Mini Review

Aleš Kvasnička, Lukáš Najdekr, Dana Dobešová, Barbora Piskláková, Eliška Ivanovová and David Friedecký*

Clinical lipidomics in the era of the big data

https://doi.org/10.1515/cclm-2022-1105
Received October 31, 2022; accepted December 16, 2022; published online January 4, 2023

Abstract: Lipidomics as a branch of metabolomics provides unique information on the complex lipid profile in biological materials. In clinically focused studies, hundreds of lipids together with available clinical information proved to be an effective tool in the discovery of biomarkers and understanding of pathobiochemistry. However, despite the introduction of lipidomics nearly twenty years ago, only dozens of big data studies using clinical lipidomics have been published to date. In this review, we discuss the lipidomics workflow, statistical tools, and the challenges of standartisation. The consequent summary divided into major clinical areas of cardiovascular disease, cancer, diabetes mellitus, neurodegenerative and liver diseases is demonstrating the importance of clinical lipidomics. In these publications, the potential of lipidomics for prediction, diagnosis or finding new targets for the treatment of selected diseases can be seen. The first of these results have already been implemented in clinical practice in the field of cardiovascular diseases, while in other areas we can expect the application of the results summarized in this review in the near future.

Keywords: big data; clinical lipidomics; cohorts; large-scale.

***Corresponding author: David Friedecký**, Laboratory for Inherited Metabolic Disorders, Department of Clinical Biochemistry, University Hospital, Olomouc, Czechia; and Faculty of Medicine and Dentistry, Palacký University Olomouc, Olomouc, Czechia, E-mail: david.friedecky@upol.cz. https://orcid.org/0000-0002-3448-9073
Aleš Kvasnička, Dana Dobešová, Barbora Piskláková and Eliška Ivanovová, Laboratory for Inherited Metabolic Disorders, Department of Clinical Biochemistry, University Hospital, Olomouc, Czechia; and Faculty of Medicine and Dentistry, Palacký University Olomouc, Olomouc, Czechia. https://orcid.org/0000-0001-5973-316X (A. Kvasnička). https://orcid.org/0000-0001-9357-414X (D. Dobešová). https://orcid.org/0000-0001-6962-0775 (B. Piskláková). https://orcid.org/0000-0003-0688-8242 (E. Ivanovová)
Lukáš Najdekr, Institute of Molecular and Translational Medicine, Palacký University Olomouc, Olomouc, Czechia. https://orcid.org/0000-0001-9559-5454

Introduction

Lipids are one of the most important elements of cells, tissues, organs, and thus the whole body. They are responsible for energy metabolism, cell membrane structure and microenvironment as well as signalling; ranging from physiological processes to cellular mechanisms of senescence. Lipidomics – analysis of hundreds of lipids in biological materials – is a bioanalytical field that has been undergoing intensive development for nearly two decades [1, 2]. Lipidomics has now reached a stage where it enables throughput workflows whereby hundreds to thousands of samples can be measured in a single study. Quantitative data on hundreds of lipids for up to ten thousand samples provides a data matrix of more than a million numerical values. To this can be added tens to hundreds of additional continuous or categorical variables within the metadata or clinical data, again multiplying the size of these complex big data matrices.

When working with large-scale lipidomics datasets containing hundreds of variables (lipids) in thousands of samples, gathering relevant information and discovering trends in the data is a challenging task. The goal of lipidomic profiling and subsequent statistical analysis lies in the selection of individual lipids – biomarkers (or panel of biomarkers) differentiating between a cohort of healthy subjects and patients. This information can be used to classify patients using linear, supervised or machine learning models or for determining a diagnosis, which can additionally make the treatment more specific. In the context of understanding the mechanisms of systemic metabolism, it is also advantageous to monitor the regulation of lipid-specific enzymes in lipidomic studies. It is of greater value if lipidomics can be integrated with other clinical omics data. However, this further complexifies the data matrices and requires high-performance computing power for statistical analysis and data processing.

The aim of this review is to summarize the progress in clinical lipidomics studies with respect to big data and to highlight the most important aspects (analytical, statistical, and overall workflow).

Lipidomic workflow

Applied analytical approaches in lipidomics

Historically, thin layer chromatography (TLC) and gas chromatography (GC) techniques were used for lipid analysis, but these usually required derivatization (which is time-consuming and chemically alters the original sample). Since the vast majority of lipidomics publications nowadays use the mass spectrometry (MS) technique for the analysis and since this technique is fully compatible with routine clinical applications (e.g., in neonatal screening or drug monitoring), it will be the main focus of our clinical lipidomics review. MS offers a sensitive, structurally detailed and high-coverage analysis of the lipidome of various biological samples, however, it also faces some shortcomings that have already been discussed in another review [3]. Additionally, for specific and key publications, alternative techniques beyond MS will also be mentioned. From an analytical point of view, clinical lipidomics is mainly performed on a MS instrument (with low or high resolution) with either direct injection or liquid chromatographic separation step. We can distinguish between two approaches applied in lipidomics: (a) untargeted and (b) targeted. The first one is mainly data-driven, and it is used for hypothesis generation. The analytical workflow (sample preparation protocol & analytical method) is optimised to detect as many molecules (lipids) as possible. Since there is no focus on a particular group of lipids, we are not eliminating the possibility of revealing the unanticipated changes, including unknown metabolites and pathways. The second approach is generally hypothesis-driven and quantitative, making it the perfect candidate for clinical (medical) practice application. It offers high sensitivity and specificity for the measurement of known lipids. With targeted lipidomics, we can establish a lipids panel that can be used for diagnostics.

