LipidSig: a web-based tool for lipidomic data analysis

Wen-Jen Lin[1,†], **Pei-Chun Shen**[2,†], **Hsiu-Cheng Liu**[2], **Yi-Chun Cho**[2], **Min-Kung Hsu**[2], **I-Chen Lin**[1], **Fang-Hsin Chen**[3,4,5], **Juan-Cheng Yang**[6], **Wen-Lung Ma**[1] and **Wei-Chung Cheng** [1,2,7,*]

[1]Graduate Institute of Biomedical Science, China Medical University, Taichung 40403, Taiwan, [2]Research Center for Cancer Biology, China Medical University, Taichung 40403, Taiwan, [3]Department of Medical Imaging and Radiological Sciences, Chang Gung University, Taoyuan 33302, Taiwan, [4]Department of Radiation Oncology, Chang Gung Memorial Hospital at Linkou, Taoyuan 33302, Taiwan, [5]Institute for Radiological Research, Chang Gung Memorial Hospital, Chang Gung University, Taoyuan 33302, Taiwan, [6]Chinese Medicine Research and Development Center, China Medical University Hospital, Taichung 40403, Taiwan and [7]The Ph.D. program for Cancer Biology and Drug Discovery, China Medical University and Academia Sinica, Taichung 40403, Taiwan

ABSTRACT

With the continuing rise of lipidomic studies, there is an urgent need for a useful and comprehensive tool to facilitate lipidomic data analysis. The most important features making lipids different from general metabolites are their various characteristics, including their lipid classes, double bonds, chain lengths, etc. Based on these characteristics, lipid species can be classified into different categories and, more interestingly, exert specific biological functions in a group. In an effort to simplify lipidomic analysis workflows and enhance the exploration of lipid characteristics, we have developed a highly flexible and user-friendly web server called LipidSig. It consists of five sections, namely, Profiling, Differential Expression, Correlation, Network and Machine Learning, and evaluates lipid effects on cellular or disease phenotypes. One of the specialties of LipidSig is the conversion between lipid species and characteristics according to a user-defined characteristics table. This function allows for efficient data mining for both individual lipids and subgroups of characteristics. To expand the server's practical utility, we also provide analyses focusing on fatty acid properties and multiple characteristics. In summary, LipidSig is expected to help users identify significant lipid-related features and to advance the field of lipid biology. The LipidSig webserver is freely available at http://chenglab.cmu.edu.tw/lipidsig

GRAPHICAL ABSTRACT
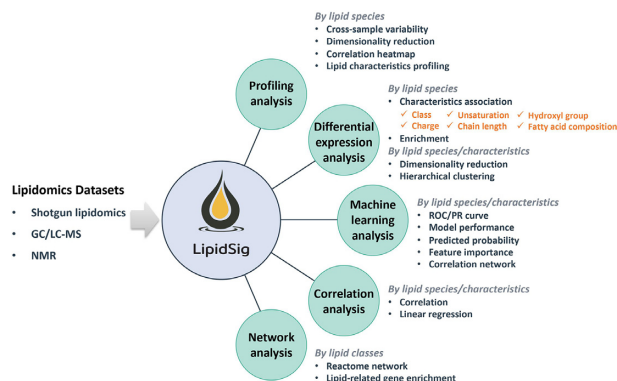
INTRODUCTION

Cells are composed, in part, of a diverse set of functional lipids with different backbones, head groups, fatty acid linkages and carbon chain compositions (1). Beyond their roles as building blocks and energy sources, lipids can mediate signalling transduction (2), protein modification (3) and cell death (4), or constitute dynamic assemblies such as lipid rafts (5), lipoproteins (6) and lipid droplets (7). Additionally, disturbances in lipid homeostasis have been linked to many disorders, including cardiovascular disease (8), obesity (9), diabetes (10) and cancer (11). Given their central role in physiological and pathological conditions, understanding the diversity and composition of lipids helps us explore the potential biological functions underlying those conditions.

With the advances in analytical chemistry and shotgun liquid chromatography–tandem mass spectrometry (LC–MS/MS), researchers have been enabled to detect hundreds

to thousands of lipid species in order to assess their phenotype associations, allowing in turn for the creation of a new field called lipidomics (12,13). Although lipidomics is one of the novel omics technologies, it has been widely applied in biomedical research, agriculture, and food-related industries (14). An exponential increase in lipidomic-related studies has also been observed among PubMed publications (15). Lipidomic profiling yields high-dimensional data such as the data yielded by genomics and metabolomics, but the relative scarcity of analysis tools remains a problem for its further development. Moreover, in contrast with other omics, each lipid species in lipidomic data can be grouped into distinct categories based on a diverse set of characteristics including its lipid class, shape, chain length, degree of unsaturation, hydroxyl groups, and fatty acid composition (1). These characteristics have been proven to dramatically affect cellular function. For example, more unsaturated (double bonds) cell membranes promote Akt clustering and activation to potentiate osteogenic differentiation (16), while LPCAT1 enhances the expression of saturated phosphatidylcholine (PC), which is required for EGFR signalling and glioblastoma survival (17). Another example is that human cytomegalovirus has been found to manipulate target cells to synthesize very-long-chain fatty acids (VEFAs) as their envelopes by inducing the production of elongase enzyme 7 (ELOVL7) (18). Recently, polyunsaturated phospholipids with hydroperoxyl or hydroxyl groups have also drawn considerable research attention due to their causal relationship with ferroptosis (19,20). Hence, analyses of lipidomics data need to consider not only lipid species but also lipid characteristics.

A range of web servers or software packages have been proposed to deal with lipidomic data (21). Of these, Lipid Data Analyzer (22), LipidBlast (23), LipidHunter (24), LipidMatch (25), are designed to identify and quantify lipids from mass spectrometry (MS) data. Only three tools, lipidr (21), ALEX (26), and Lipid Mini-On (27), are able to perform lipid characteristics analysis. However, some of these have strict formatting requirements on lipid names, and thus cannot process data from all MS platforms. The others provide profiling functions to visualize changes in different characteristics but lack statistical analysis tools. More importantly, they make it difficult to undertake characteristics analyses beyond lipid class, chain length, and double bonds. There is, therefore, an urgent need to develop a robust tool that can easily analyze lipidomic data based on lipid-specific characteristics, and interrogate their associations with underlying cellular mechanisms.

Here, we present LipidSig, the first web-based platform that provides an integrated, comprehensive analysis for streamlined data mining of lipidomic datasets. LipidSig gives users more flexibility in input data format, and allows them to define a variety of characteristics for exploration. We provide five main functions, namely Profiling, Differential Expression, Correlation, Network, and Machine Learning, to evaluate the importance of lipid species or characteristics in different experimental groups. To the best of our knowledge, LipidSig is the first tool to fill the gaps in the lipid-specific analysis essential for lipid biology. Furthermore, interactive plots with downloadable images and corresponding tables are created to support interpreting from

multiple perspectives. Collectively, LipidSig enables users to perform intensive lipid analysis easily and efficiently, and will promote the development of lipidomics.

