $s_{jk} = \frac{\sum_{l=1}^{\vartheta} \beta_{jl} c_{jk}^{(l)} + \sum_{l=1}^{\vartheta} \beta_{kl} c_{kj}^{(l)}}{2}$ providing weights for the network.

A number of other supervised machine learning methods have been explored as possibly powerful ways to derive feature edges. Applications of LASSO regularization method and Bayesian network inference have been recently reviewed [16]. With LASSO, regularization is explored as a way to reduce model complexity. L1 regularization of LASSO is used as a way to push edge coefficients toward zero, in a way providing a variable selection thereby reducing model complexity. Bayesian networks are directed acyclic graphs providing both dependence and causality between features. In this approach, Markov-Chain-Monte-Carlo (MCMC) procedure is used to estimate precision matrix by searching for the best fit with the data of large space of possible graph configurations in total $2^{\frac{M(M-1)}{2}}$, where M is number of nodes, features. Recently, Graph Neural Network (GNN) approach was used by Alghamdi et al. [2] to model cell-wise metabolic flux from single-cell RNA-seq data. The scFEA method assumes that the modeling of the flux variations of metabolic modules can be performed using nonlinear function of the changes in enzyme levels obtained using transcriptomics and that in all single cells total intermediate substrates flux imbalance is minimized. Using scRNAseq data and GNN, this approach can model flux through metabolic network from transcriptomics data while at the same time providing graph of metabolic modules. Application of neural network analysis directly to metabolomics and lipidomics data for network derivation or analysis is thus far only done by a handful of authors, with an example of deep learning use presented by [5]. Further applications of modern classification, machine learning, and neural network methodologies in metabolomics and lipidomic network derivation and analysis are desired.

5 Analysis of a Network

Once network is built, it motivates exploration using a variety of methods from graph analysis and data mining. Generally, learning based on networks can be broadly divided into node classification, link, that is, edges prediction, network classification, and embedding [29].

Node classification can find a role in the prediction of function of biomolecules using semi-supervised learning by grouping nodes within the entire network. In the context of metabolic or lipidomic networks, node classification can be used to obtain functional similarities between metabolites in one network, for example, biological system, using information known from other biosystems. Similarly, link, that is, edge prediction can be performed as an ML task where known edges are used to train the model that is then used to predict additional, missing links from network data. Graph classification or regression is utilized to predict properties of graphs. When graph is a representation of a molecule, this approach can be used to predict molecular properties. In the context of lipidomic or metabolomics network, this approach can be used to determine similar metabolic outcomes. Graph embedding is most often a preprocessing step that is used to devise representation of nodes or graphs as fixed size vectors making subsequent machine learning analysis easier. Graph Neural Networks (GNNs) are a class of deep learning AI methods designed to analyze network, graph data, unlike regular deep learning approaches appropriate for analysis of vector data. Examples of some GNNs used in biological network data analysis are recently reviewed in [29].

6 Network Comparison Methods

Increasing sizes of datasets and abundance of network development methods and models introduces the next challenge of trying to derive biological information from networks. Analysis can be either aimed at specific characteristics of nodes and edges or overall network structure or, more often, investigation of similarities and differences between networks in different conditions, that is, health and disease or treatment and placebo or changes in the network during time course analysis. Comparison between networks in the context of metabolome or lipidome can be broadly divided into three different goals:

- Comparison of the overall network equality through analysis of the distance between complete networks. In this approach, the goal is to provide numeric estimate of the change for the whole network.
- Determination of the major changes between nodes through either/or analysis of differences in the number of edges or edge weights, where in the context of metabolic network, this would be an indication of changes in metabolite or lipid behavior.

- Determination of the major changes in edges through analysis of the changes in the edge weights. In a metabolic network, this would be an indication of changes in reactions, that is, enzymatic functions.

7 Overall Network Comparison

With network analysis becoming a staple in variety of areas, there is an abundance of methods for network comparison, and the main issue is the selection of the most appropriate approach for the dataset and analysis goals. In an effort to help select optimal methodology, Tantardini et al. [39] have recently presented an appraisal of several popular network comparison methods for mostly undirected, unweighted graphs as well as few methodologies for comparison of directed or weighted networks. Although most methods have been developed to deal with significantly larger networks than what is generally seen in metabolomics and lipidomics, they can easily be applied in these areas as well. Network analysis methods deal with either networks with the same node sets (known node correspondence methods, KNC) or networks with possibly different node sets (unknown node correspondence methods, UNC). UNC methods can be of interest in, for example, comparison of metabolic networks between different species, while KNC methods provide direct comparison of networks derived from data exploring related sample types (e.g., disease vs. control).

The initial task in network comparison is the determination of optimal distance metric for graph analysis. An obvious approach is to directly compute differences between adjacency matrices between networks using any of the distance calculation methods (Euclidean, Jaccard, weighted Jaccard, etc.). Alternatively, the method based on the direct node distance comparison is DeltaCon [22]. DeltaCon compares similarity between all node pairs in two graphs using Matusita distance:

$d = \left(\sum_{N}^{i,j=1} \left(\sqrt{s_{ij}^{A}} - \sqrt{s_{ij}^{B}} \right)^2 \right)^{1/2}$, where $S^A = \left[s_{ij}^{A} \right]$ and $S^B = \left[s_{ij}^{B} \right]$ are similarity matrices for network A and B defined as $S = [I + \epsilon^2 D - \epsilon A]^{-1}$ and A is network adjacency matrix, $D =$ diag (k_i) is degree matrix of node degree (k_i), and ϵ is a small positive constant. Computational cost of DeltaCon is (N^2) thus for networks in metabolomics and lipidome that rarely have more than few hundred nodes this is acceptable. This approach provides more significant change for larger weight changes or for removing edges, while random changes favorably lead to a smaller impact on distance measure.

An alternative method to direct distance analysis is the network alignment or graph isomorphism analysis used to directly compare networks in order to determine conserved and missing nodes and edges across two, pairwise, or multiple network comparisons. Alignment can be performed locally or globally, where local alignment tries to align small regions accurately risking failing in finding large, conserved connections between subgraphs. Global alignment searches for one-to-one mapping of nodes in different networks aiming to overcome shortcomings of

local alignment methods. In general, all alignment methods define an objective function, measure, or a score of alignment quality and utilize a search algorithm that tries to find an optimal solution.

An interesting method developed for metabolic network alignment is H-GRAAL [28] specifically designed for the comparison of metabolic networks between different species. H-GRAAL and a number of related methods are based on the original GRAAL algorithm, which detects statistically significantly similar topological regions in large networks in order to highlight conserved or missing nodes and edges between two or multiple networks. GRAAL approach introduces concept of graphlets that include more detailed description of nodes by incorporating consideration of its degree based on its local neighborhood of connections. Graphlet similarity search is performed over a pair of aligned nodes independently, locally, of all other nodes. Large-scale networks make prioritizing of curation challenging and with uncertainty in the parts of the network that need further consideration and make comparison of networks as well as simulation of systems difficult. Medlock and Papin [27] have recently introduced a ML-based approach for automated metabolic model ensemble-driven uncertainty elimination using statistical learning (AMMEDEUS) as a way to guide curation of genome-scale metabolic models as well as databases.

Clustering of network provides modules of nodes, in our case modules of lipids or metabolites, where molecules within a same module are connected by a short edge paths and strong connections. Node modules can be determined for any network regardless of the method for network derivation. Kujala et al. have shown module cluster comparison method based on PLS-derived association scores [24]. The differences between modular structures in two networks is calculated as $N = 1 - \frac{1}{|L_0|} \sum_{j \in L_0} \frac{\left| F_{1r(j)} \cap F_{2r(j)} \right|}{\left| F_{1r(j)} \cup F_{2r(j)} \right|}$, where $F_{kr(j)}$ is the module, L_0 in network k that contains lipid j. $|L_0|$ shows the number of lipids that belong to modules in both networks. If $N = 0$, modular structures of the two networks are identical; otherwise, p-value for the statistical significance of the modules difference can be calculated as $p(N) = \frac{1}{P} \sum_{\pi} I\left(N(\pi) \geq N\right)$, with sum taken over all P permutations. This approach provides information about the changes in network modules. Differential connectivity for a single node can be obtained using mean absolute distance statistic as $d(j) = \frac{1}{p-1} \sum_{j' \in L, j' \neq j} \left| \hat{s}^1_{jj'} - \hat{s}^2_{jj'} \right|$. With this approach, it is possible to obtain differences bws of nodes as well as for each individual node.

Analysis of maximally dysregulated subnetworks, proposed by Mamano and Hayes [26], uses simulated annealing-supported local search for biological network alignment. In this approach, simulated annealing was shown to provide the optimal solution with better node pairing between networks and good topological and functional similarity scores. Simulated annealing, as a metaheuristic algorithm, is not developed for any specific problem and can be applied to any optimization problem as long as there is a defined objective function and neighbor relationship and there

are solutions for different states. Although the example presented by [26] is in protein-protein interaction network, similar approach can be utilized in metabolomics and lipidomics network comparisons, where score function can combine various topological and biological similarity measures and simulated annealing approach can provide global optimization solution. Once again, methods developed for network comparison in different applications present themselves to metabolomics and lipidomics applications, but it is up to the analyst to select the most appropriate comparison approaches for the network type, size, and analytical question.

8 Network Visualization

Visualization of networks and network components is an extremely important, intuitive way for the interpretation of results, but with variety of network sizes, data types, and applications, there is no single solution. Visualization methods range from the simplest ones showing adjacency matrices to more complex methods that are visualizing force directed layouts in 2D or 3D. Some examples of freely available network visualization software application are listed in Table 2 in addition to many libraries available in different programming languages dedicated to network visualization.

We recommend to the reader to freely explore many possible ways for the visual presentation of networks as the complexity and size of metabolic network necessities optimization of visualization for each application.

9 Conclusions

A number of methods for knowledge-based, chemical ontology, or data analyses network derivation combined with methods for network analysis, comparison, and visualization provide abundance of possibilities, all with their strengths and weaknesses. Knowledge-based networks are clearly limited by gaps in current

Table 2 Examples of free software tools for network visualization

Software application	Brief outline	Reference and Site
Cytoscape	Network visualization and analysis tool with a number of applications developed for bioinformatics	https://cytoscape.org/
OmicsNet	WebGL-based method	https://omicsnet.ca [43]
Gephi	General open graph visualization platform.	https://gephi.org/
Arena3D	Interactive 3D visualization of multimodal networks particularly appropriate for polyomics datasets	http://bib.fleming.gr:3838/Arena3D [19]

information and can create incomplete networks. Ontologies can possibly represent different concepts at different levels of representation. Data-driven networks in general only show co-behaviors and cannot ensure representation of metabolic relationships or direct interdependencies. Major developments are under way and further improvements are absolutely required before modeling of complete metabolic network becomes possible.

Only by combining knowledge, large and diverse datasets and appropriate statistical, machine learning, and modeling tools, we will be able to ultimately obtain truly a complete in silico representation of biological systems' weird and wonderful metabolic network.

References

1. Aimo, L.*, et al.* The SwissLipids knowledgebase for lipid biology. *Bioinformatics* 2015;31(17):2860–2866.
2. Alghamdi, N.*, et al.* A graph neural network model to estimate cell-wise metabolic flux using single-cell RNA-seq data. *Genome Res* 2021;31(10):1867–1884.
3. Amara, A.*, et al.* Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation. *Frontiers in Molecular Biosciences* 2022;9.
4. Bansal, P.*, et al.* Rhea, the reaction knowledgebase in 2022. *Nucleic Acids Res* 2022;50(D1):D693–d700.
5. Baranwal, M.*, et al.* A deep learning architecture for metabolic pathway prediction. *Bioinformatics* 2020;36(8):2547–2553.
6. Barupal, D.K. and Fiehn, O. Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Scientific Reports* 2017;7(1):14567.
7. Barupal, D.K.*, et al.* MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics* 2012;13(1):99.
8. Benedetti, E.*, et al.* A strategy to incorporate prior knowledge into correlation network cutoff selection. *Nature Communications* 2020;11(1):5153.
9. Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995;57(1):289–300.
10. Bernal, V.*, et al.* Exact hypothesis testing for shrinkage-based Gaussian graphical models. *Bioinformatics* 2019;35(23):5011–5017.
11. Brunk, E.*, et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology* 2018;36(3):272–281.
12. Buyukozkan, M.*, et al.* Integrative metabolomic and proteomic signatures define clinical outcomes in severe COVID-19. *iScience* 2022;25(7):104612.
13. Fabregat, A.*, et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res* 2018;46(D1):D649–d655.
14. Gaud, C.*, et al.* BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS. *F1000Res* 2021;10:4.
15. Hastings, J.*, et al.* The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Research* 2012;41(D1):D456–D463.
16. Hawe, J.S., Theis, F.J. and Heinig, M. Inferring Interaction Networks From Multi-Omics Data. *Front Genet* 2019;10:535.
17. Huynh-Thu, V.A. and Geurts, P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci Rep* 2018;8(1):3384.