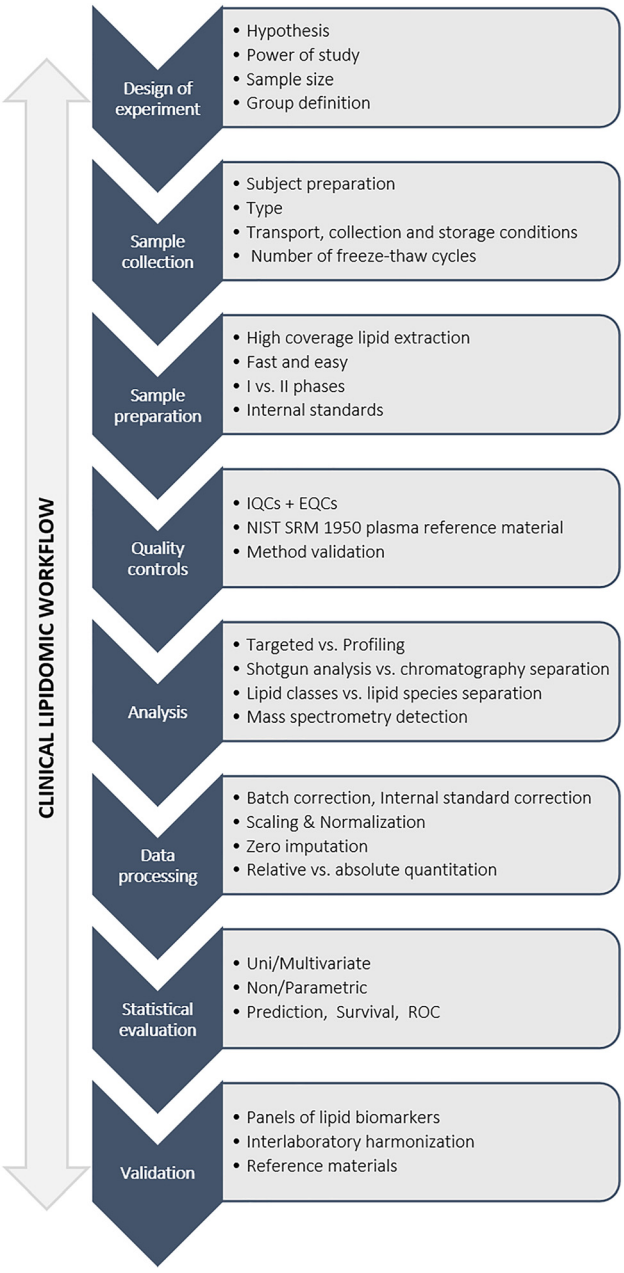

Figure 1: Overview of the clinical lipidomic workflow. IQCs, internal quality controls; EQCs, external quality controls; ROC, receiver operating characteristics.

Overview of general workflow and data acquisition

The general workflow in clinical lipidomics, which typically includes the design of experiment, sample collection, sample preparation, quality control, analysis, data processing, statistical evaluation, and validation, is summarized in Figure 1. The first step – the preanalytical phase/collection of the samples – cannot be underestimated. Many lipids (e.g. polyunsaturated fatty acids (PUFAs)) are prone to oxidative changes or other biological transformations/degradation. Thus, the storage and time between the sample collection and analyses are crucial. In terms of sample preparation, collection and storage, preceding the lipidomic analysis itself, many factors must be taken into account that may affect the obtained lipid profile and its reproducibility. Recommendations regarding these factors, the clinical trial design or reporting of results, have already been summarised and discussed in another review [4]. The non-polar compounds – lipids – are extracted from biological material (plasma, serum, tissues) by the simple addition of organic solvent, centrifuged and analysed by, in most cases, liquid

chromatography coupled with mass spectrometry. The complex mixture of lipid molecules is then separated by either the length of their non-polar side chains and the number of double-bonds (reversed-phase separation) or the polar head group (hydrophilic interaction chromatography known as HILIC) [2]. We can even skip the separation part for high-throughput quantitative assays and use the flow-injection analyses coupled with high-resolution mass spectrometry (FIA-HRMS). Molecules such as free cholesterol (FC), cholesterol esters (CE) [5] or diacylglycerols (DG) and triacylglycerols (TG) can be analysed this way.

The analyses of large sample cohorts bring many challenging difficulties. The main issue is the stability of the analytical platform over a long period. Usually, data for extensive clinical studies need to be acquired over time in multiple batches. Analytical variations and batch effects are taking a severe impact on such data. Applying quality control (QC) and quality assessment (QA) samples is strongly encouraged to overcome this obstacle. The QC samples in the individual batches acquired across multiple weeks will differ in their signal intensities, although the samples are the same. Such an effect needs to be corrected either by the application of a machine learning algorithm (e.g., Systematic Error Removal using Random Forest (SERRF)) [6] or by piecewise polynomial interpolation (e.g., SPLINE). Often only corrections to stably labeled internal standards are used to eliminate the above trends, but this is not able to fully cover the variability in trends even within lipid classes.

After processing the data and obtaining the areas of the raw peaks, filtering and zero imputation are commonly applied and the identified lipids can be quantified at the level of absolute quantity or semiquantity using stable labeled standards. Lipid concentrations in biofluids (most commonly in plasma) are usually reported in molar concentration (nmol/L), but standardization to total lipids in the sample or percentage of lipids relative to lipid class is often used.

Data file size, mainly generated by the untargeted approach, largely depends on the length of chromatographic separation and the number of acquired data points (scanning speed of the instrument). Thus, using backup storage devices with multiple mirroring options across physical storage platforms is highly recommended. The generation of checksums files (e.g., MD5, SHA-256, etc.) should also be applied to ensure correct data transfer between the research groups.

A separate problem is the software dedicated to targeted or untargeted lipidomics. Due to the variety of analytical platforms utilized in lipidomics, many different software solutions for processing lipidomics data have been developed hand-in-hand. Currently, there are many commercial software solutions provided by vendors competing with rapidly evolving free-to-use software, which is driven by an active lipidomic community.

Data handling and statistical methods in big data lipidomics

Due to the complex nature of lipid data containing hundreds and thousands of compounds, data handling and statistical analysis are key steps providing the basis for the subsequent interpretation of results. This overview in Figure 2 was created based on the use of statistical tools in the publications listed in Table 1. At the beginning of each project, a power of study should be calculated to define the relationships between sample numbers and the expected statistical significance of each test. This should be followed by an evaluation of the nature and quality of the data with respect to normality, skewness, outliers, presence of missing values and breadth of scale. This step goes hand in hand with descriptive statistics and logarithmic transformation, which provides a Gaussian distribution for subsequent use of parametric tests. Most statistical analyses then involve univariate methods, usually based on hypothesis testing or correlation analyses, in either parametric or non-parametric settings. In clinical big data projects, regression models are often applied, allowing a combination of ordinal and categorical data types. More recently, multivariate methods based on different variants of supervised discriminant analysis have emerged allowing the creation of statistical models description and classification of samples. The systematic names of lipids contain information about their structure (number of double bonds and number of carbons), these chemical and physical ontologies can be used to find and visualize trends at many different levels. The list presented in Figure 2 does not aim to be a comprehensive overview of all statistical methods but rather represents the trends and options that are currently used in data evaluation. In the near future, we can expect to see greater use of Bayesian statistics and the deployment of artificial intelligence and machine learning, which may bring new advanced tools for data evaluation.