MATERIALS AND METHODS

Workflow

The LipidSig workflow is illustrated in Figure 1 and consists of four parts: (i) Data upload, (ii) Lipid characteristics transformation, (iii) Data processing, and (iv) Functionality and Visualization. Users can explore the platform through demo datasets, or submit their own files to specify lipid expression, lipid characteristics, group assignment, or confounding variables for multivariate regression according to each analysis section. After the required matrices are uploaded, LipidSig provides two analysis pipelines for lipid species and lipid characteristics that can be freely converted between by a transformation function. This built-in function classifies and sums lipid species from 'Lipid expression data' into different categories in 'Lipid characteristics table'. Take Figure 1 as an example, the expression of Lipid1 and Lipid2 in sample1 are 20 and 10, respectively. The two lipid species belong to PC class and yield an expression of 30 for PC. Users are encouraged to employ different data processing strategies in response to data quality and desired methods, making the data suitable for downstream analysis. Missing values due to detection limits can be addressed by missing-value exclusion and by using the imputation options. LipidSig also supports lipid percentage transformation to reduce sample variance. Log transformation and data scaling are offered to improve the performance for further statistical and machine learning analysis. For lipidomic data mining, five powerful functions allow users to profile lipid expression, identify significant lipid features, or construct an informative network. These are useful to analyse lipidomic change at a different resolution levels, and to create concise visualizations for effective interpretation. Also, LipidSig enables users to select computation methods and determine statistical significance through an easy-to-use interface. A short description of the methods and the examples adopted is given below.

Input

Users can use our demo datasets or upload two to four tables for different analysis sections, including 'Lipid expression data', 'Group information'/'Condition table', 'Lipid characteristics table' and 'Multivariate adjusted table'. 'Lipid expression data' encompasses the expression levels for all lipid species in all samples. The data can derive from absolute quantification (picomole) or relative quantification (lipid percent). 'Group information' or 'Condition table' assigns the samples into different groups while 'Lipid characteristics table' records the descriptions of the characteristics for each lipid species. 'Multivariate adjusted table' is used to adjust the confounding effects and only available in the 'Correlation' analysis section. LipidSig accepts comma-separated and tab-separated values (CSV, TSV) formats, and each table should meet basic requirements for minimum sample size and lipid count, column specification, variable type, and cell content. Format is checked as soon as users
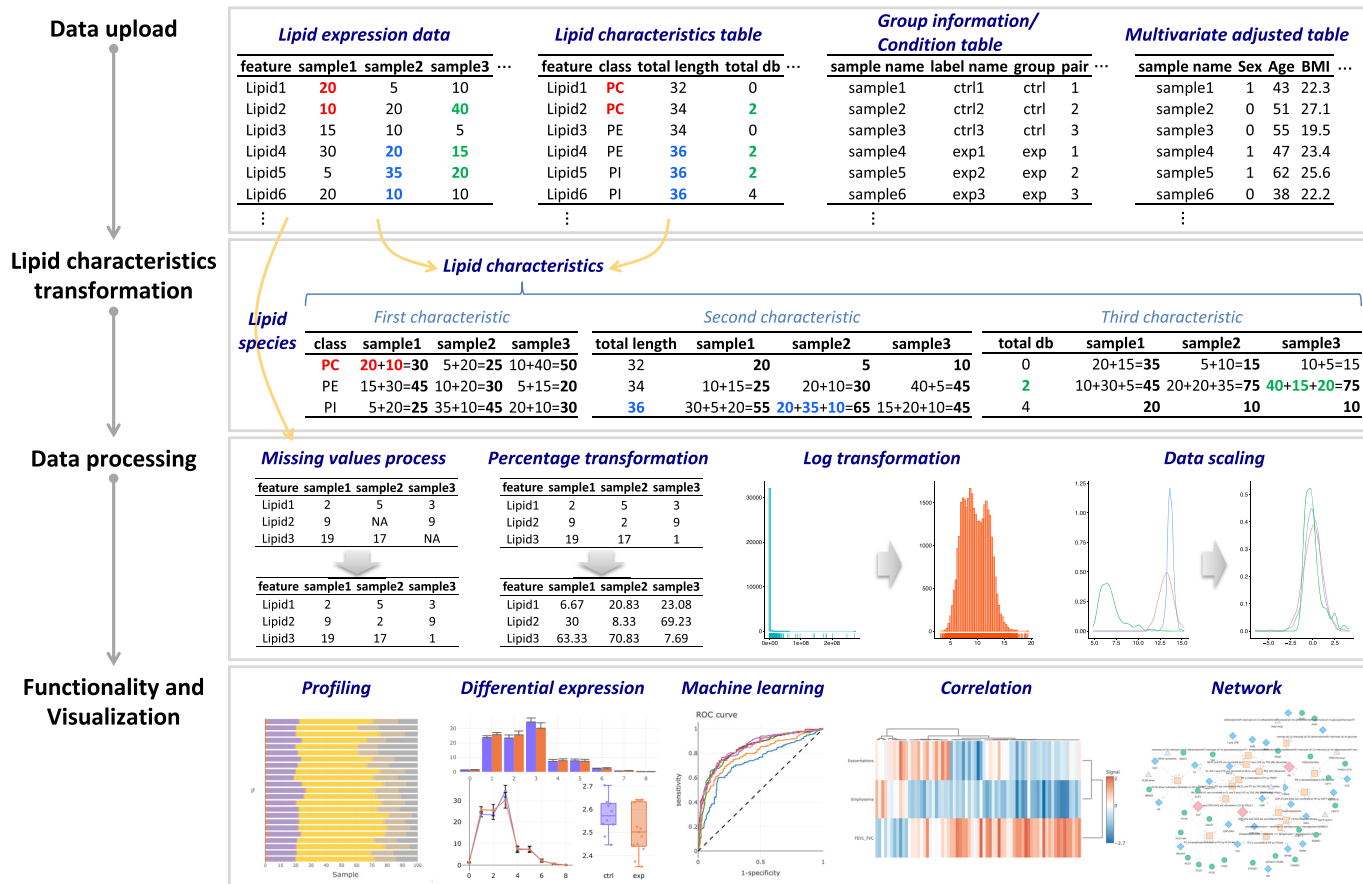
Figure 1. Workflow of the LipidSig web server. The LipidSig workflow is composed of four steps: (i) data upload, (ii) lipid characteristics transformation, (iii) data processing and (iv) functionality and visualization. Four tables at most can be uploaded according to different analysis section. LipidSig provides two analysis pipelines focusing on lipid species or lipid characteristics. The transformation processes between species and characteristics are labelled using different colors. Four data processing strategies enable users to transform and make the data suitable for following analysis. To identify significant lipid features and explore their relationships, five useful functions are offered to comprehensively analyse lipidomic change, and to create effective visualizations.

press the 'Upload' button, and the results will be shown immediately, as the analyses cannot be conducted if files do not meet the standard. Users are encouraged to replicate the formats from the example datasets available on the LipidSig web server.