18. Kanehisa, M. and Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27–30.
19. Karatzas, E.*, et al.* Arena3Dweb: interactive 3D visualization of multilayered networks. *Nucleic Acids Research* 2021;49(W1):W36–W45.
20. Karp, P.D.*, et al.* Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* 2019;22(1):109–126.
21. Köhler, N.*, et al.* Investigating Global Lipidome Alterations with the Lipid Network Explorer. *Metabolites* 2021;11(8).
22. Koutra, D.*, et al.* DeltaCon: Principled Massive-Graph Similarity Function with Attribution. *ACM Trans. Knowl. Discov. Data* 2016;10:28:21–28:43.
23. Krumsiek, J.*, et al.* Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Systems Biology* 2011;5(1):21.
24. Kujala, M.*, et al.* Differential network analysis with multiply imputed lipidomic data. *PLoS One* 2015;10(3):e0121449.
25. Langfelder, P. and Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 2008;9(1):559.
26. Mamano, N. and Hayes, W.B. SANA: simulated annealing far outperforms many other search algorithms for biological network alignment. *Bioinformatics* 2017;33(14):2156–2164.
27. Medlock, G.L. and Papin, J.A. Guiding the Refinement of Biochemical Knowledgebases with Ensembles of Metabolic Networks and Machine Learning. *Cell Syst* 2020;10(1):109–119 e103.
28. Milenkovic, T.*, et al.* Optimal network alignment with graphlet degree vectors. *Cancer Inform* 2010;9:121–137.
29. Muzio, G., O'Bray, L. and Borgwardt, K. Biological network analysis with deep learning. *Brief Bioinform* 2021;22(2):1515–1530.
30. Opgen-Rhein, R. and Strimmer, K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology* 2007;1(1):37.
31. Petralia, F.*, et al.* Integrative random forest for gene regulatory network inference. *Bioinformatics* 2015;31(12):i197–205.
32. Robinson, J.L.*, et al.* An atlas of human metabolism. *Science Signaling* 2020;13(624):eaaz1482.
33. Schäfer, J. and Strimmer, K. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology* 2005;4(1).
34. Schläpfer, P.*, et al.* Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiology* 2017;173(4):2041–2059.
35. Schmid, R.*, et al.* Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nature Communications* 2021;12(1):3832.
36. Shen, C., Priebe, C.E. and Vogelstein, J.T. From Distance Correlation to Multiscale Graph Correlation. *Journal of the American Statistical Association* 2020;115(529):280–291.
37. Song, L., Langfelder, P. and Horvath, S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics* 2012;13(1):328.
38. Székely, G.J. and Rizzo, M.L. Brownian distance covariance. *The annals of applied statistics* 2009;3(4):1236–1265.
39. Tantardini, M.*, et al.* Comparing methods for comparing networks. *Sci Rep* 2019;9(1):17557.
40. Toubiana, D. and Maruenda, H. Guidelines for correlation coefficient threshold settings in metabolite correlation networks exemplified on a potato association panel. *BMC Bioinformatics* 2021;22(1):116.
41. Zhang, B. and Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;4:Article17.
42. Zhang, T. and Wong, G. Gene expression data analysis using Hellinger correlation in weighted gene co-expression networks (WGCNA). *Computational and Structural Biotechnology Journal* 2022;20:3851–3863.
43. Zhou, G.*, et al.* OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics. *Nucleic Acids Res* 2022;50(W1):W527–533.

298. Huber K, Feuchtinger A, Borgmann DM, Li Z, Aichler M, Hauck SM, et al. Novel Approach of MALDI Drug Imaging, Immunohistochemistry, and Digital Image Analysis for Drug Distribution Studies in Tissues. Analytical Chemistry. 2014;86(21):10568–75.
299. Sugihara Y, Végvári Á, Welinder C, Jönsson G, Ingvar C, Lundgren L, et al. A new look at drugs targeting malignant melanoma—An application for mass spectrometry imaging. PROTEOMICS. 2014;14(17–18):1963–70.

Future Perspectives of Metabolomics: Gaps, Planning, and Recommendations

Vijay Soni, Nicholas Bartelo, Annalise Schweickart, Yogesh Chawla, Ankita Dutta, and Sonia Jain

Abbreviations

AD	Alzheimer's disease
ANN	Artificial neural network
AUC	Area under the curve
CE-MS	Capillary electrophoresis - Mass Spectrometry
CGM	Continuous glucose monitor
CNN	Convolutional neural network
DI-MS	Direct infusion-mass spectrometry
DL	Deep learning
EI-MS	Electron ionization mass spectrometry

Authors: "Vijay Soni" and "Nicholas Bartelo" have equally contributed to this chapter.

V. Soni (✉)
Division of Infectious Diseases, Weill Department of Medicine, Weill Cornell Medicine, New York, NY, USA
e-mail: vis2032@med.cornell.edu

N. Bartelo
Department of Physiology and Biophysics, Weill Cornell Medicine, Institute for Computational Biomedicine, Englander Institute for Precision Medicine, New York, NY, USA

A. Schweickart
The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA

Tri-Institutional Computational Biology and Medicine Program, Weill Cornell Medicine, New York, NY, USA

Y. Chawla
Department of Hematology, Mayo Clinic, Rochester, MN, USA

A. Dutta
International AIDS Vaccine Initiative, Gurugram, Haryana, India

S. Jain
Department of Radiation Oncology, Mayo Clinic, Rochester, MN, USA

V. Soni, T. E. Hartman (eds.), *Metabolomics*,
https://doi.org/10.1007/978-3-031-39094-4_14

ELISA	Enzyme-linked immunosorbent assay
FDR	False discovery rate
GC-FID	Gas chromatography-flame ionization detection
GC-MS	Gas chromatography-mass spectrometry
hCG	Human chorionic gonadotropin
HDL	High-density lipoprotein
IEM	Inborn errors of metabolism
LAESI	Laser ablation electrospray ionization
LC-HRMS	Liquid chromatography coupled to high-resolution mass spectrometry
LC-MS	Liquid chromatography-mass spectrometry
LDL	Low-density lipoprotein
LDTs	Laboratory-developed tests
MALDI	Matrix-assisted laser desorption/ionization
mGWAS	Metabolic genome-wide association studies
ML	Machine learning
mQTL	Metabolites and quantitative trait loci
MSI	Metabolomics Standards Initiative
NMR	Nuclear magnetic resonance
PMRN	Pharmacometabolomics Research Network
QSP	Quantitative and systems pharmacology
SIMS	Secondary ion mass spectrometry
SNPs	Single nucleotide polymorphisms
VOC	Volatile organic compound

1 Introduction

Metabolites are the connecting link between the genome and the environment. With the development of novel technologies and bioinformatics approaches, it is now possible to study the global metabolic changes in any organisms and cells. Metabolomes represent an organism's physiological state and can be used to help diagnose and treat a variety of diseases. Metabolomics as a field emerged in the late 1990s with the advent of proteomics [1] and is now rapidly evolving. It is the study of the metabolome that comprises the entire repertoire of small molecules with molecular weights of <1000 Da or <1500 Da excluding biopolymers like proteins or nucleic acids [2]. The small molecules are also referred to as metabolites and are present in human cells, tissues, and body fluids. They can be studied using large-scale detection, quantification, and analysis methodologies. Metabolites are organic and inorganic chemicals and are either reactants, intermediates, or end products generated during biological enzymatic reactions or may be of xenobiotic origin (i.e., the chemicals that are found in living organisms however are not produced by them) and are known to bridge gene functions and nongenetic or phenotypic end points [3–5]. They exhibit variable chemical properties that range from polar hydrophobic compounds and hydrophilic compounds including carbohydrate moieties to nonpolar hydrophobic molecules such as lipids.

Metabolomics has made significant progress in the past two decades. Nevertheless, several aspects of this field are still in the development phase and restrict its application in various domains. These limitations include a restricted detection range, bulk analysis with precise molecular features, a lack of chromatographic methods for better resolution, and the high cost of analytical devices such as mass spectrometers and NMR. However, despite these constraints, metabolomics continues to be a valuable research tool in translational biology, pharmacological medicine, biomarker discovery, and diagnosis. By analyzing the metabolic profiles of patients with different diseases, researchers can gain insights into the underlying mechanisms and identify potential targets for therapy. Furthermore, it has tremendous potential to monitor disease progression and treatment response and detect the side effects of therapies. Besides, the broad range of metabolomics is enabling research in other areas such as agriculture, environmental surveillance, and nutritional biology.

In the coming years, advancements in basic research and healthcare technologies are expected to surpass our current understanding of living organisms. Metabolomics is one such field that holds great promise. Therefore, in this chapter, we aim to provide an overview of the future perspectives of metabolomics in various areas, including precision medicine, personalized nutrition, disease diagnosis, biomarker discovery, single-cell metabolism, the development of novel AI-/ML-based tools for data integration, applications in translational biology, and therapeutic development. Additionally, we will touch upon the topic of metabolic sensors and wearables for disease surveillance. Finally, we have provided recommendations to consider while developing new technologies using metabolomics results. Overall, the future of metabolomics looks bright, with the potential to revolutionize our understanding of biological processes and identify new strategies for enhancing human welfare.

2 Metabolomics for the Masses

While the field of metabolomics has many applications in a clinical setting, there is also a clear use for bringing metabolomics into everyday life. Technologies such as pregnancy tests, breathalyzers, and blood glucose monitors are familiar examples of tools implemented for simple readouts of our physiological state from biomolecular readouts. However, these technologies are only designed for the detection of single molecules: pregnancy tests detect human chorionic gonadotropin (hCG) in urine, breathalyzer tests measure alcohol levels from a breath,[1] and blood glucose monitors measure instantaneous plasma glucose levels. While these narrow searches for

[1] Breathalyzer tests have also been designed for diagnosis of viral and bacterial infections through volatile organic compound detection. Secondary electrospray ionization-mass spectrometry (SESI-MS) on mouse breath could detect infection as well as distinguish between different pathogens and strains [6], and a diagnostic breath test using gas chromatography-mass spectrometry (GC-MS) was approved for emergency use in the Covid-19 pandemic [7, 8].

specific molecules achieve the designed task, the conclusions that can be drawn using these technologies are equally as narrow.

Recently, there has been an emergence of consumer products that expand biomolecule detection through large metabolomic analyses to provide wider insights into our physiological state. The promise of these products includes early disease detection, lifestyle recommendations to improve metabolic health, and future health outcome predictions, all without the need to visit a doctor's office. While consumer products may bring metabolomics closer to everyday life, many challenges still need to be addressed to scale up these analyses.

2.1 Physiological Assessment Through Endogenous Molecules

The accessibility of high-throughput metabolomics has grown, and therefore associations between individual, endogenously derived metabolites and physiological states have become easier to identify [9]. These findings can be leveraged for inference on a new individual's physiology. However, high-throughput metabolomics on the same biological source must be used for the inference to be accurate.

Biofluids commonly used for such analyses are blood (either plasma or serum [10]), saliva, and urine [11]. While the latter two fluids are the easiest to obtain noninvasively, saliva is highly affected by external factors such as hygiene and food intake, and urine is a waste product, meaning its metabolic contents will heavily represent molecules being excreted rather than those being actively used [12]. Blood is therefore the fluid most reflective of bodily processes. Blood circulates in all organs and tissues, making it a "reasonably good metabolic proxy for the entire organism" [13]. As such, businesses building analytical pipelines to bring metabolomics to the masses have opted for blood draw devices to collect samples for further analysis.

Reflecting the methods of published metabolomics-disease associations again, blood is then profiled through liquid chromatography-mass spectrometry (LC-MS) methods, and the resulting spectra are mined for known physiological and disease biomarkers. From plasma draws, there are ways to discern the general function of our organs as each has a unique metabolomic footprint. Organs consume and excrete different metabolites, such as the liver's production of bile acids, the thyroid's production of thyroxine, or the adrenal glands' production of epinephrine and cortisol [13]. Such unique metabolomic footprints allow health-based conclusions to be drawn per organ from metabolomics analysis on a plasma sample.

There have been numerous studies identifying metabolites as biomarkers for non-organ-derived disease as well [14–16]. While such studies have faced criticism due to low sample size and statistical power [17], a 2021 study from Pietzner et al. established robust metabolomic effects in a population of just under 12,000 individuals [9]. Associations were found between both identified and unidentified metabolites and metabolic, heart, and lung diseases, as well as an array of cancer types. The detection of abnormally high or low levels of these metabolites in an

individual's blood could potentially lead to the early detection of disease. This prospect is what drives the establishment of metabolomics tools for the masses – with a single blood draw, a person could discover potentially life-saving information and seek intervention much earlier than they would with classical screening techniques. However, this promise has yet to be realized, and only time will tell if the masses are willing to participate in and act on metabolomics-based health insights.

2.2 Limitations of Current Technologies

2.2.1 Exogenous Molecule Identification

Beyond endogenous compounds, exogenous compounds that enter the body through the environment can also provide insight into the disease state, as discussed in chapter "Exploring Ecometabolomics Landscapes: Progress, Applications, Challenges, and Future Recommendations". In fact, most complex human diseases, from cancer to cardiovascular disease, can be attributed to the environment and the interplay between an individual's genes and their environment [18, 19]. Unfortunately, these environmentally derived compounds, collectively known as our exposome, come from a vastly larger pool of chemicals than endogenous metabolites, and only a small fraction of these are known and have been measured in human tissues [20].

To make matters worse, the exposome is spatiotemporally dependent, meaning exogenous molecules in the blood will depend on the location and time of biofluid collection. Additionally, the physiological response to these molecules varies widely between individuals [21]. This, along with the only very recent development of high-throughput exposure data collection and experimental pipelines, means the number of known associations between exogenous molecules and disease lags behind those of endogenous molecules [22, 23]. The exposome contains contaminants, toxins, pollutants, and carcinogens, all molecules with potentially serious consequences to human health. Due to the aforementioned technological difficulties, potential exogenous biomarkers for disease will be missed with current tools. Therefore, technologies for exposome data collection, identification, and quantification are needed to bring the further benefit of metabolomics to the masses.

2.2.2 Commercial Affordability and Interest

Full metabolomics screens are not currently available as a part of everyday healthcare, and therefore access by individuals to their full metabolomics profiles must come from a direct-to-consumer business. From a commercial perspective, to provide helpful metabolomics-derived information to the masses, the masses must want, and therefore pay for, their metabolomics readouts. However, a viable business model for a direct-to-consumer metabolomics product has yet to emerge. The multi-omics profiling company *Arivale* shut its doors in 2019, four years after its

founding, due to an inability to fund its consumer-based scientific wellness program which included genome, metabolome, and microbiome profiling [24]. Beyond the high technological costs, CEO Clayton Lewis also cited the high costs of customer acquisition and a lack of interest in this data as a regular part of healthcare [25]. While competitors have either lowered consumer costs or made licensing agreements with large healthcare companies that could make molecular profiling more widespread, the future availability of metabolomics for the masses is going to depend on translational utility and the willingness of the masses to independently invest in health insights derived from metabolomics data [26].