Harmonization of lipidomics

Mass spectrometry-based workflows have become more technically robust and user-friendly for their easier implementation in routine clinical laboratories and thus for an

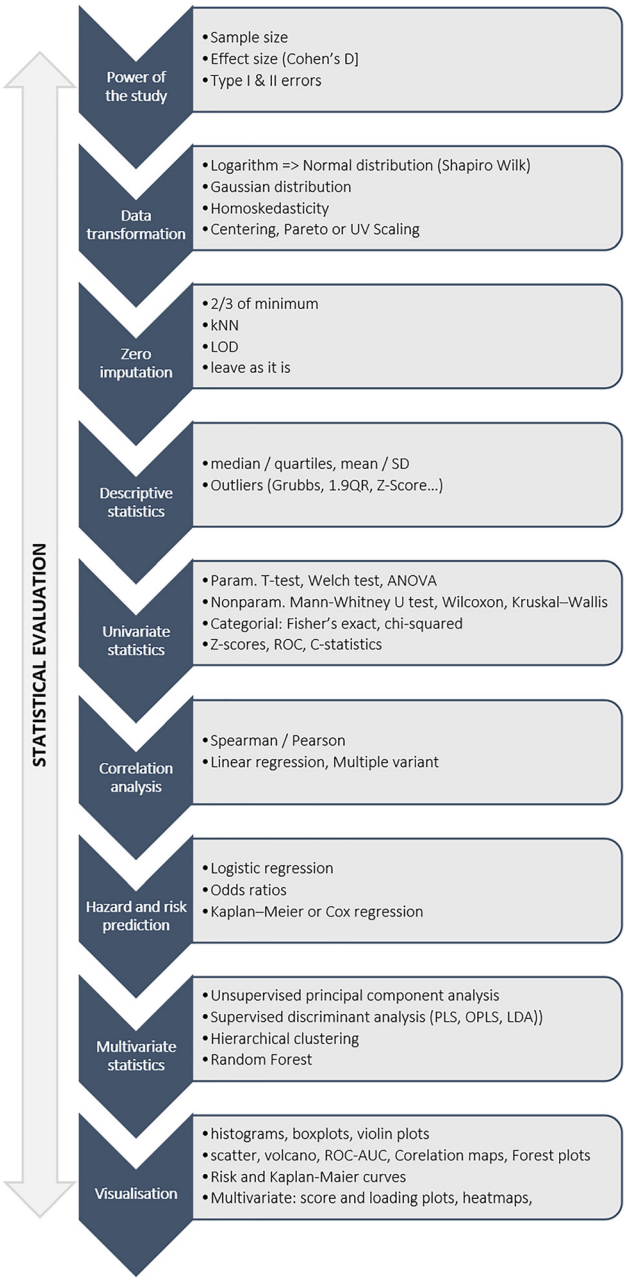

Figure 2: Statistical evaluation of big data clinical lipidomics. UV scaling, unit-variance scaling; kNN, K-nearest neighbours algorithm; LOD, limit of detection; SD, standard deviation; QR, quartile range; ANOVA, analysis of variance; ROC, receiver operating characteristics; PLS, partial least squares discriminant analysis; OPLS, orthogonal projections to latent structures discriminant analysis; LDA, linear discriminant analysis; AUC, area under the curve.

expansion of the number of lipids measured beyond only cholesterol and triacylglycerols. A wide range of analytical platforms are used in studies for mass spectrometry-based lipidomics, but simplified and streamlined workflows which are compatible with routine clinical laboratory settings are in demand. The same issues apply to sample collection and storage, for example, the lack of community-approved analytical protocols brings a challenge in the translation of lipidomics results (i.e. biomarker panels) into the clinical setting [7].

These challenges can be overcome by the efforts of the "Clinical Lipidomics Interest Group" (CLIG) within the International Lipidomics Society (ILS), which was established in 2019. The CLIG is focused on managing the inter-laboratory ring studies that were initiated in 2022. Shared materials (e.g. human plasma reference material) are being analysed by all participating laboratories according to well-described mass spectrometry-based workflows (using validated analytical methods according to FDA and EMA guidelines) [8]. After finding out major gaps in current knowledge, a constructive dialogue between representatives from all disciplines: lipidomics and biomedical researchers, analytical chemists, representatives from regulatory agencies (i.e. FDA, EMA, ILS), and data scientists will be needed to propose solutions to these pitfalls. It is the lack of harmonization and standard reference measurements that is the current challenge in lipidomics. A similar situation occurs with other body fluids such as urine, tears, saliva, or sweat. The standardization of the measurement of these alternative body fluids is another problem, due to uncontrolled procedures for collection, transport, storage, and measurement. Therefore, to further ensure that lipidomics studies are reproducible, a reporting checklist has recently been published, which will help to standardize the description of lipidomic studies [9].

Big data in clinical lipidomics

Lipidomics is increasingly pointing to lipid dysregulation with respect to the pathogenesis of many diseases [10]. Clinical lipidomics is a new extension of lipidomics, focused on the study of lipid profiles, pathways and ontology by characterizing and quantifying the lipid profiles in cells, biopsies and body fluids of patients. To this date, a large number of classes and species of lipids linked to many diseases has been discovered [11]. Large-scale clinical lipidomic studies involving hundreds to thousands of samples are being published, showing how lipidomics could be routinely applied to a huge number of samples.

In the following sections of this review, the focus will be on individual clinical areas where lipidomics has been applied with promising results and the potential of translation into clinical routine.

Table 1: Overview of the large-scale clinical lipidomic studies.

Medical condition/ number of samples	Analytical platform/statistical approach	Study [reference]	Significantly altered lipids
Ageing, lifestyle factors n=10,339	LC(RP)-ESI-MS/MS(QqQ) β-coefficients	Beyene et al. [12]	Sex-specific differences – SM with d18:2 sphingoid base were increased in females. – LPC and ether-linked phospholipids were increased in males. – Menopausal status was associated with a shift in the lipidomic profile. Ageing – PC O-, PC P-, PE O-, TG O- were negatively associated with age. – LPC, LPC O-, LPC P-, PC O-, PC P-, PE O-, PE P- were more strongly negatively associated with age in males than females. – Differential effect of ageing on peroxisomal ether-phospholipid biosynthesis in males and females was found. BMI – LPC, Hex(1/2/3)Cer, GM3 ganglioside and sulfatides were negatively associated with BMI. – SM, (dihydro- and deoxy-)Cer were positively associated with BMI.
Ageing n=980	LC(RP)-ESI-HRMS(QTOF) linear regression	Slade et al. [13]	– Hex2Cer were decreased in females with increasing age. – PC and PE were increased with age in females. – In males, alterations of lipidome were generally less pronounced with age compared to females.
Ageing n=1,882	LC(RP)-ESI-HRMS(QTOF) linear regression, logistic regression, pearson correlation	Cesare et al. [14]	– Carnitine, linoleic acid, and α-linoleic acid were positively correlated with age in females. – LPC were negatively correlated with age both in males and females. – Ratios of selected PC, LPC and MG were sex-dependently correlated with age.
Ageing, lifestyle factors n=1,076	LC(RP)-ESI-MS/MS(QqQ) correlation analysis	Weir et al. [15]	Sex-specific differences – Males have significantly higher plasma levels of Cer than females. – Elevated levels of LPC and LPC O-, in males relative to females. BMI – All dihydroCer were associated with BMI. – Hex(1,2,3)Cer/SM were negatively/positively associated with BMI. – LPC were strongly negatively associated with BMI. Ageing – Cer were strongly associated with age both in males and females. – SM were strongly positively associated with age in females more than in males.
Obesity n=1,061	DI-ESI-HRMS(Orbitrap) correlation analysis, regression models, predictive modeling	Gerl et al. [16]	– Lipidomics was the best predictor for body fat percentage compared to BMI, waist circumference or waist-hip ratio. – SM, which differed by only 1 double bond (SM 34:1 compared to SM 34:2) were the strongest positive and negative predictors.
Obesity n=2,491	LC(RP)-ESI-MS/MS(QqQ) β-coefficients, PCA	Mir et al. [17]	– Several PUFA-containing species PC(16:1_20:4), PC(18:0_20:3) and PE (18:0_20:3) as well as plasmalogen