Lipid characteristics analysis

The massive degree of structural diversity contributes to lipids' functional and characteristic variety. Variations can be subtle (chain length and double bond number) or major (head group and backbone). Lipid Characteristics Analysis evaluates the alterations of a group of lipids categorized by one or more lipid characteristics. LipidSig provides an automatic transformation function to convert the expression of lipid species into lipid characteristics. Firstly, lipid species can be summarized into specific characteristics based on the 'Lipid characteristics table' (Figure 2). For instance, four lipid species in Figure 2A are classified into three categories, lipids with 0, 1 or 6 double bonds. Using this information, the original 'Lipid expression data' (Figure 2B) is then transformed into a new characteristic expression table (Figure 2C). According to user-selected char-

acteristics, the corresponding tables are produced and applied in all analysis sections except 'Network'. Two further functions are derived from the Lipid Characteristics Analysis: Fatty Acid Analysis and Multi-Characteristics Analysis. Fatty Acid Analysis is a special transformation method because it calculates characteristic expression on the basis of fatty acids instead of on whole lipid species. For example, phosphatidylethanolamine (PE) 16:0:0/22:6:1 is a lipid species belonging to PE class with two fatty acids separated by a slash (Figure 2A). The three numbers denote, respectively, the numbers of carbon atoms, double bonds, and hydroxyl groups. Both the total chain length of 38 and the fatty acid chain lengths of 16 and 22 can be supplied via the 'Lipid characteristics table'. These two descriptive conventions offer an overall or partial perspective on lipid characteristic change. When declared in fatty acid format, the summations follow a prior decomposition of the lipid species into fatty acid parts. In Figure 2D, fatty acid chain length in PE has three categories (fatty acids with 16, 20, or 22 carbons) and their expressions in ctrl2 are 0.6 (0.3 + 0.3), 0.3 and 0.3, respectively.

On the other hand, Multi-Characteristics Analysis can be undertaken to explore interactions among numerous char-
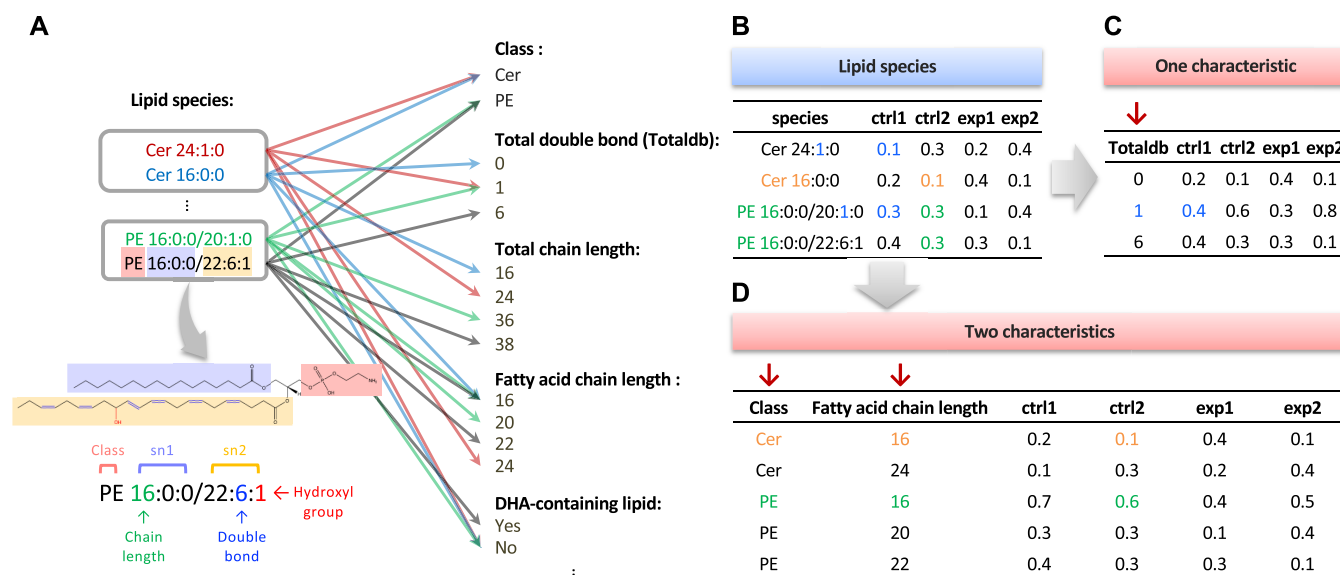
Figure 2. The characteristics transformation function in Lipid Characteristics Analysis. (**A**) A diagram showing how to categorize lipid species into different lipid characteristics. (**B**) Original 'Lipid expression data' uploaded by users. (**C**) A new characteristics expression table for total double bond (Totaldb). (**D**) An expression table combining two characteristics, specific to 'Differential Expression'. Font colors in (**B**) to (**D**) can be used to track the transformation processes between species and characteristics. Cer, ceramdie; PE, phosphatidylethanolamine.

acteristics. In 'Differential Expression', users may split the data using one particular characteristic before performing computations based on another characteristic. As an example, it is able to analyze fatty acid chain length in ceramide (Cer) or PE class specifically (Figure 2D). It is also possible to combine more than two characteristics to build a prediction model in 'Machine Learning'. Multiple characteristics tables are formed and used as predictor variables to provide additional structural or functional information of lipidomic data. The important characteristics will be selected and ranked in the resulting model.

MAIN FUNCTIONS OF LipidSig

LipidSig offers five main functions, namely 'Profiling', 'Differential Expression', 'Correlation', 'Network' and 'Machine Learning' for assessing lipid effects on biological mechanisms. In 'Profiling', an overview of comprehensive analyses allows researchers to efficiently examine data quality, clustering of samples, correlation between lipid species, and composition of lipid characteristics. 'Differential Expression' integrates many useful lipid-focused analyses, assisting users to identify significant lipid species or lipid characteristics. 'Correlation' analysis is designed to illustrate and compare the relationships between different clinical phenotypes and lipid features. The 'Network' function constructs (i) lipid metabolism pathways based on the Reactome database and (ii) a pathway enrichment network queried with lipid-related genes. 'Machine learning' provides a broad variety of feature selection methods and classifiers to build binary classification models. Subsequent analyses then help users to evaluate the learning algorithm's performance and to explore important lipid-related variables. In general, except 'Network', which is generated

based on the human database, all analysis sections make use of uploaded lipid expression and lipid characteristics data to compute the results. Hence, there is no limitation to the source of lipidomic data in most analyses. Detailed documentation for each section is given in the Supplementary Methods.