3 Future Opportunities and Challenges in Translational Metabolomics

The emerging focus of personalized medicine is greatly due to the explosion of omics data: genomics, transcriptomics, proteomics, metabolomics, etc. While bountiful research has shown the potential for biomarker detection, disease subtyping, drug repurposing and discovery, and other useful applications for patient care, the development of widespread healthcare tools has not followed suit equally for each type of omics data. Genomics has dominated clinical implementation, with more than 75,000 genetic tests available by 2017 [27]. On the contrary, 2 years later in 2019, transcriptomics and proteomics were the basis for only five assays and one assay, respectively [28], in a clinical setting. Even more astounding, metabolomics is still without an FDA-approved test in 2023 [29]. While there are no current clinical tools for metabolomics, there are many areas in healthcare research conducting extensive experiments that can benefit from their development.

Numerous branches of medicine have recognized metabolomics as a potential strategy to identify predictive, diagnostic, or prognostic markers of disease. Oncology has already made strides in this respect, as the ability to find metabolite biomarkers in serum and image-based applications has been explored. In breast cancer patients, the metabolome is representative of over 30 endogenous metabolites, characteristic of low glucose, low glycerophosphocholine, and increased tCho levels [30]. The mapping of metabolic signatures has been conducted for additional cancers including ovarian [31], lung [32], and endometrial [33]. In addition, the detection of breast cancer tissue from noninvolved adjacent tissue using metabolomics with simultaneous measurement of tumor size, lymph node status, hormone status, and histology was determined with accuracy, sensitivity, and specificity all around 90% [34]. As an increasing number of studies are performed to validate existing and discover new biomarkers of disease, the chemical fingerprint of the phenotype may become increasingly specific. In addition to diagnostic usage, metabolomics has been used to predict treatment outcomes, such as sensitivity and resistance of either chemotherapy- or hormonal therapy-treated samples of human glioma cell cultures [35]. This study demonstrates the ability to create diagnostic

tools using metabolomics that are detectable before changes in phenotype are evident using conventional imaging techniques [36].

Type 2 diabetes and other diseases under the umbrella of endocrinology have bountiful potential for metabolomic techniques to aide in our understanding of their etiology. Due to the current global type 2 diabetes epidemic and its worrying projections, a large focus has been placed on developing diagnostic biomarkers. An example of type 2 diabetes-associated metabolite, 2-aminoadipic acid, was reported by Wang et al. [37]. Using the Framingham Heart Study, 2-aminoadipic acid was discovered to be increased in diseased individuals up to 12 years before onset, and was not well correlated with other metabolites, suggesting a distinct metabolic pathway for risk assessment of phenotype outcome. As the nature vs nurture debate continues in the diabetes field, it has been demonstrated that metabolic markers are more predictive of type 2 diabetes development than genome-wide association studies or other genetic data [28]. This proposes a greater environmental role for disease onset, suggesting lifestyle changes are a viable solution. In rheumatology, 20 metabolites have been identified to decently discriminate rheumatoid arthritis from ankylosing spondylitis, Behcet's disease, and gout with an area under the receiver operating characteristic curve (AUC) of 0.812 [38]. As additional metabolomics studies targeting the classification between two groups are conducted with larger sample sizes, the validation and refinement of metabolic profiles associated with specific diseases will be attainable.

In neurological disorders, such as Alzheimer's disease (AD) and Parkinson's disease, the quest for metabolic biomarkers for early diagnosis and subtyping is of interest to many. As early AD diagnosis and AD treatment have had very limited success, metabolomics may provide novel insights into the underlying mechanisms driving AD development and progression. Metabolomics of plasma samples from AD cases compared with controls identified a higher abundance of free cholesterol in small HDL associated with a lower risk of AD and higher levels of glutamine associated with increased AD risk [39]. Additional metabolites were discovered to be correlated with general cognition. These results demonstrate potential biomarkers for further study which could be indicative of AD development and cognitive decline. If these metabolites are valid, it may be possible to create tools to aid in the diagnosis of AD and other cognition-centered conditions. In addition to biomarker detection in neurological disease, the stratification between patients with Parkinson's disease versus controls, and Parkinson's with dementia versus Parkinson's without dementia, has been demonstrated by Han et al. with AUC of 0.955 and 0.862, respectively [40]. The ability to distinguish between neurodegenerative stages provides a clinical application for diagnosing disease severity and necessary treatment.

A pediatric study focusing on the volatile organic compound (VOC) abundance for children with and without asthma has concluded an 80–100% accuracy of diagnosis with a combination of VOCs [41]. While still in its infancy, further focus on VOCs may lead to better risk assessment for identifying children with the greatest risk of adverse events. The metabolic profile of cardiovascular disease has shown promise for the discovery of new biomarkers for the diagnosis of heart-related adverse events. A significant increase in the discrimination between 150 individuals

developing atherosclerosis and 1445 not developing atherosclerosis between a 6-year time interval resulted when low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL), docosahexaenoic acid, and tyrosine were added to traditional biomarkers [42]. This conclusion demonstrates the capability of metabolites to assist in the predictive power of traditional diagnostic methods.

As the metabolome is composed of thousands of compounds, there is a high potential for the discovery of metabolites indicative of early detection for many diseases. Vallejo et al. revealed perfect separation between plasma samples from patients with atherosclerosis and controls, and patients with acute coronary syndrome and controls, using gas chromatography-mass spectrometry [43]. This suggests a metabolite-specific assay could be created to confidently identify patients with these heart problems. Heart failure patients and controls have been demonstrated to be separable based on the measurement of pseudouridine, 2-oxoglutarate, 2-hydroxy 2-methyl propanoic acid, erythritol, and 2,4,6-trihydroxy pyrimidine [44]. This research concluded that there are novel metabolic biomarkers of heart failure which can be further investigated to discover their potential to be used in prognosis.

Taken together, metabolomics has been utilized for many applications in a wide variety of medical fields. A list of metabolite biomarkers for use as diagnostics is located on the Mayo Clinic website (https://www.mayocliniclabs.com/) [28]. As the deployment of metabolomics continues to offer promising results with respect to the identification of novel predictive, diagnostic, and prognostic biomarkers that aid in the overall understanding of the biological mechanisms underlying a phenotype, new studies will result in the precise identification and refinement of a metabolic fingerprint of many diseases, which can be measured to make clinical assessments.

As interest in metabolomics increases in research, industrial efforts are aiding in the future translational capability by focusing on the creation of simpler and better LC-MS/MS systems [28]. Both Sciex and Waters have created instruments for use in clinical laboratories. In addition, enzyme-linked immunosorbent assay (ELISA) is a method by which targeted metabolites can be quantified for clinical use, although there are still limitations with this approach that need to be addressed [45]. Also, as the use of mobile device data and wearable data rapidly grows, metabolomic measurements can accompany these sources to create a foundation for the metabolome of both diseased and healthy individuals on a massive scale. The influx of this data may be used to create personalized recommendations for numerous applications including exercise and nutrition.

As mentioned earlier, diseases such as type 2 diabetes have been suggested to be driven, on average, more by environmental than by genetic components. Therefore, the culmination of multiple modalities of data for a large population could lead to the identification of necessary lifestyle requirements for the prevention of disease development. Nutritional metabolomics focuses on how chronic or acute food intake causes a response in an organism's metabolism [46]. Research involving medical foods and dietary supplements has shown promise as a solution to treat many inborn errors of metabolism, dietary deficiency diseases (such as rickets,

scurvy, and goiter), and other medical conditions such as coeliac disease (through gluten-free diets) and epilepsy (through ketogenic diets) [47].

Although cancer prevention guidelines have stated the association between the consumption of red meat, processed meat, and sugary drinks with the development of cancer, these products are still overeaten by some of the US population. Research focused on the resulting metabolic changes may aid in the discovery of direct mechanisms responsible for the correlation between these foods and drinks and cancer. As with smoking, as the evidence amounts to and becomes popularized, an incentive for the government to intervene and create a policy limiting the ability of the population to consume these harmful products may be implemented. As more studies are conducted with a focus on medicinal food, an additional application may be the incorporation of a supplemented diet in addition to standard treatment in healthcare. As nutrition influences overall well-being, the potential of precision nutrition to create a healthier population has an enormous beneficial consequence and may promise a large market as technological advances demonstrate a positive impact.

Personalized medicine, drug discovery, and minimization of risk for blood contamination are potential candidates for translational use of metabolomics. Laboratory-developed tests (LDTs) are defined by the FDA as "in vitro diagnostic tests that are manufactured by and used within a single laboratory," which can measure either individual or multiple analytes [29]. Abnormalities in metabolic pathways and biomarkers unable to be detected by other means are measured by Metabolon's Meta UDx™ test. For hereditary metabolic disorders, the diagnostic tests Meta IMD™ and Meta IMD™ + (Plus) were developed. Although not approved by the FDA, these LDTs may provide information for new tests that can be used in the clinic to gain more information about a patient than the current standard of care.

With the development of the chemical fingerprints of metabolic changes resulting from disease development, tailored recommendations can guide treatment of an individual given the metabotype of the patient [28]. There are many opportunities to expand metabolomics in the future to new sources, such as cerebrospinal fluid, human saliva, bronchoalveolar lavage, sweat, feces, semen, and amniotic fluid [48]. These studies will provide answers to current questions in multiple healthcare fields and may lead to the ability to investigate new topics. Also, by comparing metabolite abundance before and after drug treatment and studying the resulting phenotype, insight to the mechanistic impacts of drugs can be elucidated. This could also guide drug developers to create more effective therapeutics, as direct experimental evidence would provide a more comprehensive understanding of the drug mechanism [49].

Additionally, a treatment's level of toxicity may be confidently measured by metabolomics, aiding in the development of optimal medications for a patient and also providing an avenue for the creation of predictive modeling of drug toxicity for the creation of new therapeutics [50]. Bacterial problems arising from contamination and antibiotic resistance may be solved through the study of metabolism. As pathogens can be transmitted by human blood and blood-derived products, metabolomics may be a tool to minimize or eradicate this risk due to its sensitivity of measurement through clinical screening [48]. Also, the rise in antimicrobial resistance

due to antibiotic use could be alleviated by the determination of metabolomic biomarkers of resistance and the creation of methods for metabolic changes that can kill the bacteria. The implementation of these tools in healthcare could prevent the spread of resistant bacteria and save many lives in the future.

Although there are an impressive number of applications for translational metabolomics, there are numerous challenges that must be overcome to create valid, precise techniques for analyzing patient samples. Many logistical challenges provide impediments to an optimal workflow for the utilization of metabolomics in the clinic. As a relatively recent approach to omics technologies, its publicity is much less than others and often clouded by the successes more advanced omics have achieved [51]. One bottleneck is the high cost of the instruments needed to measure the samples and the laborious sample preparation methodologies [36]. The resources and armamentarium needed to store and measure metabolites before they undergo transformation and/or degradation may not be available at many healthcare institutions and must be purchased to allow for metabolomic analyses to be made possible [51]. As trends in other omics have shown decreased time and cost for the generation of data [52], it is possible that as new measurement techniques are developed, economic and temporal barriers will be less of a factor. An additional issue is a current need for the culmination of experts in different research areas including biologists, analytical chemists, statisticians, data scientists, and bioinformaticians to successfully conduct and interpret a metabolomics-based experiment in its entirety [28]. The reason for this is that the data output from traditional metabolomics platforms is rich and complex.

To be used in clinical settings, the number of metabolites must be reduced greatly for a clear interpretation of the results, thus making a risk assessment, diagnosis, and prognosis easier for the clinician. To select biomarkers for use in healthcare from large metabolomic datasets, there is a need for a user-friendly platform that can process, statistically interpret, and determine straightforward conclusions about data, demonstrating the direct effect of a change in phenotype on the metabolome. The market will also drive the availability of clinical tests. To make a product commercially viable, it must be profitable, which will depend on an estimate of how many people will use it [28]. To increase the probability of incorporating new technology in healthcare, researchers can work with industrial organizations to develop easy-to-use, clinician-approved tools.

Numerous technical aspects of metabolomics need to be overcome to create reliable metabolite biomarkers for unique metabotyping of disease. One major obstacle for untargeted metabolomics is overcoming its semiquantitative nature. As data generation relies on the normalization of a signal, the definition of the normal concentration of metabolites is needed for reliable conclusions regarding the ability of a compound to be used as a biomarker of a phenotype [48]. A showcase example is the comparison of two studies analyzing roughly 45 total Crohn's disease and ulcerative colitis patients. One concluded that there is no discrimination using metabolomics between ulcerative colitis and Crohn's disease [53], while the other concluded that choline, lipoprotein, and N-acetylated glycoprotein levels were able to separate the conditions significantly with an AUC greater than 0.9 [54].

Due to the wide range of biological variation in the metabolome, different cohorts may exhibit wide-ranging metabolic profiles. Therefore, absolute quantification of metabolite concentration is necessary for accurate benchmarking [28]. To ensure bias is not impacting the results of benchmarking experiments, validation needs to be repeated in multiple populations with large sample sizes. Currently, many biomarkers determined through metabolomics are results from studies limited in validity, statistical robustness, and experimental design [48]. Quantification coupled with repeated validation must be performed to develop an understanding of the true metabolic pathway alterations characteristic of a disease. Another major challenge to translational metabolomics is the inability to identify metabolites and the difficulty of pathway mapping [28]. There are many metabolites unavailable for measurement in commercial products and/or cannot be identified using current spectral libraries such as METLIN or mzCloud. Thus, current methods lack the complete metabolome as a whole. An extension of this is the inability to identify metabolic pathways perturbed by a disease, hindering the potential use of metabolomics for the proper determination of changing biological mechanisms, biomarker identification, and therapeutic development.

The future of translational metabolomics is contingent on the creation of standardized protocols for experimental design and measurement, simplification of data analysis and results, and the development of robust quantization methods leading to the reliable identification of metabolites. As groups such as the Metabolomics Standards Initiative (MSI) continue to meet and perfect the current procedures in metabolomic analysis, metabolomics continually progresses toward translational applications. Still, in its infancy, numerous unanswered questions in biology will be elucidated by metabolomics as it develops and its utilization increases, making it one of the most exciting technologies of the present. There is an enormous opportunity for the study of the metabolome to influence global healthcare.