Table 1: (continued)

Medical condition/ number of samples	Analytical platform/statistical approach	Study [reference]	Significantly altered lipids
			species PE(P-16:0/20:3), PE(P-16:0/20:5) and PE(P-18:0/20:3) were positively associated with BMI in the case of children (6 years old). – Additionally, SM(d18:2/14:0), CE (16:1), PI (16:0_16:1) and several PUFA containing TG also showed a significant positive association with BMI in children whereas in adults, TG containing palmitoleic acid, stearic acid, and oleic acid were the most positively associated lipids. – Several omega-6-containing lipid species were significantly associated with ppBMI, birth weight, and child BMI, suggesting these represent intergenerational signatures of obesity risk.
T2DM n=2,032	LC(RP)-ESI-MS/MS(QqQ) correlation analysis	Chew et al. [18]	– Significant associations were found between sphingolipids and obesity/T2DM characteristics, specifically, those of HexCer and sphingolipids with d16:1 and d18:2 backbones. – SM(d16:1/18:0) and SM(d18:1/18:0) were associated with a higher risk of T2DM.
T2DM n=692	LC(RP)-ESI-HRMS(Orbitrap) PCA, cox regression	Razquin et al. [19]	– LPC, LPE, PC P-, PC O-, SM and CE were negatively associated with risk of T2DM. – TG, DG and PE were positively associated with T2D risk.
T2DM n=954	LC(RP)-ESI-HRMS(QTOF) linear mixed-effect models	Suvitaval et al. [20]	– Higher levels of TG and monounsaturated PC and lower levels of PC O- and LPC were associated with the progression of T2DM.
T2DM, CVD n=4,067	DI-ESI-HRMS(Orbitrap) risk scores, HCA	Lauber et al. [21]	– All TG and DG and most SM species showed elevated concentrations in the highest T2DM risk individuals and contrary most PC O- species showed decreased concentrations.
CVD n=10,803	LC(RP)-ESI-MS/MS(QqQ) cox regression	Hilvo et al. [22]	– Increased levels of Cer(d18:1/16:0), Cer(d18:1/18:0), Cer(d18:1/24:1) and PC(16:0/16:0) and decrease of PC(16:0/22:5) and PC(14:0/22:6) and their ratios were observed in individuals at risk of CVD event or death. – Cer(d18/24:0) was not independently significantly associated with CVD outcomes but was used in ratios with other lipids. – A ceramide- and phospholipid-based risk score (CERT2) was established to efficiently predict residual CVD event risk in patients with coronary artery disease.
CVD, T2DM n=3,779	LC(RP)-ESI-MS/MS(QqQ) cox regression	Alshehry et al. [23]	– Cer, Hex(1,2,3)Cer, mono- and di-unsaturated PC O-, PC P-, LPC, LPC O- and CE were positively and polyunsaturated PC, PC P- and TG were negatively associated with CVD events and death in individuals with T2DM.
CVD n=980	LC(RP)-ESI-MS/MS(QqQ) cox regression	Wang et al. [24]	– Increased levels of Cer(d18:1/16:0), Cer(d18:1/22:0), Cer(d18:1/24:0) and Cer(d18:1/24:1) were associated with higher CVD risk.
CVD n=1,797	LC(RP)-ESI-MS/MS(QqQ) cox regression	Laaksonen et al. [25]	– Cer(d18:1/16:0), Cer(d18:1/18:0), Cer(d18:1/24:0), Cer(d18:1/24:1) and their ratios can be used as predictors of CV death both in patients with stable CAD and ACS.
CVD n=4,249	LC(RP)-ESI-MS/MS(QqQ) cox regression	Lemaitre et al. [26]	– Associations of higher plasma levels of Cer(d18:1/16:0) and SM(d18:1/16:0) with increased risk of heart failure and higher levels of Cer(d18:1/22:0), SM(d18:1/20:0), SM(d18:1/22:0), and SM(d18:1/24:0) with decreased risk of heart failure.

Table 1: (continued)