EXAMPLES

Discovery of novel driver lipids for ferroptosis in cancer cells

Ferroptosis is iron-dependent, non-apoptotic, oxidative cell death that is implicated in various diseases including neurodegeneration, ischemic injury in many organs, and cancer resistance (28,29). Ferroptosis is closely related to polyunsaturated phospholipid peroxidation but the underlying mechanism, for example whether it happens on global or specific lipids, is poorly understood. Here, we utilized Zou, Yilong et al.'s dataset as an example (30). The expression table for this dataset contains six OVCAR-8 cell samples and 202 lipid species. The species are categorized into different lipid classes in the lipid characteristics table. Through genome-wide CRISPR–Cas9 screens, they found alkylglycerone phosphate synthase (AGPS) was a critical mediator in driving ferroptosis susceptibility. Thus, we separated the samples into control and AGPS knockout groups. In 'Profiling', OVCAR-8 cells (*n* = 3) expressing control sgRNAs (sgNC) or sgRNAs targeting AGPS (sgAGPS) can be perfectly distinguished using PCA on lipidomic data (Figure 3A). In addition, our characteristics profiling reveals an obvious change in lipid class composition, especially the reduction of ether-linked PC (PC O–) and PE (PE O–) in the AGPS knockout group (Supplementary Figure S1A). Significant lipid species were then further explored using differential expression analysis. In Figure 3B, the volcano
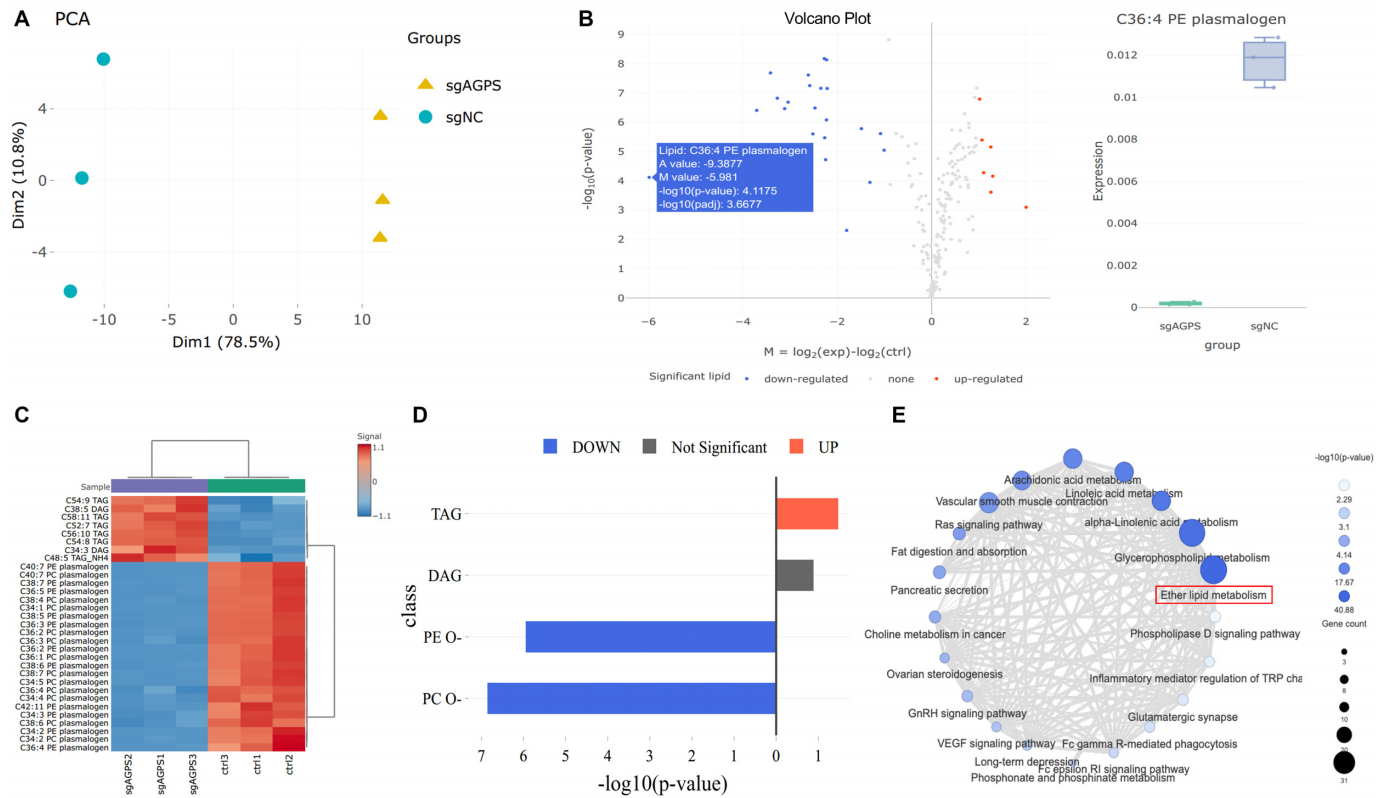
Figure 3. An example of LipidSig being used to identify critical lipids driving ferroptosis in OVCAR-8 cells. (**A**) The PCA plot of lipidome in the OVCAR-8 cells expressing control sgRNAs (sgNC) and sgRNAs targeting AGPS (sgAGPS). (**B**) An interactive volcano plot showing the differentially expressed lipid species of OVCAR-8 cells expressing sgNC or sgAGPS. $n = 3$ biological replicates. Two-tailed Student's t-tests with Benjamini–Hochberg correction method were used to calculate the p-values. (**C**) Heatmap of hierarchical clustering for significant lipid species ($P < 0.05$) with sample group labels in the top. (**D**) Enrichment analysis of significant lipid species was performed using over representation analysis based on lipid class. Bars indicate $-\log_{10}(P\text{-value})$. (**E**) Enrichment network built from KEGG pathway analysis presents the significantly altered pathways ($P < 0.05$) associated with PC O– and PE O– related genes. Nodes are filled according to $-\log_{10}(P\text{-value})$ and their sizes represent the lipid-related gene number involved in the pathway. Line width indicates the value of gene similarity between the pathways. PC O-, ether-linked phosphatidylcholine; PE O–, ether-linked phosphatidylethanolamine.

plot indicates statistical significance versus magnitude of fold change for each lipid species. Still more information can be dug into using LipidSig. We obtained detailed statistical information, for example, as well as characteristic distributions to emphasize the importance of ether lipids (Supplementary Figure S1B and S1C). Likewise, a hierarchical clustering (Figure 3C) and an enrichment using over representation analysis (Figure 3D) clearly revealed that significantly down-regulated lipid species were enriched in the PC O– and PE O– categories. The expression of different lipid classes can be compared through Lipid Characteristics Analysis, and used to corroborate the results from the preceding evaluations (Supplementary Figure S1D). Besides, ether lipid metabolism was successfully ranked as the top hit in KEGG pathway enrichment analysis, queried with PC O– and PE O– related genes (Figure 3E). These data indicated that depletion of AGPS remodeled the intracellular lipidome, especially the biosynthesis of ether lipids, which might be critical to the pro-ferroptotic status. Taken together, LipidSig can not only identify differential expressed lipid species but also offer more practical functions to associate them with specific lipid characteristics and potential metabolic pathways.