4 Metabolomics as a tool to Accelerate Therapeutics and Novel Drug Discoveries

To enable capturing of the diverse array of metabolites and their dynamic cellular concentrations, the detection of these molecules is primarily based on two technologies, namely, nuclear magnetic resonance (NMR) spectroscopy (1H or 13C) and mass spectrometry (MS). Mass spectrometry is often coupled with capillary electrophoresis (CE-MS), gas chromatography (GC-MS), gas chromatography-flame ionization detection (GC-FID), direct infusion-mass spectrometry (DI-MS), or liquid chromatography (LC-MS). Due to the large chemical diversity and limited knowledge on metabolism despite the implementation of these technologies coupled with a range of analytical methods, less than 5% of the metabolome is annotated [55].

As discussed in previous chapters, there are primarily two approaches for investigating the metabolome, i.e., targeted and global approach (also referred to as

non-targeted). The former refers to the identification and measurement of the well-defined groups of chemically characterized metabolites that have been biochemically annotated as well using appropriate internal standards. Measurement is quantitative, and metabolite concentrations are expressed in molar units. This approach enables studying novel associations between the metabolites under variable physiological conditions [56]. Owing to the high sensitivity of this approach, it is often used for studying the flux metabolic pathways and for the detection of well-defined chemical compounds or validation of known biomarkers [57, 58]. Alternatively, the untargeted approach provides opportunities for comprehensive data analysis as it is possible to detect all measurable compounds or analytes within a given sample including putative annotated metabolites or chemically unknown samples. Due to this variability in detection, the data is examined in a semiquantitative or relative manner using multivariate analysis, wherein the extensive dataset is divided into smaller datasets of manageable signals. Owing to the relative quantification of the readouts as chromatographic peak areas, the data generally is expressed in terms of the intensity of ions or arbitrary units [56–59].

Moreover, these peak areas are determined by the experimental conditions under which the detection is performed using NMR, GC-MS, or liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS). Therefore, these variations make it difficult to directly compare the data from experiments done at different time points within the same laboratory or by different laboratories. The current challenges associated with the non-targeted approach include the nonavailability of standardized workflows primarily for data generation, complexity of signatures detected, lower sensitivity of detection owing to lower abundance, identification of metabolites (only a small proportion of metabolites and their annotated features are known), automated processing of data through feature detection and integration with other omics data, and finally the availability of only limited well-defined interoperability frameworks. Moreover, the current platforms for detection and analysis are highly expensive. Therefore, these together lead to nonoptimal reuse or interoperability of the data. Additionally, the identification of medically important molecular signatures and the demand for participatory medicine will also impact the establishment of methodologies for simplifying the complex data and accelerating research in the field, and will subsequently catalyze the development of affordable and accessible alternative analytical methods for the nonspecialized end users [11, 60]. Currently, the non-targeted approach is being used for the discovery of biomarkers.

There are multifaceted applications of metabolomics in various fields of biology [61]. For example, in environmental research, it is being used for addressing ecotoxicological issues [62], and in plant biology and agricultural science, it is being used for understanding cellular functions and discovering biomarkers, for diagnostics and phenotyping (*specific metabolites and quantitative trait loci (mQTL) & metabolic genome-wide association studies (mGWAS)*), and for predicting the metabolite-genome correlations [63, 64]. The current applications of metabolomics have also expanded to microbiome research [65–68]; animal health [69–71]; human healthcare including toxicology [72], epidemiology [73], cancer biology [36, 74,

75], cardiovascular diseases [36, 76], gastrointestinal diseases, [36, 77–79], aging research [80], and infectious diseases [81–83]; and nutrimetabolomics [83–85]. Toward biomedical research, metabolomics has furthered systems biology and systems medicine which has paved a pathway for personalized medicine (also referred to as precision medicine).

Personalized medicine or precision medicine is aimed at developing both disease prevention and clinical care strategies that account for variability in individuals that is affected by their environment, genetics, lifestyle, and molecular phenotype (determined by both genotype & metabolome) [60, 86]. It follows the concept of "P4 medicine," i.e., preventive, predictive, personalized, and participatory [87] in nature. The approach relies on the characterization of genetic, epigenetic, and clinical information of individuals and provides adept tailored medical treatment, which will consequently be safe and effective. As a trickle-down effect, it may enable reducing time and financial expenditure on healthcare, improve quality of life, and reduce side effects of the given treatment. Personalized medicine may have several implications including early detection of disease using medically relevant biomarkers and identification of key genetic and epigenetic parameters during the initiation and progression of the disease [60, 86, 88]. Overall, it is promising in providing deeper insights into the mechanism of disease emergence and progression and facilitates using noninvasive methods and easy-to-obtain clinical samples like body fluids (blood, sweat, urine, etc.) or volatile breath components for diagnostic purposes and stratifying disease propensity. Consequently, it is promising in laying a foundation for pharmacogenomics and targeted drug discovery, thereby enabling the measurement of well-being.

Medical decision-making is based on the examination of biochemical parameters, clinical assays, imaging scans, and rarely genetic markers. The drugs used either for treatment or for the drug discovery process are based on cellular proteins, for example, enzymes, receptors, transporters, etc. Thus, comprehensive measurement of metabolites generated through this process may provide deeper insights [89]. Moreover, these metabolic signatures have started to emerge as new biomarkers for diseases and for responding to treatment [90–109]. Furthermore, the discovery of new biomarkers can also be based on the co-metabolism of the gut microbiome along with that in humans, which have shown to modulate the levels of drugs in the blood and their effects, i.e., altering their pharmacokinetic (PK) profiles. Thus, these may be useful signatures for PK studies and PK modeling studies [110].

Toward the implementation of personalized medicine, research in pharmacometabolomics has accelerated in the last decade. It is aimed at the identification of the detailed biochemical roadmap to facilitate understanding intraindividual heterogeneity for a given disease (especially for depression and cardiovascular disorders) and their variation in response to drug treatment [98, 100, 102, 106, 111–115]. It has been also shown that urinary drug metabolite profile before treatment or at baseline can inform about the metabolism of the drugs and their toxicity [116]. One of the key initiatives has been funded by the National Institutes of Health (NIH) through the Pharmacometabolomics Research Network (PMRN) (http://pharmacometabolomics.duhs.duke.edu/) in partnership with the Pharmacogenomics Research

Network (PGRN; https://www.pgrn.org/) [89]. Studies from the same have shown that a patient's genetic and metabolic data alone or their combinations are crucial in informing the treatment outcomes as well as the underlying cause for their variation in response to treatment including the contribution of ethnicity, sex, etc. [117].

Other studies have demonstrated that metabolomics can effectively complement genomics data for assessing the risk of a given disease and for its monitoring and management to enable precision care [118]. This has led to the emergence of pharmacometabolomics-informed pharmacogenomics for addressing various diseases [100], wherein metabolic profiles are analyzed and further linked to the clinical phenotypic manifestations and with relevant genetic variants (single nucleotide polymorphisms (SNPs)) to enable the identification of novel genetic variants or SNPs that are associated with these varied drug response phenotypes [115]. Moreover, this approach seems to be useful, especially in the case of complex diseases, where similar phenotypes may arise owing to pathophysiologic processes and information from genomics data is not sufficient [115].

The upscaling of data generated in clinical pharmacology and the integration of knowledge from systems biology have led to the emergence of quantitative and systems pharmacology (QSP) [119]. This was led by the National Institute of General Medical Sciences (NIGMS) with the engagement of domain experts in pharmacology, systems biology, pharmacokinetics/pharmacodynamics, and computer modeling. QSP is enriched by data from both pharmacometabolomics and pharmacogenomics datasets [89, 115]. Owing to the low success rate of the drugs that progress from preclinical to first in human studies, the data from QSP is based on an understanding of biological pathways, disease progression, and drug mechanisms. This feeds into informing this translation that is critical for pharmaceutical R&D [119]. Thus, the combination of information from the metabolic and genetic markers can be used as unique identifiers for novel biomarker discovery.

Thus, metabolomics and the recent tools being developed are crucial in facilitating the identification of diseases through unique metabolic fingerprints or signatures. The culmination of this information with genomic data will contribute to novel biomarker discovery. Besides, the examination of clinical characteristics and their variability will enable patient stratification for informing personalized drug treatment and inform clinical trial designs including their inclusion criteria. Additionally, the clinical characterization may enable the identification of new pathways for therapeutic discovery as well as provide novel insights into mechanisms of drug actions. This will provide scope for enhancing treatment outcomes by integration of the metabolomics data with fluxomics. Advanced methodologies and the integration of omics data along with computational methods and systems biology may enable higher success rates critical in drug discovery, development, and translation. Moreover, a comparative analysis of the metabolome under baseline versus the treatment and environmental variations (including knowledge of host gut microbiome) would further provide confirmatory results for the success of personalized medicine or treatment.

5 AI-/ML-Based Approaches for Metabolomics Data Mining and Analysis

Machine learning (ML) and deep learning (DL) applications encompass everyday life, including product recommendations, spam filtering, language translation, and even customer service chatbots. The artificial intelligence revolution has gained considerable interest in healthcare, and its implementation of omics data is evident from a multitude of studies [120–122]. However, metabolomics analysis pipelines are much less developed than other omics, such as genomics and transcriptomics, which have a plethora of validated databases and tools at their disposal [123, 124]. While many challenges need to be overcome in the field of computational metabolomics, numerous applications of ML and DL are working toward solutions. In addition, research has already begun to reveal the utility of ML and DL using metabolomics, and future work has the potential to transform our understanding of health and nature.

Two similar issues facing metabolite measurements are the inability to annotate metabolites and the misidentification of similar metabolites from the raw spectral data output by the mass spectrometer. The study of the entire metabolome is ongoing, and there are many databases with metabolomic information (https://metabolomicsna.org/index.php/resources/databases). However, a majority of small compounds have not yet been added to the databases. This leads researchers to a dilemma, and the following questions are inevitable: Does one throw away metabolites unable to be annotated given current databases to increase statistical power for identifiable metabolites? Alternatively, does one include all metabolites and keep them as m/z ratios when reporting results, discussing the need for unidentifiable compounds to be found in future analyses? Regardless of the researcher's decision, valid identification of metabolites is paramount to the ability to interpret the findings and create diagnostic tools.

By employing ML and DL, researchers are devising different strategies to increase the robustness of metabolomic annotations. D.D. Matyushin, A.Y. Sholokhova, and A.K. Buryak created a deep convolutional neural network (CNN) to rank small molecules for identification using low-resolution electron ionization mass spectrometry (EI-MS) [125]. This model used the NIST 17 database to train the CNN, and the validation sets were the Golm Metabolome Database, Human Metabolome Database, and FiehnLib. The CNN outperformed other methods in ranking the metabolites. This work demonstrates the superiority of DL approaches over other methods for this specific case of metabolite identification. Multiple reviews mention convolutional neural networks developed to automate the peak-picking process [126, 127]. Kantz et al. created a CNN which removed about 90% of false positives from a conventional peak-picking pipeline [128]. By precisely identifying true metabolites in noisy mass spec data, robust biological findings can be more readily discovered. Lauren M. Petrick and Noam Shomron discuss multiple ML and DL models for peak picking, including ML models WiPP and MetaClean, and DL models Peakonly, NeatMS, NPFimg, and Eva [129]. However, some of

these models were developed to work best for specific types of peaks and therefore may not be generalizable to all analyses. Nevertheless, the DL model Eva obtained a classification accuracy for good and bad peak shapes greater than 90% when applied to 22 publicly available LC-MS metabolomics datasets. This result is very promising and shows the power of DL when large datasets have culminated. Additional models have been developed to predict the absence of a mass spectrum in a database [130, 131]. This type of analysis allows for enhanced sensitivity of metabolite annotation, as only matched metabolites are kept for further analysis. If researchers are only interested in the metabolites that are already annotated, there will be widespread use of these tools and their successors. It is possible that attributes from all these strategies using ML and DL will be combined to create a robust metabolomic analysis preprocessing pipeline.

One area of enormous potential that metabolomic analyses using ML and DL have already impacted is the safety and optimization of food. Wang et al. created a deep artificial neural network (ANN) to classify pathogenic and nonpathogenic microbes commonly found in food [132]. Although the plot of the linear dimensionality reduction PCA showed overlap between different microbes, the ANN was able to discriminate all microbe types in a model using only metabolite signals that increase during the incubation time with an accuracy of 99.2% [132]. Through the screening of microbes in food, it is possible to greatly reduce the risk of illness due to the large-scale distribution of infected food. Asakura et al. performed metabolomic profiling on eight fish species and, using an ensemble deep neural network, revealed that there were metabolites that correlated with fish size [133]. This study demonstrates the potential to engineer animals with desired traits using the metabolomic composition as biomarkers. Therefore, future applications using other animals could reshape current farming techniques and increase food supply.

Healthcare diagnoses and biomarker discovery using ML and DL with metabolomic data have the potential to revolutionize healthcare standards. A widely cited study of DL using metabolomics was conducted by Alakwaa et al. [134]. One DL and six ML models were fit to ER+ and ER- breast cancer tissue metabolites. The DL model significantly outperformed the ML models, achieving an area under the receiver operating characteristic curve of 0.93. It also identified important metabolites for separation of ER+ and ER- samples which were not identified by the ML algorithms, signifying DL's superiority to find more complex relationships within the data. The tumor microenvironment is a topic of increasing interest. Metabolomic sampling can provide a snapshot of the small compounds in contact with a patient's tumor. As databases become larger, it may be possible to identify better diagnostic markers of cancer and develop personalized therapies specific to cancer progression.