Medical condition/ number of samples	Analytical platform/statistical approach	Study [reference]	Significantly altered lipids
CVD n=5,776	LC(RP)-ESI-MS/MS(QqQ) cox regression	Peterson et al. [27]	– Ratios of Cer(d18:1/24:0)/Cer(d18:1/16:0) and Cer(d18:1/22:0)/Cer(d18:1/16:0) were inversely associated with CVD risk, HF risk, and all-cause mortality.
CVD n=1,260	LC(RP)-ESI-MS/MS(QqQ) cox regression	Katajamäki et al. [28]	– CERT2 risk score, based on increased levels of Cer(d18:1/16:0), Cer(d18:1/18:0), Cer(d18:1/24:1) and PC(16:0/16:0) and decrease of PC(16:0/22:5) and PC(14:0/22:6) and their ratios (also with Cer(d18:1/24:0)), was strongly associated with CVD, CAD, and stroke mortality in the elderly, while the association between these events and conventional lipids was weak.
Cancer n=830	Multiple MS-based methods PCA, OPLS-DA, ROC, Survival analysis	Wolrab et al. [29]	– Dysregulation in some very long chain sphingomyelins (SM 41:1, SM 42:1, SM 42:2), ceramides (Cer 41:1, Cer 42:1) and LPC 18:2 in plasma was observed in pancreatic cancer patients. The sensitivity and specificity to diagnose pancreatic cancer using lipidomics was over 90%. – Prognostic biomarkers were identified such as LPC 18:2, which was positively correlated, and in contrast, Cer 36:1, Cer 38:1, Cer 42:2, PC 32:0, PC O-38:5, and SM 42:2 which were negatively correlated with survival.
Cancer n=553	DI-ESI-HRMS(FTICR-MS) PLS-DA, ROC	Guo et al. [30]	– Using oleamide, long-chain acylcarnitines, LPC(18:1), LPC(20:4), LPC(20:3), LPC(22:6), and SM(d18:1/16:0) as a serum biomarker panel resulted in a separation between the lung cancer patients and healthy controls with sensitivity and specificity of 100 and 91%, respectively.
Cancer n=1,449	DI-ESI-HRMS(FTICR-MS) Mann–Whitney U test, PLSDA, ROC	Guo et al. [31]	– Increased levels of SM(34:1), PC(34:2), PC(34:1), PC(36:3), and PC(36:2) were observed in the plasma of lung cancer patients.
Cancer n=674	MALDI-HRMS(Orbitrap) Mann–Whitney U test, ROC, logistic regression	Jirásko et al. [32]	– Elevated lactosylsulfatides and decreased sulfatides with hydroxylated fatty acyls were found in the tumour, plasma and urine of RCC patients compared to controls, showing a diagnostical potential of this lipid panel.
Cancer n=1,440	MALDI-HRMS(FTICR-MS) Mann–Whitney U test, ANOVA, ROC	Ren et al. [33]	– Increased levels of total fatty acids FA16:0, FA16:1, FA18:0, FA18:1, FA18:3, FA20:3, and FA 22:6 and decreased levels of FA20:5 in serum were observed in lung cancer patients.
Cancer n=400	MALDI-HRMS(TOF), LC(HILIC)-MS(IT) Welch test, logistic regression, ROC	Ros-Mazurczyk et al. [34]	– Downregulation of lysophosphatidylcholines (LPC18:2, LPC18:1 and LPC18:0) in serum samples from lung cancer patients was observed.
Neurodegenerative disorders n=1,912	LC(RP)-ESI-MS/MS(QqQ) cox regression	Hyuhn et al. [35]	– Lipid signatures associated with prevalent AD arising from altered lipid pathways including; ether lipids and plasmalogens (such as PE P-18:0/22:6), sphingolipids (notably Cer and GM3 gangliosides) were identified. Multivariate lipid-based models improved the classification and prediction of AD.
Neurodegenerative disorders n=274	LC(RP)-ESI-HRMS(QTOF) PLS-DA, ROC	Sinclair et al. [36]	– Alterations in lipid metabolism related to the carnitine shuttle, sphingolipid metabolism (Cer and HexCer), arachidonic acid metabolism and fatty acid biosynthesis (FA 26:0, FA 22:0) were found in the sebum of PD patients.
Neurodegenerative disorders n=198	LC(HILIC)-ESI-MS/MS(QqQ) logistic regression, ROC, PCA	Hwangbo et al. [37]	– SM(d18:1/18:1), CE(16:1), CE(20:1) and PC(18:0/20:3) were positively associated and PE(P-18:0/22:6), PE(18:0/20:4), PE(18:0/22:6) were negatively associated with AD compared to controls.

Table 1: (continued)

Medical condition/ number of samples	Analytical platform/statistical approach	Study [reference]	Significantly altered lipids
			– PE(P-16:0/18:1), HexCer(18:0), FA(16:1), SM(d18:1/18:1), FA(24:0), PC(16:0/20:2), FA(20:2), CE(20:1), DG(20:0/20:0), PE(16:0/22:6) LPC(18:1) were positively associated with PD. – PC(18:1/18:2), FA(18:0), PE(18:1/18:1), FA(24:1), PC(18:1/20:4), PC(18:0/22:6), PC(18:1/16:1 were negatively associated with PD.
Liver disease n=467	Multiple LC-MS-based methods Wilcoxon rank sum test, Kruskal–Wallis rank sum test, ROC, random forest	Barr et al. [38]	– DG and TG species were elevated in steatosis patients, as compared to normal liver subjects. – PE O-, PE P-, PC O- and PC P- were decreased in steatosis patients as compared to normal liver subjects, with elevated levels in NASH.
Liver disease n=535	Multiple LC-MS-based methods Student's t-test, Volcano plot, HCA	Alonso et al. [39]	– Serum lipidomic and metabolomic analysis of patients with NAFLD revealed 2 major subtypes (M and non-M) of NAFLD and markers that differentiate steatosis from NASH in each subtype were identified. – A metabolic signature which was able to separate NAFLD patients of the M subtype into NASH and simple steatosis included aminoacids (serine, asparagine, methionine); LPE: LPE(16:0), LPE(18:0); PC: PC(16:0/22:6), PC(20:0/20:4), PC(18:0/22:5); LPC: LPC(22:6), LPC(20:1), LPC(18:1); LPC P-/LPC O-: LPC(P-18:0), LPC(O-22:0), LPC(O-24:1) and SM: SM(33:1), SM(39:1), SM(d16:1/24:1), SM(d18:2/22:0).
Liver disease n=679	LC(RP)-ESI-HRMS(qTOF) Bayesian-model-based clustering, logistic and ridge regression, ROC	Orešič et al. [40]	– A serum-lipid signature comprising of TG(16:0/18:0/18:1), phosphatidylcholine PC(18:1/22:6), PC(O-24:1/20:4) was developed to estimate the percentage of liver fat with the sensitivity of 69.1% and specificity of 73.8% when applied for diagnosis of NAFLD.
Liver disease n=375	DI-ESI-HRMS(Orbitrap) random forest, PR and ROC curves	Vvedenskaya et al. [41]	– Accumulation of DG, TG and CE in the liver of NAFLD and NASH patients was observed, while the bulk composition of glycerophospho- and sphingolipids remained unchanged. – Further stratification identified SM species comprising n24:2 fatty acid moieties as membrane lipid markers of NAFLD. – Normalized relative abundance of SM 43:3; O2 and SM 43:1; O2 containing n24:2 and n24:0 fatty acid moieties, respectively, showed opposite trends during NAFLD progression and distinguished NAFL and NASH lipidomes from the lipidome of nonsteatotic livers.

T2DM, type 2 diabetes mellitus; CVD, cardiovascular disease; LC, liquid chromatography; MS, mass spectrometry; RP, reversed-phase; HILIC, hydrophilic interaction chromatography; ESI, electrospray ionisation; DI, direct injection; MALDI, matrix-assisted laser desorption/ionization; MS/MS, tandem mass spectrometry; QqQ, triple quadrupole; IT, ion trap; HRMS, high-resolution mass spectrometry; QTOF, quadrupole time-of-flight; FTICR-MS, Fourier-transform ion cyclotron resonance mass spectrometry; PCA, principal component analysis; HCA, hierarchical clustering; PLS-DA, partial least squares discriminant analysis; OPLS-DA, orthogonal projections to latent structures discriminant analysis; ROC, receiver operating characteristic; PR, precision recall; ANOVA, analysis of variance; SM, sphingomyelin; LPC, lysophosphatidylcholine; LPC O-/P-, lysophosphatidylcholine with alkyl/alkenyl ether substituent; LPE, lysophosphatidylethanolamine; PC O-/P-, phosphatidylcholine with alkyl/alkenyl ether substituent; PE, phosphatidylethanolamine; PE O-/P-, phosphatidylethanolamine with alkyl/alkenyl ether substituent; TG, triacylglycerol; TG O-, triacylglycerol with alkyl ether substituent; DG, diacylglycerol; MG, monoacylglycerol; Hex(1/2/3)Cer, mono-/di-/tri-hexosylceramide; GM3, monosialodihexosylganglioside; Cer, ceramide; (dihydro- and deoxy-)Cer, dihydro-/deoxy-ceramide; CE, cholesteryl ester; PI, phosphatidylinositol; PUFA, polyunsaturated fatty acid; BMI, body mass indes; ppBMI, pre-pregnancy body mass index; ACS, acute coronary syndromes; HF, hearth failure; CAD, coronary artery disease; RCC, renal cell cancer; FA, esterified or non-esterified fatty acid; AD, Alzheimer's disease; PD, parkinson's disease; NASH, nonalcoholic steatohepatitis; NAFLD, nonalcoholic fatty liver disease.