Connecting lipid characteristics with biological function

Lipid Characteristics Analysis is one of the important features in LipidSig. It can be carried out using our transformation function to build a new characteristics expression table for further analysis (Figure 2). In 'Differential Expression', users are allowed to focus on one specific characteristic, or a combination of two characteristics, to access the cellular phenotypes or disease associations. We collected two examples to demonstrate this application in lipid unsaturation (double bond) and chain-length analysis. First case is to study mammalian membrane homeostasis in response to dietary lipid perturbations (31). The expression table for the dataset includes 390 lipid species in six mice fed with fish oil or corn oil. All lipid species are categorized according to the number of total double bond, fatty acid double bond, fatty acid chain length in the lipid characteristics table. Using LipidSig, we obtained an almost same lipid unsaturation profile in the paper. The fish oil (FO) diet, enriched with eicosapentaenoic acid (EPA; 20:5) and docosahexaenoic acid (DHA; 22:6), led to a striking increase of glycerophospholipids (GPLs) with six double bonds in mouse cardiac tissue compared to the corn oil (CO) diet (Figure 4A). This response can be compensated by the induction of saturated
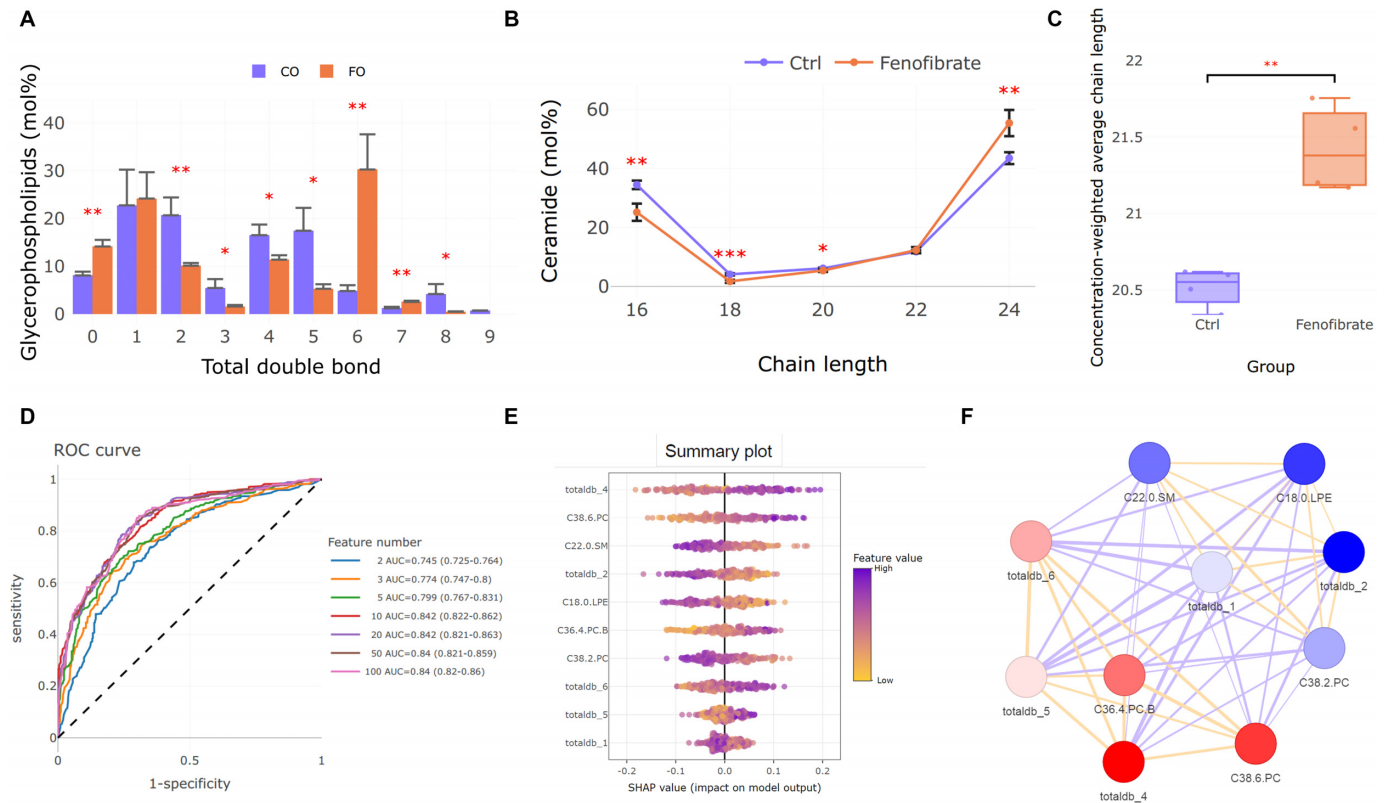
Figure 4. Applications of lipid characteristics analysis. (**A**) Lipid unsaturation (double bond) profile in membrane glycerophospholipids (GPLs) isolated from mouse cardiac tissue treated with corn oil (CO) diet versus fish oil (FO) diet. $n = 3$ biological replicates. (**B**) Lipid chain trend plot in ceramide (Cer) from murine pancreatic tissue after feeding with fenofibrate. $n = 4$ biological replicates. (**C**) The concentration-weighted average chain length of Cer is significantly up-regulated upon fenofibrate treatment. (**D**) Receiver operating characteristic (ROC) curve and area under curve (AUC) for machine learning models with different number of features. (**E**) Feature importance of the top 10 contributing features under SHAP (SHapley Additive exPlanations) analysis in the best ten-feature model. (**F**) Pearson correlation network of the lipid predictors in the best ten-feature model. Nodes are filled according to SHAP feature importance and their colors indicate direction of impact. Line width indicates the correlation coefficients, with purple for negative correlation and orange for positive correlation. Two-tailed Student's *t*-tests were used to calculate the p-values in (A) to (C). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. totaldb, total double bond.

GPLs and the reduction of other polyunsaturated GPLs (2–5 double bonds). Using the Fatty Acid Analysis, we found an expected increase in fatty acids with 22 carbons and six double bonds, corresponding to DHA (22:6) (Supplementary Figure S2A and B). Due to LipidSig's flexible handling of user-defined lipid characteristics, a significant incorporation of DHA into the membrane can also be seen in the FO-fed group (Supplementary Figure S2C).