Breast cancer is one of the many phenotypes in which researchers have used ML or DL approaches with metabolomics data to classify samples by group. This type of research has already been conducted in tuberculosis [135], preterm delivery [136], colorectal cancer [137], influenza [138], renal cancer [139], acute myocardial ischemia diagnosis [140], systemic lupus erythematosus [141], NAFLD [142], Covid-19 [143], and depression [144]. Many of these analyses follow a similar

format such that metabolites from the optimal ML or DL approach which are most influential to the discrimination of the groups studied are validated by conducting literature searches for previous work correlating the metabolites with the disease. Proper use of ML and DL with sufficient sample sizes can produce powerful results because they quantify the extent to which potential sets of metabolites can be used as diagnostic markers of disease. After repeated validation using multiple datasets, diagnostic tools should be developed and deployed to better stratify risk for patients, and tailored treatments should be designed to provide patients with optimal recovery.

While the potential benefit of DL and ML using metabolomic data in the future is compelling, there are challenges that the field must overcome to advance to its greatest potential. As in all omics data analyses, there must be a comprehensive understanding of the precise question being asked and how the data will be measured to answer this question. In addition, the possibility of batch effects and confounding due to different species, sex, etc. within samples must be considered. ML and DL papers should provide and discuss numerous evaluation metrics. For example, classification problems should report area under the receiver operating characteristic curve, area under the precision-recall curve, F1 score, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, etc. to grant the audience a better ability to understand how the model correctly makes predictions and where it lacks predictive power.

One reason metabolomics has fallen behind other omics is that there are not large, standard benchmarking datasets for many of the analyses. Therefore, it has been difficult to reliably compare different DL and ML techniques for data analysis, making a standardized pipeline, such as DESeq2 for transcriptomics, nonexistent. A universal feature of DL models is the need for large amounts of data. Large studies with metabolomic data are not widely available and are not easy to use or lack a high coverage of the metabolome. While there are openly accessible data on platforms such as MetaboLights and Metabolomics Workbench, the abundance matrices are often not shared, resulting in time-consuming replication using the raw data. Without an understanding of the files necessary to process, and which tools are available for processing the raw data, it is extremely difficult to correctly reproduce the data used in the study's analysis. A recommendation would be to require the abundance matrix to be added to all studies, along with any necessary metadata to replicate the analysis that has been performed. This would likely bring a lot of attention to the metabolomics field from bioinformaticians and computer scientists who are interested in applying their knowledge to a new discipline, greatly accelerating the advancement of new techniques, technology, and the overall use of metabolomics. As an increasing attraction to the field of metabolomics continues, an already AI-influenced world will inevitably incorporate new metabolomic technologies that will contribute to a healthier, happier population.

6 Single-Cell Metabolomics

Cells are widely recognized as the most minimal, basic units of life. In biological systems, cells differentiate based on genetic expression to create heterogeneous populations which can then intercommunicate and organize from complex structures like tissues [145]. Biological function can vary greatly across cells, and therefore bulk analyses of pooled cells lose the ability to discern the differential function of distinct populations. To avoid losing this information, single-cell omics analyses have become increasingly prevalent as they have the potential to detect cellular heterogeneity within tissues [146–148]. Single-cell omics allow for the extraction and measurement of biomolecules specific to individual cells, which can then be compared to other cells to identify discrete populations that are invisible in bulk analyses [149]. While single-cell transcriptomics has seen the most rapid development of all single-cell omics technologies, there has been a push for single-cell proteomics and metabolomics data to make the functional connection between single-cell genotype and molecular phenotype [150].

6.1 Current Single-Cell Metabolomics Technologies

Probe-based mass spectrometry, also known as mass spectrometry imaging, has emerged as the most useful technique for biomolecular profiling in single-cell metabolomics. MSI can detect the levels and localization of biomolecules using a probe, such as an ion beam or laser, to perform in situ chemical desorption and/or ionization [145, 151]. By overlaying MSI probe ablation coordinates with cell images from the same sample slide, mass spectra can be assigned to cells in tissue [152]. This circumvents the need for single-cell isolation, a costly, time-consuming process on which single-cell transcriptomic methods heavily rely [153].

In using MSI for biological applications, the spatial resolution of the probe is incredibly important as cells vary greatly in scale: most eukaryotic cells are 10–20 μm, while bacterial cells are only 1–2 μm [145]. In addition, single cells can contain a large variety of metabolites at very low abundances as compared to bulk analyses, such that ion competition among molecules could lead to the detection of only the most abundant metabolites in the cell. Table 1 gives an overview of current MSI-based techniques to perform single-cell metabolomics measurements, along with their resolutions, sensitivities, and current areas for development.

Table 1 An overview of MSI-based techniques for single-cell metabolomics

Technique	Resolution (μm)	Sensitivity(fmol)	Technique description	Areas for development and application
Secondary ion mass spectrometry (SIMS)	0.05–200	$>10^{-4}$	A primary beam of positive or negative ions is focused on a sample, providing energy to ionize molecules in its focus. These "secondary ions" are then accelerated into a mass spectrometer [154]	Primary ion impact energies are high compared to bond energies within the analytes. This leads to molecular fragmentation, which complicates downstream data analysis [154] The development of high-lateral resolution SIMS (NanoSIMS) makes this technology the best for smaller organisms such as microbes, with spatial resolution as low as 30 nm [155]
Matrix-assisted laser desorption/ ionization (MALDI)	1–25	>1	A sample is covered with a chemical matrix. A laser, generally ultraviolet (infrared in IR-MALDI), is then focused on a point in the sample. The matrix absorbs energy from the laser, causing analytes to be desorbed and ionized into the gas phase, which are then measured by a mass spectrometer [156]	MALDI is the most popular technique for biological application due to its "soft" ionization technique that reduces fragmentation, leading to measurements of biomolecules with a wide range of molecular weights [157] Methods to reduce probe size for a higher resolution in MALDI include transmission geometry MALDI (TG-MALDI) and scanning microprobe MALDI (SMALDI, [145] Methods to increase sensitivity of MALDI include MALDI-2 which integrates laser post-ionization to ionize molecules in the gas phase [158]

(continued)

Table 1 (continued)

Technique	Resolution (μm)	Sensitivity(fmol)	Technique description	Areas for development and application
Laser ablation electrospray ionization (LAESI)	>30	>0.6	A mid-infrared laser is focused on a sample. The resulting ablation plume is intercepted by a highly charged aqueous spray (electrospray) to post-ionize the ablated molecules, which are then funneled into a mass spectrometer [159]	LAESI allows for sampling in ambient conditions as compared to SIMS and MALDI which occur in a vacuum The IR wavelengths used by LAESI lead to large probe diameters. To increase resolution, LAESI uses optical fiber (f-LAESI) for IR laser transmission to the sample surface [160]
Desorption electrospray ionization (DESI)	>50	>0.5	Under ambient conditions, an electrospray is aimed at a sample where it desorbs and ionizes analyte molecules on the sample surface. These now ionized analytes then travel through the air into a mass spectrometer [161]	DESI is a combination of electrospray (ESI) and desorption ionization Nanospray-DESI (NanoDESI) uses capillary action to desorb analytes which improves sensitivity and lateral resolution [162]

6.2 The Future of Single-Cell Metabolomics

The past decade has shown rapid technological advancement in the realm of MSI, addressing concerns about spatial resolution and biomolecular sensitivity, and recent developments are bringing this advancement into the coming decade. Preliminary research has already shown alternative ionization approaches for laser desorption through nanostructured surfaces and stable isotope tracking to detect metabolite incorporation into metabolic pathways at the single-cell level [145].

Beyond the development of single-cell metabolomics technologies, their application is inevitable. Already, the most popular of these MSI techniques, MALDI, has produced mass spectra on tens of thousands of cells to define subpopulations with distinct metabolic states in human hepatocytes [152]. The distinction of metabolic differences between cell populations could impact the fields of cancer research, as tumors are composed of different cell types, each with cell-type-specific metabolism [163, 164], and viral infection, as elucidating the cell-specific metabolic pathways required in viral replication could propose new therapeutic targets for antiviral mechanism [153]. Single-cell metabolomics is the next step in understanding

cellular diversity in complex biological organisms. With the development of these technologies, we will finally be able to see the genetic and phenotypic profiles of individual cells in tandem.[2]

7 Metabolic Sensors and the Future of Healthcare

The development of metabolomic-based technology is leading to a new revolution in healthcare. Recent research has identified metabolites whose abundances are indicative of changes in phenotype. For example, one group of compounds receiving focus as diagnostic markers is volatile organic compounds (VOCs). This trend has been accompanied by sensors engineered to quickly measure unique metabolites. The combination of these advances has the potential to produce personalized recommendations for nutrition, early detection of disease, food desirability, and many other applications.

VOCs are produced by a change in normal physiology and metabolic pathways in disease-affected tissues of the GI tract [166]. VOCs are measured using noninvasive techniques and could be key elements in the early detection of many diseases. Electronic nose (e-nose) instruments are tools developed that can measure many VOCs, utilizing many different sensor arrays. GI tract diseases detected using e-noses include colorectal cancer, Crohn's disease, ulcerative colitis, irritable bowel syndrome, and cholera [166]. The development of disease-specific e-nose devices has increased specificity and sensitivity. A preprint describes work by Gladding et al. in which they demonstrate VOC patterns of heart failure using a unique breath sensor that was optimized to detect acetone [167]. As acetone is an early signal of future heart failure, this technology provides a noninvasive, inexpensive diagnostic tool that can be used to assess a user's risk. Panebianco et al. conducted a study to compare the results of an untargeted GC-MS approach to GC-olfactometry (GC-O), a faster biomarker identification, on healthy and gastrointestinal cancer patients [168]. Their analyses showed that GC-O identified differentially abundant odor-active compounds that were not discovered using the GC-MS method. The targeted approach of GC-O exemplifies an increased sensitivity to compounds of interest, resulting in the identification of more biomarker candidates. These studies illustrate the potential for VOCs to be used in finding metabolites that correlate with a disease, leading to the ability to create screening and early diagnosis of several diseases.

We have already incorporated wearables, such as Fitbit, into our daily lives. These devices have sparked an interest, and sometimes an obsession, with personalized health, as consumers can continuously track some health markers. The perception of increased longevity, the ability to live a longer, healthier life, through personal tracking of biomarkers, has become the forefront with no age

[2] Metabolomics analysis has been performed on an isolated mouse-embryonic fibroblast cell by sucking a cell's contents into a nano-electrospray ionization tip and sent through a mass spectrometer to measure compounds of low molecular weight [165].

discrimination. Whereas previously the sick and elderly had access to continuous monitoring of biomarkers, all ages now have the ability, although still in a limited capacity compared to the standard of care in a medical center, to understand their current health and how it changes over time. With a focus on metabolomics, new technology is under development that will transform our understanding of personalized health through the identification of biomarkers for various purposes, leading to recommendations for fitness, nutrition, early detection of disease, and optimal treatments for a consumer.

As a pioneer of healthcare wearables, one of the major successes is the continuous glucose monitor (CGM) [169]. These wearables are typically inserted into the interstitial fluid (ISF) in the skin and repeatedly measure the consumer's glucose level at regular intervals. A CGM is traditionally worn by patients with type 1 or late-stage type 2 diabetes. Patients with diabetes must monitor their blood glucose levels because their pancreas does not produce insulin efficiently, resulting in the need for insulin intake through injection. A CGM monitors the healthy range for a patient, and some designs can send alerts when the blood glucose level is predicted to move outside the desired range. This intervention has reshaped the treatment landscape for diabetes, as patients can receive real-time readings of their glucose and, depending on the CGM provider, receive personalized information and recommendations about diet and exercise. As increasing interest has been given to CGM devices, there has now been a noninvasive CGM designed, called GlucoWatch, which uses reverse iontophoresis to obtain glucose samples on the skin [170]. The development of a noninvasive CGM lays the foundation for future wearables which measure biomarkers traditionally through blood to design new ways to record these markers.

As CGMs have provided clear evidence of the success of metabolomic sensor deployment worldwide, new wearable technologies are being developed, targeting salivary and tear fluid metabolites. Both vectors are of great interest because they also provide noninvasive means of biomarker measurements. Mannoor et al. measured bacteria in saliva using a dental tattoo [171], demonstrating the ability to noninvasively detect harmful pathogens. Kim et al. have created multiple biosensors using noninvasive mouthguards, which have successfully measured lactate [172] and uric acid [173]. Google entered the CGM space in 2014 using tear fluid metabolite measurements [174]. The demand for new wearables is evident, and there is a large potential benefit for both societal health and profits in emerging biotechnology. This research is pioneering the metabolomics-centered wearable field with success. As studies are validated, and new questions are asked and solved, an explosion of biomarker technology is likely imminent.

The wearables movement has largely been driven by private companies, as opposed to government agencies. As a benefit, the data that is collected can be used by these companies for internal research, new algorithms can be quickly developed, and more personalized recommendations for lifestyle changes can be given as a result. Conversely, consumers are, sometimes unknowingly, agreeing to share their health information with a source that could use this information to negatively affect them, for example, through increased healthcare costs if the company shares

information with insurance companies. We should be conversing about the potential benefits and drawbacks of continuous health monitoring and discussing possible regulations that should be enforced to keep consumer data secure and to help the consumer optimize health.

In addition to wearables, new devices which measure metabolites will impact a multitude of fields. To make metabolite measurement faster than current methods, Heinemann et al. created a microfluidic system that consists of a metabolite extraction chip (MEC) integrated with an automatic sampler, micropumps, and LC-MS detection [175]. Whole blood and urine samples can be analyzed in 7 min and 5 min, respectively. This innovation could severely enhance the possibility of clinical metabolomics becoming a reality, as samples could be taken, results could be received, and diagnoses could be made in the same patient visit. Measuring stool samples allows for a noninvasive method to determine a patient's current nutritional status and future needs. Auggi, a startup acquired by Seed Health, aims to create an AI algorithm that uses the collection of stool samples over time to create connections between a consumer's triggers and symptoms to suggest dietary needs [176]. This platform has the potential to increase biological knowledge about the influence of diet on overall gut health and demonstrate the use of stool as a method of biomarker development. It is possible that as more companies like Auggi are created, personalized healthcare usage of stool could become normalized as health benefits are discovered. Another area metabolomic devices have infiltrated is the criminal justice system. Abdelshafi et al. created a miniaturized device that can detect cocaine in bodily fluids using saliva and urine samples [177]. This technology may influence similar diagnostic tests to be developed for other drugs, increasing law enforcement's ability to correctly determine users under the influence. Finally, one amazing application of metabolomic devices in food is called ripeSense® [178]. This tool is the world's first sensor that changes color to indicate how ripe a fruit is, allowing consumers to choose the fruit which is most suitable for their eating schedules. Through the engineering of new technologies designed to solve numerous problems using metabolomics, an applied metabolomics revolution is likely to begin as these exciting developments become part of our daily lives.