Cardiovascular diseases

The number of patients with cardiovascular disease (CVD) is increasing due to an ageing population and improving survival rate after CVD events due to better healthcare. However, at the individual level or in specific or comorbid subpopulations, the risk of an event can vary widely, requiring risk estimation tools to improve patient management. Effective risk stratification should identify at-risk individuals requiring more intensive or different treatment. Conversely, patients with favourable prognosis should be identified to avoid the overuse of medications and associated adverse events. For all these reasons, clinical lipidomics studies have been carried out to search for lipids with prognostic value as markers of CVD-related events and death. Different plasma ceramide ratios have been identified as significant predictors of CVD-related death in patients with stable coronary artery disease and acute coronary syndrome [25]. Additionally, polyunsaturated phosphatidylcholines (PC) in plasma, in combination with ceramides, can be used to further enhance the predictive ability and accuracy of prediction models [22]. Based on these and other studies, the CERT2 score for cardiovascular risk and event prediction using plasma ceramides (with possible improvement by including polyunsaturated PC) has been developed and its usefulness has already been summarized in previous reviews [42, 43]. It has also been shown that the CERT2 score is strongly associated with CVD, coronary artery disease, and stroke mortality in the elderly, while the association between these events and conventionally measured lipids (HDL-C, LDL-C, TG, total cholesterol) was weak [28]. Moreover, several mono-, di- and trihexosylceramides, alkyl- and alkenyl- phosphatidylcholines and other lipids improved current lifestyle and clinical risk factors for prediction of CVD events or death in patients with type 2 diabetes mellitus (T2DM) [23]. Multiple large cohort studies have provided similar results with regard to Mediterranean dietary intervention [24], heart failure risk [26] or all-cause mortality [27].

Ageing, obesity, diabetes and lifestyle factors

Changes in the lipidome due to ageing, lifestyle, obesity, and T2DM have been the focus of many lipidomic studies. Lipidomic fingerprints have been shown to be related to age with sex-specific differences (also influenced by the postmenopausal stage of life in females) in multiple studies [12, 13]. Some studies point to the interconnection of physiological factors (age, sex, BMI) to the alterations of lipidome [15]. Ratios between lipids can be used as potential biomarkers of age, as these sometimes describe the lipid metabolism dynamics better than sole lipid abundances [14]. Obesity is caused by excessive accumulation of fats, i.e. various lipids, which have also historically been determined as biochemical parameters in routine clinical tests. However, a detailed view using lipidomics can help understand how obesity relates to increased mortality or risk of developing T2DM or CVD. By using lipidomic analysis of human plasma, it was possible to predict BMI, total body fat percentage, waist circumference and waist-hip ratio and its associations with sphingomyelin differing by 1 double bond (suggesting involvement of specific desaturases) have been shown [16]. A mother-offspring study has pointed to the change in lysophospholipids in relation to birth weight and multiple other lipid classes were altered with respect to obesity risk [17]. Several other studies have shown that lipidomics can identify a subset of individuals at high risk for T2DM (even in combination with CVD) years before the onset of the disease, and thus the potential utilization of a lipid panel to identify these at-risk individuals should be considered [19–21, 23].

Cancer

Alterations in the lipidome have been demonstrated in almost all types of cancer. Not just in cancer cell lines *in vitro*, but also in patients at the level of the tumour, surrounding tissues and in peripheral biofluids [44]. However, only a limited number of studies have demonstrated the diagnostical potential of lipidomics in routine analysis or screening. Based on changes in serum/plasma lipids, it was possible to distinguish a group of colorectal and gastric cancer patients from controls [31] or lung cancer, including benign lung disease [30, 33, 34]. Another recent study showed increased levels of lactosylsulfatides and decreased levels of sulfatides with hydroxylated fatty acyls in patients with renal cell carcinoma compared to controls not only at the tumour tissue level but also in plasma and urine [32]. One of the most promising studies, which results are supported by several parallel methods carried out in multiple laboratories, showed dysregulation of very long-chain sphingomyelins, ceramides and (lyso)phosphatidylcholines in the plasma of pancreatic cancer patients [29]. The sensitivity and specificity of these markers exceed 90%, outperforming even conventional CA 19-9, especially at an early stage, and the accuracy was comparable to established diagnostic imaging methods [29]. Additionally, we would like to mention that

another direction in the clinical lipidomics of cancer points to a significant impact on the metabolism of fatty acids, such as sapienic acid [45] and others, whereas the results and potential of these studies have already been summarized previously in another review [46].

Neurodegenerative disorders

To date, not many large-scale lipidomics studies focused on neurodegenerative diseases such as Alzheimer's or Parkinson's disease have been carried out. One study discovered lipid patterns associated with Alzheimer's disease resulting from changes in ether lipids, sphingolipids (especially GM3 gangliosides), phosphatidylethanolamines, and triacylglycerols. Constructed subsequent multivariate lipid models improved the classification of disease [35]. Additionally, metabolites and lipids identified in cerebrospinal fluid combined into an elastic net regression model were able to discriminate between AD, PD and healthy controls [37]. In another study, the non-invasively collected sebum from people with Parkinson's disease revealed a difference in its lipid composition compared to controls, independent of dopaminergic medication [36].

Non-alcoholic fatty liver disease

The mechanisms by which non-alcoholic fatty liver disease (NAFLD) progresses from simple steatosis to steatohepatitis (NASH) is not well understood. It is known that obesity is a major cause of NAFLD, but it is not known why fatty liver develops in lean individuals nor why NASH develops only in a small part of NAFLD patients. A study focusing on serum lipidomic analysis of patients (with histologically confirmed NAFLD based on liver biopsy) showed that a BMI-dependent lipidomic profile can be used to reliably distinguish patients with NASH from steatosis, and highlighted new targets for therapeutic intervention [38]. Analysis of serum lipidome of patients with NAFLD and MAT1A-KO mice with steatohepatitis revealed three subtypes of NAFLD (M, non-M and indeterminate subtype) and identified markers that differentiate steatosis from NASH in each subtype [39]. It has been shown that individuals with NAFLD had increased serum levels of triacylglycerols with low carbon number and double-bond content while lysophosphatidylcholines and ether phospholipids were decreased compared to healthy controls [40]. By lipidomic analysis of the patient's liver tissues, it was possible to describe the transition from a healthy liver to NAFLD to NASH. Changes were observed in the concentrations of unsaturated cholesteryl esters, di- and triacylglycerols and sphingomyelin ratios [41].