The second example examines the effects of fenofibrate, a potential drug for type-1 diabetes, on the mouse pancreatic lipidome (32). One hundred ninety-six lipid species with lipid class and chain length assignment are used to analyze the lipidomic alteration in the pancreas from four control and four fenofibrate-treated mice. Our Multi-Characteristics Analysis provides a summary table that clearly highlights significant chain length categories in the Cer class (Supplementary Figure S2D). On the trend plot, a chain length shift from C16 to C24 Cer is uncovered in the fenofibrate-treated group, as previously described (Figure 4B). C16 Cer is associated with apoptosis and insulin resistance (33,34), which might explain fenofibrate's anti-diabetic effect. Besides, we calculate the concentration-

weighted average index to reflect an overall change in chain length (31). In brief, 'Lipid expression data' is transformed into a new expression table for chain length. Then, each chain length was multiplied by its proportion and summed up to get the index. In LipidSig, an increased concentration-weighted average chain-length index in Cer can also be detected after feeding with fenofibrate (Figure 4C). Overall, our Lipid Characteristics Analysis in 'Differential Expression' unravels the impact of one or two characteristics and links their changes to different biological functions.

Lipid-related biomarker exploration

Recently, distinct dysregulation of lipid metabolism has been discovered in many diseases (8–11). Lipidome profiling can provide new insights into specific signatures, making them compelling biomarkers or candidates for investigating causal effects. Here, we introduce two datasets, a lipidomic study for chronic obstructive pulmonary disease (COPD) (35) and an association analysis linking cancer metabolomic alterations to genetic features (36), to illustrate our 'Corre-

lation' and 'Machine Learning' functions as biomarker explorers.

In the COPD dataset, a multivariate linear regression model is applied to associate sphingolipids with clinical subphenotypes of COPD. The dataset detects 69 distinct plasma sphingolipid species in 129 current and former smokers. Three COPD subphenotypes including exacerbations, emphysema, and FEV1/FVC (ratio of the forced expiratory volume during the first second to the forced vital capacity) are recorded in the condition table. Additionally, confounding factors such as age, sex, smoking status, body mass index, and FEV1 are provided in the multivariate-adjusted table. We then build a heatmap to easily compare the lipidomic fingerprints identified in different groups (Supplementary Figure S3A). The most significant negative association is between emphysema and sphingomyelins, consistent with the paper's result. In the other cancer metabolomic research, cell lines with high polyunsaturated lipids have been found more resistant to the knockout of stearoyl-CoA Desaturase (SCD), a key enzyme in monounsaturated fatty acid formation. Therefore, we now aim at predicting cancer cell lines' sensitivity to SCD knockout using the provided lipidomic data, and corroborating the key unsaturation feature by machine-learning modeling. We extract 89 lipid species from the original metabolite data and build the lipid characteristics table containing the information about lipid class, double bond number, and chain length. 228 cancer cell lines in the first and last quarter according to sensitivity to SCD knockout are selected for further analysis. Lipid species and multiple characteristics (including lipid class, chain length, and double bond) are introduced to separate two groups—sensitive (labelled as 0) or resistant (labelled as 1) cell lines. We apportion the data into training and testing sets, with a ratio of 2:1. A random forest model is adopted to perform feature selection, model training, and performance evaluation using 10 times Monte-Carlo cross-validation. The receiver operating characteristic (ROC) and precision-recall (PR) curves reveal that ten features reached a plateau in model performance, which is supported by the elbow point in the accuracy curve (Figure 4D, Supplemental Figure S3B and S3C). A further feature importance analysis by SHAP (SHapley Additive exPlanations) points out that unsaturation-related variables (lipids with 1, 2, 4, 5 and 6 double bonds, totaldb) account for half in the ten-feature model (Figure 4E). The directions of feature values and Shapley values in the SHAP summary plot show that lipids with 4, 5 and 6 double bonds contributed to SCD knockout resistance, while lipids with one and two double bonds increased the sensitivity. Furthermore, the correlation network of the predictors displays two positive correlation clusters containing polyunsaturated and less-unsaturated (double bond) lipid features, respectively (Figure 4F). In conclusion, 'Machine Learning' allows users to introduce multiple lipid characteristics as feature variables and we highlight the effect of lipid unsaturation on cell lines' sensitivity to SCD knockout. To summarize, this analyses illustrate the feasibility of using the 'Correlation' and 'Machine Learning' functions to discover novel lipid biomarkers, made up of species, characteristics, or both.

DISCUSSION

LipidSig is the first web server dedicated to intensive lipid-focused analyses and giving detailed insights into the changes of lipid species and characteristics. It supports robust data processing, data mining, and visualization to functionally interpret complex lipidomic data. We provide a comprehensive analysis workflow to identify significant lipid features in different cellular or clinical phenotypes (Figure 3 and Supplemental Figure S1). Our web server also places great importance on Lipid Characteristics Analysis, designed to investigate the changes of diverse characteristics in different experimental groups. Most existing tools with similar functions require users to follow specific naming rules, for example 'PC 36:2' meaning a lipid species belonging to PC class with chain length of 36, and two double bonds. Then, they extract characteristics information from this format. To the contrary, LipidSig allows users to upload a characteristics table that defines characteristics for each lipid species by themselves. The customary format limitation is thus mitigated, allowing the handling of more lipid characteristics, class, shape (cone, cylinder), hydroxyl group, chain length, double bond, fatty acid composition, etc. These characteristics are unique for lipids and usually correspond to specific biological functions. Our transformation function can convert expression data from lipid species to lipid characteristics, and can be applied in many analysis sections. In the ferroptosis case study, Lipid Characteristics Analysis in 'Profiling' and 'Differential Expression' revealed a clear decrease of the PC O– and PE O– classes in the AGPS knockout group (Supplementary Figure S1A and S1D). The change in lipid class is considered a good indicator of altered lipid metabolism pathways. The enrichment result in 'Network' supported this point, indicating that ether lipid metabolism might be affected (Figure 3E).

Moreover, Lipid Characteristics Analysis also examines the changes of double bond, chain length, and hydroxyl group that constitute fatty acid diversity. In Figure 4B and C, we demonstrated the shift of double bond and chain length between the control and the treatment group. Double bond change represents membrane unsaturation degree or fluidity, and has been connected to signaling transduction (16,17). Lipid chain length has been thought to be a measure of fatty acid elongation or oxidation (18,37), but recent findings have revealed that changes in chain length (C16 up and C24 down in ceramide) could lead to insulin resistance in obesity, unravelling a new role for specific chain length (Figures 4C) (32,38). Further, owing to the flexibility of characteristics definition in LipidSig, users are able to observe changes in a specific lipid group according to their own experimental requirements. In Supplemental Figures S2C, whether lipids contain DHA was recorded in the 'Lipid characteristics table' and calculated. In addition, the accumulation of lipid hydroperoxides on membrane is recognized as the hallmark of ferroptosis (29). A redox-lipidomics analysis can also be performed to evaluate alterations of hydroxyl and hydroperoxyl lipids as ferroptotic signals (Supplemental Figures S2E) (19). It's noted that Lipid Characteristics Analysis uses a direct add-up function for selected characteristics regardless of data source.

Although this method has been widely used in many studies (16–18,31,32), users should consider whether their data are suitable for this process when analyzing certain lipid characteristics.