8 Recommendations

In the foreseeable future, major scientific endeavors will be focused on personalized care. Personalized care assumes that each of us has a biomolecular variation pattern determining the disease outcome entailing personalized medical interventions. This is best highlighted in twin studies; e.g., a multi-year study comparing pairs of monozygotic and dizygotic twins found quantifiable differences in selected features of plasma proteome, which could not be explained, alone, by genetic similarity [179]. Similar studies have been conducted on metabolites [180], suggesting longitudinal and inter-individual phenotypic variability to differing degrees. It is now believed that certain biomolecular features with complex variation patterns may be

discovered which make a different person respond to treatments differently. The impact of the metabolomic approach versus the traditional biochemical approach can be appreciated by the work of Liu N. et al. wherein the use of untargeted metabolomics is employed to identify inborn errors of metabolism (IEM) [181]. Our next frontier to conquer in personalized medicine will be to identify physiological features which can reasonably predict the upcoming pathological changes to optimize individual health. This would be achieved by both preventative biomarker utilization to therapy response and monitoring biomarker discovery. Wearable devices which monitor heart rate, physical activity, sleep pattern, and other parameters are a start to this journey.

Metabolomics arguably holds more potential than other omics-based technologies, in detecting features (metabolites) mirroring the physiological state, with metabolites being a natural culmination of DNA-RNA-enzyme-metabolite dogma. However, we must be aware that metabolites change not only in disease vs healthy conditions but by age (citrate levels increase with age even in healthy controls), sex (hormonal differences along with level change with age), food habits, and population niche as well. Something seemingly simple such as increased water uptake by the subject may alter the relative concentration of crucial metabolites and can have bearing on the interpretation of the data. This requires adopting a more careful approach, which can provide us with more information mirroring the person's health.

To advance the field further, improvements need to be made at every step of the process. This involves sample collection, data acquisition, data processing, analysis, data storage, and sharing. Automation and standardization in sample collection practices must be followed. Sample type, collection method, storage conditions, and processing reagents all play a crucial role in the final output of experiments. Currently, absolute quantitation of metabolites (to achieve molar differences in key metabolites like cAMP or cGMP in healthy versus disease state) and untargeted metabolomics (to achieve complete metabolome endeavor) are two challenging aspects of metabolomics research. Absolute quantitation will open the field of biomarker discovery wherein key metabolites get perturbed in healthy versus disease state. Finding an array of key signature metabolites, which are altered, will be determining aspects of the success of metabolomics translation into clinics.

Further, to move the field forward, untargeted quantitative metabolomics is going to be of primary focus. For total metabolite detection and absolute metabolite quantification, the approach of generating synthetic metabolite standards, including isomeric metabolites, will be a key aspect. In this endeavor, MSI is going to play a central role. Untargeted metabolomics is pursued to expand the breadth and totality of metabolome profiling. This can be realized with a rigorous and exhaustive pool of reference metabolites. Future work should lean toward collaborative approaches for metabolite synthesis, thorough characterization by atomic spectroscopy and/or NMR, and then inclusion into the reference metabolites list. A consortium with worldwide access will be essential to make these reference metabolites available for analytical, and reference material used. This would accelerate the field of biomarker discovery by identifying unknown metabolites, secondary metabolites, and personalized metabolic profiles for disease prognosis and treatment. To achieve this, the

research community, sample collection centers, analytical facilities/companies, statutory organizations, private device companies, and overseeing committees (e.g., bodies like MSI) need to come together and work for a common outcome. False discovery rate (FDR) is another factor that must be minimized by increasing the sample size to appropriate numbers (minimum 5, as recommended by MSI), with least perturbation by sample collection and preparation. This study discusses at length the promises and challenges in untargeted metabolomics [182].

Second, during the scientific discovery period, data acquisition of samples should be performed by utilizing different mass spectrometry modalities, which can increase the coverage of diverse types of metabolites from hydrophobic to hydrophilic, uncharged to charged, cationic to anionic, and to different isomeric metabolites, so that an exhaustive pool is generated, making the repository more exhaustive. This will feed into the system, and more and more metabolites will be discovered and characterized. We must appreciate that no single acquisition platform can achieve a global/total breadth of the metabolome. Hence, new discoveries ranging from diverse chromatography techniques, ultrasensitive mass spectrometers, and technological advancement will take us closer to our aim of profiling near-complete metabolome.

Third, the analysis of obtained data is one of the most crucial aspects of success in the metabolomics endeavor. For this, filtering out the most common metabolites and performing longitudinal studies wherein key metabolite is absent or more abundant in condition A versus B. For this, user-friendly software, training human resource, and automated data processing are the way forward. Open-sourcing the platforms and powerful analysis software would make it robust and conclusive. Further, combining the results from three different omics approaches (transcriptomics, proteomics, and metabolomics) and integrating them to understand biologically relevant questions are crucial, and platforms like MetaCore, MetaboAnalyst, InCroMAP, and 3Omics are useful tools to analyze the metabolomic data in a standalone or integrated manner. More robust open and connected platforms with robust statistical methodologies will greatly enhance the reach of the metabolomics approach.

Further, accessibility to the masses will be a key theme to bring metabolomics closer to life. For this, small yet sensitive instrumentation and pocket-size devices with the availability of reliable and easy-to-use detection kits will be a key advance. Metabolomics has a huge role to play in public safety like airports, sports administration, and control (from detecting controlled/banned substances to measuring athlete performance markers). This can be achieved by miniaturizing the instruments and making them available at the site of use. This will greatly translate the metabolomic prowess to real-world use. Many of these current challenges, metabolomic technology updates, and opportunities for the future are discussed in good detail in different review articles for further reading [28, 61, 183].

9 Conclusion

According to Gary Patti (one of the leading scientists in the field), "Metabolomics is like a molecular-level snapshot of what's happening inside a cell or organism. It provides a unique perspective on the metabolic pathways that are active at a particular moment in time" [57]. It provides a molecular-level understanding of biology and connects it with the environment. A single human cell contains more than 42,000 metabolites [184], and most of them are still uncharacterized. The use of advanced analytical technologies, coupled with the increasing availability of large-scale datasets, has enabled researchers to identify novel pathways that are associated with a wide range of biological phenomena and diseases. With these facts in mind, we believe that metabolomics would have a spectacular impact on biology and healthcare developments.

One of the key trends in metabolomics research is precision medicine, which aims to provide personalized treatment options based on an individual's unique genetic makeup and metabolic profile. In this chapter, we have summarized the potential of metabolomics in precision medicine by identifying biomarkers that can be used to predict an individual's response to different treatments. In the same direction, personalized nutrition is another area where metabolomics can have a significant impact by identifying dietary biomarkers that can be used to develop personalized dietary recommendations. We have also reviewed the development of novel computational tools and their applications to integrate the metabolomics data with other omics (such as genomics, transcriptomics, and proteomics) and provide a more comprehensive understanding of biological systems. A comprehensive segment delineated the connection between translational biology and therapeutic progress, stemming from the comprehension of metabolic levels in diverse disease models and pathogens.

We hold the belief that interdisciplinary partnerships are vital for the advancement of metabolomics research. Collaborating with experts from various fields, such as biology, chemistry, computer science, mathematics, and engineering, can uncover fresh insights into biological systems and facilitate the identification of novel biomarkers and pathways linked to disease. Moreover, the significance of metabolomics research in public health cannot be overstated. It has the potential to revolutionize healthcare by allowing for early disease detection, more precise diagnosis, customized treatment, and nutrition options. Nonetheless, some obstacles must be overcome to fully realize the potential of metabolomics in enhancing human health. These challenges involve standardizing sample collection, data acquisition, data processing, data analysis, data storage, and sharing. To sum up, the prospects for metabolomics research are bright and full of promise. With the continuous progress of advanced analytical technologies, along with the growing availability of large-scale datasets and the use of artificial intelligence and machine learning, researchers and clinicians will be able to make faster decisions.

References

1. Oliver, S.G., et al., *Systematic functional analysis of the yeast genome.* Trends Biotechnol, 1998. **16**(9): p. 373–8.
2. Junot, C., et al., *High resolution mass spectrometry based techniques at the crossroads of metabolic pathways.* Mass Spectrom Rev, 2014. **33**(6): p. 471–500.
3. Shin, S.Y., et al., *An atlas of genetic influences on human blood metabolites.* Nat Genet, 2014. **46**(6): p. 543–550.
4. Suhre, K., et al., *Human metabolic individuality in biomedical and pharmaceutical research.* Nature, 2011. **477**(7362): p. 54–60.
5. Foster, M., et al., *Uncovering PFAS and Other Xenobiotics in the Dark Metabolome Using Ion Mobility Spectrometry, Mass Defect Analysis, and Machine Learning.* Environ Sci Technol, 2022. **56**(12): p. 9133–9143.
6. Zhu, J., et al., *Detecting bacterial lung infections: in vivo evaluation of in vitro volatile fingerprints.* J Breath Res, 2013. **7**(1): p. 016003.
7. Exline, M.C., et al., *Exhaled nitric oxide detection for diagnosis of COVID-19 in critically ill patients.* PLoS One, 2021. **16**(10): p. e0257644.
8. McKinney, J., *Coronavirus (COVID-19) Update: FDA Authorizes First COVID-19 Diagnostic Test Using Breath Samples*, in *Test provides results in less than three minutes*. 2022, FDA.gov.
9. Pietzner, M., et al., *Plasma metabolites to profile pathways in noncommunicable disease multimorbidity.* Nat Med, 2021. **27**(3): p. 471–479.
10. Yu, Z., et al., *Differences between human plasma and serum metabolite profiles.* PLoS One, 2011. **6**(7): p. e21230.
11. Trivedi, D.K., K.A. Hollywood, and R. Goodacre, *Metabolomics for the masses: The future of metabolomics in a personalized world.* New Horiz Transl Med, 2017. **3**(6): p. 294–305.
12. Do, K.T., et al., *Network-based approach for analyzing intra- and interfluid metabolite associations in human blood, urine, and saliva.* J Proteome Res, 2015. **14**(2): p. 1183–94.
13. Wishart, D.S., *Metabolomics for Investigating Physiological and Pathophysiological Processes.* Physiol Rev, 2019. **99**(4): p. 1819–1875.
14. Hollywood, K., D.R. Brison, and R. Goodacre, *Metabolomics: current technologies and future trends.* Proteomics, 2006. **6**(17): p. 4716–23.
15. Armitage, E.G. and C. Barbas, *Metabolomics in cancer biomarker discovery: current trends and future perspectives.* J Pharm Biomed Anal, 2014. **87**: p. 1–11.
16. Kell, D.B., *Metabolomic biomarkers: search, discovery and validation.* Expert Review of Molecular Diagnostics, 2014. **7**(4): p. 329–333.
17. Xia, J., et al., *Translational biomarker discovery in clinical metabolomics: an introductory tutorial.* Metabolomics, 2013. **9**(2): p. 280–299.
18. Cui, Y., et al., *The Exposome: Embracing the Complexity for Discovery in Environmental Health.* Environ Health Perspect, 2016. **124**(8): p. A137–40.
19. Rappaport, S.M., *Genetic Factors Are Not the Major Causes of Chronic Diseases.* PLoS One, 2016. **11**(4): p. e0154387.
20. Andra, S.S., et al., *Trends in the application of high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the environmental chemical space of the human exposome.* Environ Int, 2017. **100**: p. 32–61.
21. Gao, P., et al., *Precision environmental health monitoring by longitudinal exposome and multi-omics profiling.* Genome Res, 2022. **32**(6): p. 1199–1214.
22. Jiang, C., et al., *Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring.* Cell, 2018. **175**(1): p. 277–291 e31.
23. Jiang, C., et al., *Decoding personal biotic and abiotic airborne exposome.* Nat Protoc, 2021. **16**(2): p. 1129–1151.
24. Arivale. 2019; Available from: http://www.arivale.com/.