Conclusions and perspectives

The upcoming challenges lie in expanding from the current clinical concept of a "single-lipid-marker" to a "multi-analyte-lipid panels". These lipidomic panels have the potential to be applied in personalised medicine in the future (for disease prediction or progression monitoring). Although multiple diagnostic biomarkers for many diseases are constantly identified, it remains challenging to implement these lipid-based molecules in everyday medicine. One of the reasons is the already discussed standatisation of lipidomic analyses. A new approaches to interpret (visualise) the results based on a combination of hundreds of lipid species is in demand. Lipidomic analyses vary in coverage (how many lipids can be measured in a single analysis) and in the depth of structural identification from the level of summary description (number of carbons and double bonds) to the composition of acyl chains. In contrast, a complete description of lipids, which also includes the position and stereochemistry of their double bonds, requires state-of-the-art instrumentation such as ion mobility mass spectrometry or new types of fragmentation (such as electron-activated dissociation). Another instrumental improvement in clinical lipidomics is a surgical scalpel called iKnife. It combines a technique called rapid evaporative ionization mass spectrometry (REIMS) with an electrosurgical knife. iKnife converts tissue constituents including lipids into ions in the gas phase, which are analysed in real-time, thus enabling differentiation of tumour from normal surrounding tissue.

Based on the results of the clinical lipidomics studies, it is evident that lipidomics has potential not only in diagnosing diseases but also in stratification of individuals at-risk and monitoring the disease progression. Thus, we can envision lipidomics as a tool where a panel of lipids (and their ratios) could be used to alert physicians to the risk of developing various diseases (CVD, T2DM, cancer, neurological disorders, liver disease) from a single analysis. Just as the targeted panel of metabolites is established in newborn screening (NBS) of inborn errors of metabolism (IEM), lipidomics is still waiting for its everyday application, where it has the great potential (summarized elsewhere [47]). There are certainly many drawbacks lying ahead in the application of clinical lipidomics in the context of big data. The importance and increasing interest in metabolomics and lipidomics by diagnostic laboratories are demonstrated by the metabolomics working group

(WG-M) established at the International federation of clinical chemistry and laboratory medicine (IFCC).

Research funding: Supported by a doctoral student grant (DSGC-2021-0098) from Palacký University Olomouc and MH CZ – DRO FNOl, 00098892. The project National Institute for Cancer Research (Programme EXCELES, ID Project No. LX22NPO5102) – Funded by the European Union – Next Generation EU. This research was supported by grants from the Czech Ministry of Education, Youth and Sports (EATRIS-CZ – LM2018133).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

References

1. Carrasco-Pancorbo A, Navas-Iglesias N, Cuadros-Rodríguez L. From lipid analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part I: modern lipid analysis. Trends Anal Chem 2009;28:263–78.
2. Teo CC, Chong WPK, Tan E, Basri NB, Low ZJ, Ho YS. Advances in sample preparation and analytical techniques for lipidomics study of clinical samples. Trends Anal Chem 2015;66:1–18.
3. Köfeler HC, Ahrends R, Baker ES, Ekroos K, Han X, Hoffmann N, et al. Recommendations for good practice in MS-based lipidomics. J Lipid Res 2021;62:100138.
4. Brenna JT, Plourde M, Stark KD, Jones PJ, Lin YH. Best practices for the design, laboratory analysis, and reporting of trials involving fatty acids. Am J Clin Nutr 2018;108:211–27.
5. Höring M, Ejsing CS, Hermansson M, Liebisch G. Quantification of cholesterol and cholesteryl ester by direct flow injection high-resolution fourier transform mass spectrometry utilizing species-specific response factors. Anal Chem 2019;91:3459–66.
6. Fan S, Kind T, Cajka T, Hazen SL, Tang WHW, Kaddurah-Daouk R, et al. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. Anal Chem 2019;91:3590–6.
7. Lipidomics Standards Initiative Consortium. Lipidomics needs more standardization. Nat Metab 2019;1:745–7.
8. Vvedenskaya O, Holčapek M, Vogeser M, Ekroos K, Meikle PJ, Bendt AK. Clinical lipidomics – a community-driven roadmap to translate research into clinical applications. J Mass Spectrom Adv Clin Lab 2022; 24:1–4.
9. McDonald JG, Ejsing CS, Kopczynski D, Holčapek M, Aoki J, Arita M, et al. Introducing the lipidomics minimal reporting checklist. Nat Metab 2022;4:1086–8.
10. Avela HF, Sirén H. Advances in lipidomics. Clin Chim Acta 2020;510: 123–41.
11. Meikle TG, Huynh K, Giles C, Meikle PJ. Clinical lipidomics: realizing the potential of lipid profiling. J Lipid Res 2021;62:100127.
12. Beyene HB, Olshansky G, Smith TAA, Giles C, Huynh K, Cinel M, et al. High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age and BMI: evidence from two large population cohort studies. PLoS Biol 2020;18:e3000870.
13. Slade E, Irvin MR, Xie K, Arnett DK, Claas SA, Kind T, et al. Age and sex are associated with the plasma lipidome: findings from the GOLDN study. Lipids Health Dis 2021;20:30.
14. Di Cesare F, Luchinat C, Tenori L, Saccenti E. Age- and sex-dependent changes of free circulating blood metabolite and lipid abundances, correlations, and ratios. J Gerontol A Biol Sci Med Sci 2022;77:918–26.
15. Weir JM, Wong G, Barlow CK, Greeve MA, Kowalczyk A, Almasy L, et al. Plasma lipid profiling in a large population-based cohort. J Lipid Res 2013;54:2898–908.
16. Gerl MJ, Klose C, Surma MA, Fernandez C, Melander O, Männistö S, et al. Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. PLoS Biol 2019;17:e3000443.
17. Mir SA, Chen L, Burugupalli S, Burla B, Ji S, Smith AAT, et al. Population-based plasma lipidomics reveals developmental changes in metabolism and signatures of obesity risk: a mother-offspring cohort study. BMC Med 2022;20:242.
18. Chew WS, Torta F, Ji S, Choi H, Begum H, Sim X, et al. Large-scale lipidomics identifies associations between plasma sphingolipids and T2DM incidence. JCI Insight 2019;5:1–14.
19. Razquin C, Toledo E, Clish CB, Ruiz-Canela M, Dennis C, Corella D, et al. Plasma lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. Diabetes Care 2018;41:2617–24.
20. Suvitaival T, Bondia-Pons I, Yetukuri L, Pöhö P, Nolan JJ, Hyötyläinen T, et al. Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. Metabolism 2018;78:1–12.
21. Lauber C, Gerl MJ, Klose C, Ottosson F, Melander O, Simons K. Lipidomic risk scores are independent of polygenic risk scores and can predict incidence of diabetes and cardiovascular disease in a large population cohort. PLoS Biol 2022;20:e3001561.
22. Hilvo M, Meikle PJ, Pedersen ER, Tell GS, Dhar I, Brenner H, et al. Development and validation of a ceramide- and phospholipid-based cardiovascular risk estimation score for coronary artery disease patients. Eur Heart J 2020;41:371–80.
23. Alshehry ZH, Mundra PA, Barlow CK, Mellett NA, Wong G, McConville MJ, et al. Plasma lipidomic profiles improve on traditional risk factors for the prediction of cardiovascular events in type 2 diabetes mellitus. Circulation 2016;134:1637–50.
24. Wang DD, Toledo E, Hruby A, Rosner BA, Willett WC, Sun Q, et al. Plasma Ceramides, Mediterranean Diet, and Incident Cardiovascular Disease in the PREDIMED Trial (Prevención con Dieta Mediterránea). Circulation 2017;135:2028–40.
25. Laaksonen R, Ekroos K, Sysi-Aho M, Hilvo M, Vihervaara T, Kauhanen D, et al. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. Eur Heart J 2016;37:1967–76.
26. Lemaitre RN, Jensen PN, Hoofnagle A, McKnight B, Fretts AM, King IB, et al. Plasma ceramides and sphingomyelins in relation to heart failure risk. Circ Heart Fail 2019;12:e005708.
27. Peterson LR, Xanthakis V, Duncan MS, Gross S, Friedrich N, Völzke H, et al. Ceramide remodeling and risk of cardiovascular events and mortality. J Am Heart Assoc 2018;7:1–11.
28. Katajamäki TT, Koivula M-K, Hilvo M, Lääperi MTA, Salminen MJ, Viljanen AM, et al. Ceramides and phosphatidylcholines associate with cardiovascular diseases in the elderly. Clin Chem 2022;68:1502–8.