Two unique functions, Fatty Acid Analysis and Multi-Characteristics Analysis, derived from Lipid Characteristics Analysis can extend more applications in LipidSig. The first of these provides an analysis focusing on fatty acid instead of lipid species in 'Differential Expression'. This allows users to examine different types of polyunsaturated fatty acids (PUFAs) that are stored in lipids, and that exert special functions. For instance, compared to calculation for total double bonds in lipid species, an increase in fatty acids with 6 double bonds was more reasonable after DHA (22:6) supplementation (Figure 4A and Supplemental Figures S2A). PUFA search through Fatty Acid Analysis can be also used to explore substrate specificity of acyltransferases (39). One example is to study arachidonic acid (AA; 20:4) incorporation by LPCAT3 in intestine and liver cells, which promotes lipoprotein assembly (40). As for Multi-Characteristics Analysis, that can be achieved in 'Differential Expression' and 'Machine Learning'. In the fenofibrate-treated case, we observed a major change in chain length in the ceramide class (Figure 4B) (32). Another study showed that membrane saturation remodeling by LPCAT1, a lysophosphatidylcholine acyltransferase, happened in the PC class (17). Newly, a similar approach has also been implemented to explore the unique pattern of chain length and desaturation in different lipid classes by probing the mitochondrial lipidome of adipose tissues (41). Thus, a two-characteristics analysis is suited to studies that concentrate on specific reactions or pathways. For working with more than two characteristics, users are encouraged to build different machine learning models to evaluate the importance of lipid-related variables. In Figure 4D–F, we demonstrated how the addition of various lipid characteristics promotes structural or functional resolution of the lipidome and links to specific phenotypes. This method is useful for processing multidimensional lipidomic data, and for discovering compelling biomarkers that can be used in clinical research. For example, there have been three large population cohorts combining machine learning and lipidome to obtain new insights into complex disease processes including coronary artery disease (8), obesity (9), and type 2 diabetes (10). Moreover, LipidSig users can also assess Human Metabolome Database (HMDB) database to connect important lipid predictors with related diseases (42).

Additionally, we developed two practical functions, 'Correlation' and 'Network', to support further analysis. 'Correlation' offers a clear heatmap to compare significant lipid features for diverse clinical phenotypes (Supplemental Figures S3A). Lipid Characteristics Analysis can also be performed here to assess the associations of a specific group of lipids. Similar methods have been used to interrogate causal effects of metabolites in nonalcoholic fatty liver disease (NAFLD) (43) and acute-on-chronic liver failure (ACLF) (44). The 'Network' page is divided into two parts. The first of these is built from the Reactome database; its purpose is to find connections between different lipids and their regulatory genes. Recently, a web-based tool called BioPAN has also been proposed to explore lipidome metabolic pathways (45). Different from our Reactome network, BioPAN requires users to upload lipid expression data and introduces a statistical model to identify activated or suppressed lipid pathways based on a ratio of product over reactant. In LipidSig, we can provide an overview of the metabolic network and guide users to explore the critical reaction steps. On the other hand, BioPAN puts more emphasis on the concept of lipid flux and also lists predicted genes, which could be involved in the reactions. The other part, the pathway enrichment network, presents significantly altered pathways based on related genes of selected lipids. This multi-omics approach helped us find ether lipid metabolism in Figure 3E and was recently adopted to identify a metabolic signature in high-grade bladder cancer (46).

In summary, we propose LipidSig as a highly flexible web server that can conduct comprehensive and deep lipidomic analysis. A strong emphasis has been placed on the functionality of Lipid Characteristics Analysis, exemplified by many published works using a wide range of datasets. Our main hope is that LipidSig will bridge fundamental aspects of lipid biology in order to contribute to the broader development of lipidomics. Currently, there are more and more interesting lipidomic datasets focusing on specific diseases (8–10) or subcellular organelles available (41,47,48). These datasets can help users grasp the essence of lipidomic analysis workflow in LipidSig. We will include them into the demo datasets for our next update.

DATA AVAILABILITY

LipidSig is freely and publicly available at http://chenglab. cmu.edu.tw/lipidsig. No login is required and detailed documentation is provided on the LipidSig website.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Harayama,T. and Riezman,H. (2018) Understanding the diversity of membrane lipid composition. *Nat. Rev. Mol. Cell Biol.*, **19**, 281–296.

2. Shimizu,T. (2009) Lipid mediators in health and disease: enzymes and receptors as therapeutic targets for the regulation of immunity and inflammation. *Annu. Rev. Pharmacol. Toxicol.*, **49**, 123–150.

3. Resh,M.D. (2016) Fatty acylation of proteins: the long and the short of it. *Prog. Lipid Res.*, **63**, 120–131.

4. Yang,W.S. and Stockwell,B.R.J.T.i.c.b. (2016) Ferroptosis: death by lipid peroxidation. **26**, 165–176.

5. Sezgin,E., Levental,I., Mayor,S. and Eggeling,C. (2017) The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat. Rev. Mol. Cell Biol.*, **18**, 361–374.

6. Nilsson-Ehle,P., Garfinkel,A.S. and Schotz,M.C.J.A.r.o.b. (1980) Lipolytic enzymes and plasma lipoprotein metabolism. **49**, 667–693.

7. Walther,T.C. and Farese,R.V. Jr (2012) Lipid droplets and cellular lipid metabolism. *Annu. Rev. Biochem.*, **81**, 687–714.

8. Poss,A.M., Maschek,J.A., Cox,J.E., Hauner,B.J., Hopkins,P.N., Hunt,S.C., Holland,W.L., Summers,S.A. and Playdon,M.C. (2020) Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease. *J. Clin. Invest.*, **130**, 1363–1376.

9. Gerl,M.J., Klose,C., Surma,M.A., Fernandez,C., Melander,O., Männistö,S., Borodulin,K., Havulinna,A.S., Salomaa,V., Ikonen,E. *et al.* (2019) Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. *PLoS Biol.*, **17**, e3000443.

10. Fernandez,C., Surma,M.A., Klose,C., Gerl,M.J., Ottosson,F., Ericson,U., Oskolkov,N., Ohro-Melander,M., Simons,K. and Melander,O. (2019) Plasma lipidome and prediction of type 2 diabetes in the population-based Malmö diet and cancer cohort. dc191199.

11. Corn,K.C., Windham,M.A. and Rafat,M. (2020) Lipids in the tumor microenvironment: from cancer progression to treatment. *Prog. Lipid Res.*, **80**, 101055.

12. Blanksby,S.J. and Mitchell,T.W. (2010) Advances in mass spectrometry for lipidomics. *Annu. Rev. Anal. Chem.*, **3**, 433–465.

13. Wang,M., Wang,C., Han,R.H. and Han,X.J.P.i.l.r. (2016) Novel advances in shotgun lipidomics for biology and medicine. **61**, 83–108.

14. Yang,K. and Han,X. (2016) Lipidomics: techniques, applications, and outcomes related to biomedical sciences. *Trends Biochem. Sci.*, **41**, 954–969.