25. Bishop, T. and J. Thorne. *Why Arivale failed: Inside the surprise closure of an ambitious 'scientific wellness' startup*. 2019 [cited 2022; Available from: https://www.geekwire.com/2019/arivale-shut-doors-inside-surprise-closure-ambitious-scientific-wellness-startup/.
26. Paquette, D. *Molecular You Signs Agreement for Roche to Use its Molecular Profiling Technology*. 2021 [cited 2022; Available from: https://www.biospace.com/article/molecular-you-signs-agreement-for-roche-to-use-its-molecular-profiling-technology/.
27. Phillips, K.A., et al., *Genetic Test Availability And Spending: Where Are We Now? Where Are We Going?* Health Aff (Millwood), 2018. **37**(5): p. 710–716.
28. Pinu, F.R., S.A. Goldansaz, and J. Jaine, *Translational Metabolomics: Current Challenges and Future Opportunities.* Metabolites, 2019. **9**(6).
29. Trifonova, O.P., et al., *Current State and Future Perspectives on Personalized Metabolomics.* Metabolites, 2023. **13**(1).
30. Spratlin, J.L., N.J. Serkova, and S.G. Eckhardt, *Clinical applications of metabolomics in oncology: a review.* Clin Cancer Res, 2009. **15**(2): p. 431–40.
31. Turkoglu, O., et al., *Metabolomics of biomarker discovery in ovarian cancer: a systematic review of the current literature.* Metabolomics, 2016. **12**(4).
32. Yu, L., K. Li, and X. Zhang, *Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review.* Oncotarget, 2017. **8**(70): p. 115774–115786.
33. Troisi, J., et al., *Metabolomic Signature of Endometrial Cancer.* J Proteome Res, 2018. **17**(2): p. 804–812.
34. Bathen, T.F., et al., *Feasibility of MR metabolomics for immediate analysis of resection margins during breast cancer surgery.* PLoS One, 2013. **8**(4): p. e61578.
35. Glunde, K., et al., *Choline metabolism-based molecular diagnosis of cancer: an update.* Expert Rev Mol Diagn, 2015. **15**(6): p. 735–47.
36. Ashrafian, H., et al., *Metabolomics: The Stethoscope for the Twenty-First Century.* Med Princ Pract, 2021. **30**(4): p. 301–310.
37. Wang, T.J., et al., *2-Aminoadipic acid is a biomarker for diabetes risk.* J Clin Invest, 2013. **123**(10): p. 4309–17.
38. Kim, S., et al., *Global metabolite profiling of synovial fluid for the specific diagnosis of rheumatoid arthritis from other inflammatory arthritis.* PLoS One, 2014. **9**(6): p. e97501.
39. van der Lee, S.J., et al., *Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies.* Alzheimers Dement, 2018. **14**(6): p. 707–722.
40. Han, W., et al., *Profiling novel metabolic biomarkers for Parkinson's disease using in-depth metabolomic analysis.* Mov Disord, 2017. **32**(12): p. 1720–1728.
41. Licari, A., et al., *Asthma Endotyping and Biomarkers in Childhood Asthma.* Pediatr Allergy Immunol Pulmonol, 2018. **31**(2): p. 44–55.
42. Wurtz, P., et al., *High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis.* Eur Heart J, 2012. **33**(18): p. 2307–16.
43. Vallejo, M., et al., *Plasma fingerprinting with GC-MS in acute coronary syndrome.* Anal Bioanal Chem, 2009. **394**(6): p. 1517–24.
44. Dunn, W.B., Broadhurst, D.I., Deepak, S.M. et al., *Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate.* Metabolomics 2007. **3**: p. 413–426.
45. Zhang, S., et al., *Predicting detection limits of enzyme-linked immunosorbent assay (ELISA) and bioanalytical techniques in general.* Analyst, 2014. **139**(2): p. 439–45.
46. Gibney, M.J., et al., *Metabolomics in human nutrition: opportunities and challenges.* Am J Clin Nutr, 2005. **82**(3): p. 497–503.
47. Wishart, D.S., *Emerging applications of metabolomics in drug discovery and precision medicine.* Nat Rev Drug Discov, 2016. **15**(7): p. 473–84.
48. Di Minno, A., et al., *Challenges in Metabolomics-Based Tests, Biomarkers Revealed by Metabolomic Analysis, and the Promise of the Application of Metabolomics in Precision Medicine.* Int J Mol Sci, 2022. **23**(9).

49. Mastrangelo, A., et al., *Metabolomics as a tool for drug discovery and personalised medicine. A review.* Curr Top Med Chem, 2014. **14**(23): p. 2627–36.
50. Nicholson, J.K., et al., *Metabonomics: a platform for studying drug toxicity and gene function.* Nat Rev Drug Discov, 2002. **1**(2): p. 153–61.
51. Sengupta, A., A. Uppoor, and M.B. Joshi, *Metabolomics: Paving the path for personalized periodontics – A literature review.* J Indian Soc Periodontol, 2022. **26**(2): p. 98–103.
52. D'Adamo, G.L., J.T. Widdop, and E.M. Giles, *The future is now? Clinical and translational aspects of "Omics" technologies.* Immunol Cell Biol, 2021. **99**(2): p. 168–176.
53. Dawiskiba, T., et al., *Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases.* World J Gastroenterol, 2014. **20**(1): p. 163–74.
54. Williams, H.R., et al., *Serum metabolic profiling in inflammatory bowel disease.* Dig Dis Sci, 2012. **57**(8): p. 2157–65.
55. da Silva, R.R., P.C. Dorrestein, and R.A. Quinn, *Illuminating the dark matter in metabolomics.* Proc Natl Acad Sci U S A, 2015. **112**(41): p. 12549–50.
56. Roberts, L.D., et al., *Targeted metabolomics.* Curr Protoc Mol Biol, 2012. **Chapter 30**: p. Unit 30 2 1–24.
57. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy.* Nat Rev Mol Cell Biol, 2012. **13**(4): p. 263–9.
58. Dunn, W.B., et al., *Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy.* Chem Soc Rev, 2011. **40**(1): p. 387–426.
59. Zhang, X., et al., *Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis.* Oncotarget, 2016. **7**(39): p. 63437–63448.
60. Castelli, F.A., et al., *Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests.* Anal Bioanal Chem, 2022. **414**(2): p. 759–789.
61. Yang, Q., et al., *Metabolomics biotechnology, applications, and future trends: a systematic review.* RSC Adv, 2019. **9**(64): p. 37245–37257.
62. Dumas, T., et al., *Environmental Metabolomics Promises and Achievements in the Field of Aquatic Ecotoxicology: Viewed through the Pharmaceutical Lens.* Metabolites, 2022. **12**(2).
63. Wu, W., et al., *Emerging applications of metabolomics in food science and future trends.* Food Chem X, 2022. **16**: p. 100500.
64. Sakurai, N., *Recent applications of metabolomics in plant breeding.* Breed Sci, 2022. **72**(1): p. 56–65.
65. Tang, Z.Z., et al., *Multi-Omic Analysis of the Microbiome and Metabolome in Healthy Subjects Reveals Microbiome-Dependent Relationships Between Diet and Metabolites.* Front Genet, 2019. **10**: p. 454.
66. Nguyen, Q.P., et al., *Associations between the gut microbiome and metabolome in early life.* BMC Microbiol, 2021. **21**(1): p. 238.
67. Bauermeister, A., et al., *Mass spectrometry-based metabolomics in microbiome investigations.* Nat Rev Microbiol, 2022. **20**(3): p. 143–160.
68. Lee-Sarwar, K.A., et al., *Metabolome-Microbiome Crosstalk and Human Disease.* Metabolites, 2020. **10**(5).
69. Tran, H., M. McConville, and P. Loukopoulos, *Metabolomics in the study of spontaneous animal diseases.* J Vet Diagn Invest, 2020. **32**(5): p. 635–647.
70. Goldansaz, S.A., et al., *Livestock metabolomics and the livestock metabolome: A systematic review.* PLoS One, 2017. **12**(5): p. e0177675.
71. Zhang, G., et al., *A Multi-Platform Metabolomics Approach Identifies Urinary Metabolite Signatures That Differentiate Ketotic From Healthy Dairy Cows.* Front Vet Sci, 2021. **8**: p. 595983.
72. Ramirez, T., et al., *Metabolomics in toxicology and preclinical research.* ALTEX, 2013. **30**(2): p. 209–25.

73. Playdon, M.C., et al., *Metabolomics Analytics Workflow for Epidemiological Research: Perspectives from the Consortium of Metabolomics Studies (COMETS).* Metabolites, 2019. **9**(7).
74. Dougan, M.M., et al., *Metabolomic profiles in breast cancer: a pilot case-control study in the breast cancer family registry.* BMC Cancer, 2018. **18**(1): p. 532.
75. De Preter, V. and K. Verbeke, *Metabolomics as a diagnostic tool in gastroenterology.* World J Gastrointest Pharmacol Ther, 2013. **4**(4): p. 97–107.
76. McGarrah, R.W., et al., *Cardiovascular Metabolomics.* Circ Res, 2018. **122**(9): p. 1238–1258.
77. Han, L., et al., *Altered metabolome and microbiome features provide clues in understanding irritable bowel syndrome and depression comorbidity.* ISME J, 2022. **16**(4): p. 983–996.
78. Vernocchi, P., F. Del Chierico, and L. Putignani, *Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health.* Front Microbiol, 2016. **7**: p. 1144.
79. Zhao, Z., et al., *Application of metabolomics in osteoporosis research.* Front Endocrinol (Lausanne), 2022. **13**: p. 993253.
80. Adav, S.S. and Y. Wang, *Metabolomics Signatures of Aging: Recent Advances.* Aging Dis, 2021. **12**(2): p. 646–661.
81. Wendt, C.H., et al., *Metabolite profiles associated with disease progression in influenza infection.* PLoS One, 2021. **16**(4): p. e0247493.
82. McCreath, G., et al., *A Metabolomics approach for the diagnosis Of SecondAry InfeCtions in COVID-19 (MOSAIC): a study protocol.* BMC Infect Dis, 2021. **21**(1): p. 1204.
83. Ulaszewska, M.M., et al., *Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies.* Mol Nutr Food Res, 2019. **63**(1): p. e1800384.
84. Reisdorph, N.A., et al., *Nutrimetabolomics reveals food-specific compounds in urine of adults consuming a DASH-style diet.* Sci Rep, 2020. **10**(1): p. 1157.
85. Rangel-Huerta, O.D. and A. Gil, *Nutrimetabolomics: An Update on Analytical Approaches to Investigate the Role of Plant-Based Foods and Their Bioactive Compounds in Non-Communicable Chronic Diseases.* Int J Mol Sci, 2016. **17**(12).
86. Mathur, S. and J. Sutton, *Personalized medicine could transform healthcare.* Biomed Rep, 2017. **7**(1): p. 3–5.
87. Hood, L. and M. Flores, *A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory.* N Biotechnol, 2012. **29**(6): p. 613–24.
88. Ginsburg, G.S. and K.A. Phillips, *Precision Medicine: From Science To Value.* Health Aff (Millwood), 2018. **37**(5): p. 694–701.
89. Kaddurah-Daouk, R., R. Weinshilboum, and N. Pharmacometabolomics Research, *Metabolomic Signatures for Drug Response Phenotypes: Pharmacometabolomics Enables Precision Medicine.* Clin Pharmacol Ther, 2015. **98**(1): p. 71–5.
90. Hesse, J., et al., *Profound inhibition of CD73-dependent formation of anti-inflammatory adenosine in B cells of SLE patients.* EBioMedicine, 2021. **73**: p. 103616.
91. Lin, X., et al., *Metabolic biomarker signature for predicting the effect of neoadjuvant chemotherapy of breast cancer.* Ann Transl Med, 2019. **7**(22): p. 670.
92. Ye, Z., et al., *A 13-Gene Metabolic Prognostic Signature Is Associated With Clinical and Immune Features in Stomach Adenocarcinoma.* Front Oncol, 2021. **11**: p. 612952.
93. Yousf, S., et al., *Metabolic signatures suggest o-phosphocholine to UDP-N-acetylglucosamine ratio as a potential biomarker for high-glucose and/or palmitate exposure in pancreatic beta-cells.* Metabolomics, 2019. **15**(4): p. 55.
94. Kaddurah-Daouk, R., et al., *Pretreatment metabotype as a predictor of response to sertraline or placebo in depressed outpatients: a proof of concept.* Transl Psychiatry, 2011. **1**(7): p. e26–.
95. Zhu, H., et al., *Pharmacometabolomics of response to sertraline and to placebo in major depressive disorder – possible role for methoxyindole pathway.* PLoS One, 2013. **8**(7): p. e68283.

96. Kaddurah-Daouk, R., et al., *Pharmacometabolomic mapping of early biochemical changes induced by sertraline and placebo.* Transl Psychiatry, 2013. **3**(1): p. e223.
97. Yao, J.K., et al., *Associations between purine metabolites and clinical symptoms in schizophrenia.* PLoS One, 2012. **7**(8): p. e42165.
98. Yerges-Armstrong, L.M., et al., *Purine pathway implicated in mechanism of resistance to aspirin therapy: pharmacometabolomics-informed pharmacogenomics.* Clin Pharmacol Ther, 2013. **94**(4): p. 525–32.
99. Abo, R., et al., *Merging pharmacometabolomics with pharmacogenomics using '1000 Genomes' single-nucleotide polymorphism imputation: selective serotonin reuptake inhibitor response pharmacogenomics.* Pharmacogenet Genomics, 2012. **22**(4): p. 247–53.
100. Ji, Y., et al., *Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics.* Clin Pharmacol Ther, 2011. **89**(1): p. 97–104.
101. Perroud, B., et al., *Pharmacometabolomic signature of ataxia SCA1 mouse model and lithium effects.* PLoS One, 2013. **8**(8): p. e70610.
102. Lewis, J.P., et al., *Integration of pharmacometabolomic and pharmacogenomic approaches reveals novel insights into antiplatelet therapy.* Clin Pharmacol Ther, 2013. **94**(5): p. 570–3.
103. Trupp, M., et al., *Metabolomics reveals amino acids contribute to variation in response to simvastatin treatment.* PLoS One, 2012. **7**(7): p. e38386.
104. Kaddurah-Daouk, R., et al., *Metabolomic mapping of atypical antipsychotic effects in schizophrenia.* Mol Psychiatry, 2007. **12**(10): p. 934–45.
105. McEvoy, J., et al., *Lipidomics reveals early metabolic changes in subjects with schizophrenia: effects of atypical antipsychotics.* PLoS One, 2013. **8**(7): p. e68717.
106. Ellero-Simatos, S., et al., *Pharmacometabolomics reveals that serotonin is implicated in aspirin response variability.* CPT Pharmacometrics Syst Pharmacol, 2014. **3**(7): p. e125.
107. Cooper-Dehoff, R.M., et al., *Is diabetes mellitus-linked amino acid signature associated with beta-blocker-induced impaired fasting glucose?* Circ Cardiovasc Genet, 2014. **7**(2): p. 199–205.
108. Wikoff, W.R., et al., *Pharmacometabolomics reveals racial differences in response to atenolol treatment.* PLoS One, 2013. **8**(3): p. e57639.
109. Kaddurah-Daouk, R., et al., *Lipidomic analysis of variation in response to simvastatin in the Cholesterol and Pharmacogenetics Study.* Metabolomics, 2010. **6**(2): p. 191–201.
110. Li, H. and W. Jia, *Cometabolism of microbes and host: implications for drug metabolism and drug-induced toxicity.* Clin Pharmacol Ther, 2013. **94**(5): p. 574–81.
111. Ji, Y., et al., *Pharmacogenomics of selective serotonin reuptake inhibitor treatment for major depressive disorder: genome-wide associations and functional genomics.* Pharmacogenomics J, 2013. **13**(5): p. 456–63.
112. Ji, Y., et al., *Citalopram and escitalopram plasma drug and metabolite concentrations: genome-wide associations.* Br J Clin Pharmacol, 2014. **78**(2): p. 373–83.
113. Gupta, M., et al., *TSPAN5, ERICH3 and selective serotonin reuptake inhibitors in major depressive disorder: pharmacometabolomics-informed pharmacogenomics.* Mol Psychiatry, 2016. **21**(12): p. 1717–1725.
114. Mitchell, B.D., et al., *The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study.* Am Heart J, 2008. **155**(5): p. 823–8.
115. Neavin, D., R. Kaddurah-Daouk, and R. Weinshilboum, *Pharmacometabolomics informs Pharmacogenomics.* Metabolomics, 2016. **12**(7).
116. Clayton, T.A., et al., *Pharmaco-metabonomic phenotyping and personalized drug treatment.* Nature, 2006. **440**(7087): p. 1073–7.
117. Beger, R.D., et al., *Metabolomics enables precision medicine: "A White Paper, Community Perspective".* Metabolomics, 2016. **12**(10): p. 149.
118. Guo, L., et al., *Plasma metabolomic profiles enhance precision medicine for volunteers of normal health.* Proc Natl Acad Sci U S A, 2015. **112**(35): p. E4901–10.