29. Wolrab D, Jirásko R, Cífková E, Höring M, Mei D, Chocholoušková M, et al. Lipidomic profiling of human serum enables detection of pancreatic cancer. Nat Commun 2022;13:124.
30. Guo Y, Wang X, Qiu L, Qin X, Liu H, Wang Y, et al. Probing gender-specific lipid metabolites and diagnostic biomarkers for lung cancer using Fourier transform ion cyclotron resonance mass spectrometry. Clin Chim Acta 2012;414:135–41.
31. Guo Y, Ren J, Li X, Liu X, Liu N, Wang Y, et al. Simultaneous quantification of serum multi-phospholipids as potential biomarkers for differentiating different pathophysiological states of lung, stomach, intestine, and pancreas. J Cancer 2017;8:2191–204.
32. Jirásko R, Idkowiak J, Wolrab D, Kvasnička A, Friedecký D, Polański K, et al. Altered plasma, urine, and tissue profiles of sulfatides and sphingomyelins in patients with renal cell carcinoma. Cancers 2022;14: 1–18.
33. Ren J, Zhang D, Liu Y, Zhang R, Fang H, Guo S, et al. Simultaneous quantification of serum nonesterified and esterified fatty acids as potential biomarkers to differentiate benign lung diseases from lung cancer. Sci Rep 2016;6:34201.
34. Ros-Mazurczyk M, Jelonek K, Marczyk M, Binczyk F, Pietrowska M, Polanska J, et al. Serum lipid profile discriminates patients with early lung cancer from healthy controls. Lung Cancer 2017;112: 69–74.
35. Huynh K, Lim WLF, Giles C, Jayawardana KS, Salim A, Mellett NA, et al. Concordant peripheral lipidome signatures in two large clinical studies of Alzheimer's disease. Nat Commun 2020;11:5698.
36. Sinclair E, Trivedi DK, Sarkar D, Walton-Doyle C, Milne J, Kunath T, et al. Metabolomics of sebum reveals lipid dysregulation in Parkinson's disease. Nat Commun 2021;12:1592.
37. Hwangbo N, Zhang X, Raftery D, Gu H, Hu S-C, Montine TJ, et al. Predictive modeling of alzheimer's and Parkinson's disease using metabolomic and lipidomic profiles from cerebrospinal fluid. Metabolites 2022;12:1–28.
38. Barr J, Caballería J, Martínez-Arranz I, Domínguez-Díez A, Alonso C, Muntané J, et al. Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. J Proteome Res 2012; 11:2521–32.
39. Alonso C, Fernández-Ramos D, Varela-Rey M, Martínez-Arranz I, Navasa N, Van Liempd SM, et al. Metabolomic identification of subtypes of nonalcoholic steatohepatitis. Gastroenterology 2017;152: 1449–61.e7.
40. Orešič M, Hyötyläinen T, Kotronen A, Gopalacharyulu P, Nygren H, Arola J, et al. Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. Diabetologia 2013;56:2266–74.
41. Vvedenskaya O, Rose TD, Knittelfelder O, Palladini A, Wodke JAH, Schuhmann K, et al. Nonalcoholic fatty liver disease stratification by liver lipidomics. J Lipid Res 2021;62:100104.
42. Meeusen JW, Donato LJ, Kopecky SL, Vasile VC, Jaffe AS, Laaksonen R. Ceramides improve atherosclerotic cardiovascular disease risk assessment beyond standard risk factors. Clin Chim Acta 2020;511: 138–42.
43. Hilvo M, Vasile VC, Donato LJ, Hurme R, Laaksonen R. Ceramides and ceramide scores: clinical applications for cardiometabolic risk stratification. Front Endocrinol 2020;11:570628.
44. Wolrab D, Jirásko R, Chocholoušková M, Peterka O, Holčapek M. Oncolipidomics: mass spectrometric quantitation of lipids in cancer research. TrAC, Trends Anal Chem 2019;120:115480.
45. Küçüksayan E, Sansone A, Chatgilialoglu C, Ozben T, Tekeli D, Talibova G, et al. Sapienic acid metabolism influences membrane plasticity and protein signaling in breast cancer cell lines. Cells 2022;11:1–16.
46. Ferreri C, Sansone A, Ferreri R, Amézaga J, Tueros I. Fatty acids and membrane lipidomics in oncology: a cross-road of nutritional, signaling and metabolic pathways. Metabolites 2020;10:1–26.
47. Ivanovová E, Pisklákova B, Dobešová D, Kvasnička A, Friedecký D. Novel LC-MS tools for diagnosing inborn errors of metabolism. Microchem J 2021;170:106654.