119. Leil, T.A. and R. Bertz, *Quantitative Systems Pharmacology can reduce attrition and improve productivity in pharmaceutical research and development.* Front Pharmacol, 2014. **5**: p. 247.
120. Martorell-Marugan, J., et al., *Deep Learning in Omics Data Analysis and Precision Medicine*, in *Computational Biology*, H. Husi, Editor. 2019: Brisbane (AU).
121. Sammut, S.J., et al., *Multi-omic machine learning predictor of breast cancer therapy response.* Nature, 2022. **601**(7894): p. 623–629.
122. Reel, P.S., et al., *Using machine learning approaches for multi-omics data analysis: A review.* Biotechnol Adv, 2021. **49**: p. 107739.
123. Tercan, B., & Leblebici, A, *Data Resources and Machine Learning for Transcriptomics Data Analysis.* Current Studies in Basic Sciences, Engineering and Technology, 2021: p. 70–85.
124. Félix Raimundo, L.M.-P., Céline Vallot, Jean-Philippe Vert, *Machine learning for single-cell genomics data analysis.* 2021. **26**: p. 64–71.
125. Pomyen, Y., et al., *Deep metabolome: Applications of deep learning in metabolomics.* Comput Struct Biotechnol J, 2020. **18**: p. 2818–2825.
126. Matyushin, D.D., A.Y. Sholokhova, and A.K. Buryak, *Deep Learning Driven GC-MS Library Search and Its Application for Metabolomics.* Anal Chem, 2020. **92**(17): p. 11818–11825.
127. Liebal, U.W., et al., *Machine Learning Applications for Mass Spectrometry-Based Metabolomics.* Metabolites, 2020. **10**(6).
128. Kantz, E.D., et al., *Deep Neural Networks for Classification of LC-MS Spectral Peaks.* Anal Chem, 2019. **91**(19): p. 12407–12413.
129. Petrick, L.M. and N. Shomron, *AI/ML-driven advances in untargeted metabolomics and exposomics for biomedical applications.* Cell Rep Phys Sci, 2022. **3**(7).
130. Sotnezova, K.M., Samokhin, A.S. & Revelsky, I.A., *Use of PLS Discriminant Analysis for Revealing the Absence of a Compound in an Electron Ionization Mass Spectral Database.* Anal Chem, 2017.
131. Samokhin, A., K. Sotnezova, and I. Revelsky, *Predicting the absence of an unknown compound in a mass spectral database.* Eur J Mass Spectrom (Chichester), 2019. **25**(6): p. 439–444.
132. Wang, D., P. Greenwood, and M.S. Klein, *Deep Learning for Rapid Identification of Microbes Using Metabolomics Profiles.* Metabolites, 2021. **11**(12).
133. Asakura, T., Y. Date, and J. Kikuchi, *Application of ensemble deep neural network to metabolomics studies.* Anal Chim Acta, 2018. **1037**: p. 230–236.
134. Alakwaa, F.M., K. Chaudhary, and L.X. Garmire, *Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data.* J Proteome Res, 2018. **17**(1): p. 337–347.
135. Hu, X., et al., *Combining metabolome and clinical indicators with machine learning provides some promising diagnostic markers to precisely detect smear-positive/negative pulmonary tuberculosis.* BMC Infect Dis, 2022. **22**(1): p. 707.
136. Bahado-Singh, R.O., et al., *Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix.* Ultrasound Obstet Gynecol, 2019. **54**(1): p. 110–118.
137. Troisi, J., et al., *A Metabolomics-Based Screening Proposal for Colorectal Cancer.* Metabolites, 2022. **12**(2).
138. Hogan, C.A., et al., *Nasopharyngeal metabolomics and machine learning approach for the diagnosis of influenza.* EBioMedicine, 2021. **71**: p. 103546.
139. Prade, V.M., et al., *The synergism of spatial metabolomics and morphometry improves machine learning-based renal tumour subtype classification.* Clin Transl Med, 2022. **12**(2): p. e666.
140. Cao, J., et al., *Combined metabolomics and machine learning algorithms to explore metabolic biomarkers for diagnosis of acute myocardial ischemia.* Int J Legal Med, 2022.
141. Niu, J., et al., *Joint deep learning for batch effect removal and classification toward MALDI MS based metabolomics.* BMC Bioinformatics, 2022. **23**(1): p. 270.

142. Oh, T.G., et al., *A Universal Gut-Microbiome-Derived Signature Predicts Cirrhosis.* Cell Metab, 2020. **32**(5): p. 878–888 e6.
143. Shen, B.e.a., *Proteomic and Metabolomic Characterization of COVID-19 Patient Sera.* Cell, 2020. **182**: p. 59–72.
144. Takahashi, Y., et al., *Improved metabolomic data-based prediction of depressive symptoms using nonlinear machine learning with feature selection.* Transl Psychiatry, 2020. **10**(1): p. 157.
145. Taylor, M.J., J.K. Lukowski, and C.R. Anderton, *Spatially Resolved Mass Spectrometry at the Single Cell: Recent Innovations in Proteomics and Metabolomics.* J Am Soc Mass Spectrom, 2021. **32**(4): p. 872–894.
146. Pelkmans, L., *Cell Biology. Using cell-to-cell variability--a new era in molecular biology.* Science, 2012. **336**(6080): p. 425–6.
147. Lee, M.C., et al., *Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing.* Proc Natl Acad Sci U S A, 2014. **111**(44): p. E4726–35.
148. Russell, A.B., C. Trapnell, and J.D. Bloom, *Extreme heterogeneity of influenza virus infection in single cells.* Elife, 2018. **7**.
149. Altschuler, S.J. and L.F. Wu, *Cellular heterogeneity: do differences make a difference?* Cell, 2010. **141**(4): p. 559–63.
150. Lanekoff, I., V.V. Sharma, and C. Marques, *Single-cell metabolomics: where are we and where are we going?* Curr Opin Biotechnol, 2022. **75**: p. 102693.
151. Hansen, R.L. and Y.J. Lee, *High-Spatial Resolution Mass Spectrometry Imaging: Toward Single Cell Metabolomics in Plant Tissues.* The Chemical Record, 2018. **18**(1): p. 65–77.
152. Rappez, L., et al., *SpaceM reveals metabolic states of single cells.* Nat Methods, 2021. **18**(7): p. 799–805.
153. Kumar, R., et al., *Single Cell Metabolomics: A Future Tool to Unmask Cellular Heterogeneity and Virus-Host Interaction in Context of Emerging Viral Diseases.* Front Microbiol, 2020. **11**: p. 1152.
154. Mas, S., et al., *Cluster TOF-SIMS imaging: a new light for in situ metabolomics?* Proteomics, 2008. **8**(18): p. 3735–45.
155. Behrens, S., A. Kappler, and M. Obst, *Linking environmental processes to the in situ functioning of microorganisms by high-resolution secondary ion mass spectrometry (NanoSIMS) and scanning transmission X-ray microscopy (STXM).* Environ Microbiol, 2012. **14**(11): p. 2851–69.
156. Aichler, M. and A. Walch, *MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice.* Lab Invest, 2015. **95**(4): p. 422–31.
157. Norris, J.L. and R.M. Caprioli, *Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research.* Chem Rev, 2013. **113**(4): p. 2309–42.
158. Soltwisch, J., et al., *Mass spectrometry imaging with laser-induced postionization.* Science, 2015. **348**(6231): p. 211–5.
159. Shrestha, B. and A. Vertes, *In situ metabolic profiling of single cells by laser ablation electrospray ionization mass spectrometry.* Anal Chem, 2009. **81**(20): p. 8265-71.
160. Taylor, M.J., et al., *Ambient Single-Cell Analysis and Native Tissue Imaging Using Laser-Ablation Electrospray Ionization Mass Spectrometry with Increased Spatial Resolution.* J Am Soc Mass Spectrom, 2021. **32**(9): p. 2490–2494.
161. Arentz, G., et al., *Applications of Mass Spectrometry Imaging to Cancer.* Adv Cancer Res, 2017. **134**: p. 27–66.
162. Roach, P.J., J. Laskin, and A. Laskin, *Nanospray desorption electrospray ionization: an ambient method for liquid-extraction surface sampling in mass spectrometry.* Analyst, 2010. **135**(9): p. 2233–6.
163. Kim, J. and R.J. DeBerardinis, *Mechanisms and Implications of Metabolic Heterogeneity in Cancer.* Cell Metab, 2019. **30**(3): p. 434–446.

164. Lau, A.N., et al., *Dissecting cell-type-specific metabolism in pancreatic ductal adenocarcinoma.* Elife, 2020. **9**.
165. Tsuyama, N., et al., *Live single-cell molecular analysis by video-mass spectrometry.* Anal Sci, 2008. **24**(5): p. 559–61.
166. Wilson, A.D., *Application of Electronic-Nose Technologies and VOC-Biomarkers for the Noninvasive Early Diagnosis of Gastrointestinal Diseases (dagger).* Sensors (Basel), 2018. **18**(8).
167. Patrick A. Gladding, R.Y., Maxine Cooper, Suzanne Loader, Kevin Smith, Erica Zarate, Saras Green, Silas G. Villas-Boas, Phillip Shepherd, Purvi Kakadiya, Eric Thorstensen, Christine Keven, Margaret Coe, Mia Jüllig, Edmond Zhang, Todd T. Schlegel, *Metabolomics and a Breath Sensor Identify Acetone as a Biomarker for Heart Failure.* 2021.
168. Panebianco, C., et al., *Cancer sniffer dogs: how can we translate this peculiarity in laboratory medicine? Results of a pilot study on gastrointestinal cancers.* Clin Chem Lab Med, 2017. **56**(1): p. 138–146.
169. Gao, W., G.A. Brooks, and D.C. Klonoff, *Wearable physiological systems and technologies for metabolic monitoring.* J Appl Physiol (1985), 2018. **124**(3): p. 548–556.
170. Potts, R.O., J.A. Tamada, and M.J. Tierney, *Glucose monitoring by reverse iontophoresis.* Diabetes Metab Res Rev, 2002. **18 Suppl 1**: p. S49–53.
171. Mannoor, M.S., et al., *Graphene-based wireless bacteria detection on tooth enamel.* Nat Commun, 2012. **3**: p. 763.
172. Kim, J., et al., *Non-invasive mouthguard biosensor for continuous salivary monitoring of metabolites.* Analyst, 2014. **139**(7): p. 1632–6.
173. Kim, J., et al., *Wearable salivary uric acid mouthguard biosensor with integrated wireless electronics.* Biosens Bioelectron, 2015. **74**: p. 1061–8.
174. DiaTribe. *Google Secures Patent for Glucose-Sensing Contact Lens*. 2015; Available from: https://diatribe.org/google-secures-patent-glucose-sensing-contact-lens.
175. Heinemann, J., et al., *Analysis of Raw Biofluids by Mass Spectrometry Using Microfluidic Diffusion-Based Separation.* Anal Methods, 2017. **9**(3): p. 385–392.
176. 2021; Available from: https://www.auggi.ai/.
177. Lapizco-Encinas, B.H. and Y.V. Zhang, *Microfluidic systems in clinical diagnosis.* Electrophoresis, 2022.
178. Jacobs, D.M., M.A. van den Berg, and R.D. Hall, *Towards superior plant-based foods using metabolomics.* Curr Opin Biotechnol, 2021. **70**: p. 23–28.
179. Liu, Y., et al., *Quantitative variability of 342 plasma proteins in a human twin population.* Mol Syst Biol, 2015. **11**(1): p. 786.
180. Bermingham, K.M., et al., *Genetic and Environmental Contributions to Variation in the Stable Urinary NMR Metabolome over Time: A Classic Twin Study.* J Proteome Res, 2021. **20**(8): p. 3992–4000.
181. Liu, N., et al., *Comparison of Untargeted Metabolomic Profiling vs Traditional Metabolic Screening to Identify Inborn Errors of Metabolism.* JAMA Netw Open, 2021. **4**(7): p. e2114155.
182. Gertsman, I. and B.A. Barshop, *Promises and pitfalls of untargeted metabolomics.* J Inherit Metab Dis, 2018. **41**(3): p. 355–366.
183. Barron, R., et al., *Twin metabolomics: the key to unlocking complex phenotypes in nutrition research.* Nutr Res, 2016. **36**(4): p. 291–304.
184. Fessenden, M., *Metabolomics: Small molecules, single cells.* Nature, 2016. **540**(7631): p. 153–155.

Index

V. Soni, T. E. Hartman (eds.), *Metabolomics*,
https://doi.org/10.1007/978-3-031-39094-4