15. de Anda-Jáuregui,G. and Hernández-Lemus,E. (2020) Computational oncology in the multi-omics era: state of the art. *Front. Oncol.*, **10**, 423–423.

16. Levental,K.R., Surma,M.A., Skinkle,A.D., Lorent,J.H., Zhou,Y., Klose,C., Chang,J.T., Hancock,J.F. and Levental,I. (2017) ω-3 polyunsaturated fatty acids direct differentiation of the membrane phenotype in mesenchymal stem cells to potentiate osteogenesis. *Sci. Adv.*, **3**, eaao1193.

17. Bi,J., Ichu,T.-A., Zanca,C., Yang,H., Zhang,W., Gu,Y., Chowdhry,S., Reed,A., Ikegami,S., Turner,K.M. *et al.* (2019) Oncogene amplification in growth factor signaling pathways renders cancers dependent on membrane lipid remodeling. *Cell Metab.*, **30**, 525–538.

18. Purdy,J.G., Shenk,T. and Rabinowitz,J.D. (2015) Fatty acid elongase 7 catalyzes lipidome remodeling essential for human cytomegalovirus replication. *Cell Rep.*, **10**, 1375–1385.

19. Zou,Y., Li,H., Graham,E.T., Deik,A.A., Eaton,J.K., Wang,W., Sandoval-Gomez,G., Clish,C.B., Doench,J.G. and Schreiber,S.L. (2020) Cytochrome P450 oxidoreductase contributes to phospholipid peroxidation in ferroptosis. *Nat. Chem. Biol.*, **16**, 302–309.

20. Kagan,V.E., Mao,G., Qu,F., Angeli,J.P.F., Doll,S., Croix,C.S., Dar,H.H., Liu,B., Tyurin,V.A., Ritov,V.B. *et al.* (2017) Oxidized arachidonic and adrenic PEs navigate cells to ferroptosis. *Nat. Chem. Biol.*, **13**, 81–90.

21. Mohamed,A., Molendijk,J. and Hill,M.M. (2020) lipidr: a software tool for data mining and analysis of lipidomics datasets. *J. Proteome Res.*, **19**, 2890–2897.

22. Hartler,J., Trötzmüller,M., Chitraju,C., Spener,F., Köfeler,H.C. and Thallinger,G.G. (2011) Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data. *Bioinformatics*, **27**, 572–577.

23. Kind,T., Liu,K.-H., Lee,D.Y., DeFelice,B., Meissen,J.K. and Fiehn,O. (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods*, **10**, 755–758.

24. Ni,Z., Angelidou,G., Lange,M., Hoffmann,R. and Fedorova,M. (2017) LipidHunter identifies phospholipids by high-throughput processing of LC-MS and shotgun lipidomics datasets. *Anal. Chem.*, **89**, 8800–8807.

25. Koelmel,J.P., Kroeger,N.M., Ulmer,C.Z., Bowden,J.A., Patterson,R.E., Cochran,J.A., Beecher,C.W.W., Garrett,T.J. and Yost,R.A. (2017) LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinformatics*, **18**, 331.

26. Husen,P., Tarasov,K., Katafiasz,M., Sokol,E., Vogt,J., Baumgart,J., Nitsch,R., Ekroos,K. and Ejsing,C.S. (2013) Analysis of lipid experiments (ALEX): a software framework for analysis of high-resolution shotgun lipidomics data. *PLoS One*, **8**, e79736.

27. Clair,G., Reehl,S., Stratton,K.G., Monroe,M.E., Tfaily,M.M., Ansong,C. and Kyle,J.E. (2019) Lipid Mini-On: mining and ontology tool for enrichment analysis of lipidomic data. *Bioinformatics*, **35**, 4507–4508.

28. Hangauer,M.J., Viswanathan,V.S., Ryan,M.J., Bole,D., Eaton,J.K., Matov,A., Galeas,J., Dhruv,H.D., Berens,M.E., Schreiber,S.L. *et al.* (2017) Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature*, **551**, 247–250.

29. Stockwell,B.R., Friedmann Angeli,J.P., Bayir,H., Bush,A.I., Conrad,M., Dixon,S.J., Fulda,S., Gascón,S., Hatzios,S.K., Kagan,V.E. *et al.* (2017) Ferroptosis: a regulated cell death nexus linking metabolism, redox biology, and disease. *Cell*, **171**, 273–285.

30. Zou,Y., Henry,W.S., Ricq,E.L., Graham,E.T., Phadnis,V.V., Maretich,P., Paradkar,S., Boehnke,N., Deik,A.A., Reinhardt,F. *et al.* (2020) Plasticity of ether lipids promotes ferroptosis susceptibility and evasion. *Nature*, **585**, 603–608.

31. Levental,K.R., Malmberg,E., Symons,J.L., Fan,Y.-Y., Chapkin,R.S., Ernst,R. and Levental,I. (2020) Lipidomic and biophysical homeostasis of mammalian membranes counteracts dietary lipid perturbations to maintain cellular fitness. *Nat. Commun.*, **11**, 1339.

32. Holm,L.J., Haupt-Jorgensen,M., Giacobini,J.D., Hasselby,J.P., Bilgin,M. and Buschard,K.J.D. (2019) Fenofibrate increases very-long-chain sphingolipids and improves blood glucose homeostasis in NOD mice. **62**, 2262–2272.

33. Grösch,S., Schiffmann,S. and Geisslinger,G. (2012) Chain length-specific properties of ceramides. *Prog. Lipid Res.*, **51**, 50–62.

34. Raichur,S., Wang,S.T., Chan,P.W., Li,Y., Ching,J., Chaurasia,B., Dogra,S., Öhman,M.K., Takeda,K., Sugii,S. *et al.* (2014) CerS2 haploinsufficiency inhibits β-oxidation and confers susceptibility to diet-induced steatohepatitis and insulin resistance. *Cell Metab.*, **20**, 687–695.

35. Bowler,R.P., Jacobson,S., Cruickshank,C., Hughes,G.J., Siska,C., Ory,D.S., Petrache,I., Schaffer,J.E., Reisdorph,N. and Kechris,K. (2015) Plasma sphingolipids associated with chronic obstructive pulmonary disease phenotypes. *Am. J. Respir. Crit. Care Med.*, **191**, 275–284.

36. Li,H., Ning,S., Ghandi,M., Kryukov,G.V., Gopal,S., Deik,A., Souza,A., Pierce,K., Keskula,P., Hernandez,D. *et al.* (2019) The landscape of cancer cell line metabolism. *Nat. Med.*, **25**, 850–860.

37. Dávalos-Salas,M., Montgomery,M.K., Reehorst,C.M., Nightingale,R., Ng,I., Anderton,H., Al-Obaidi,S., Lesmana,A., Scott,C.M., Ioannidis,P. *et al.* (2019) Deletion of intestinal Hdac3 remodels the lipidome of enterocytes and protects mice from diet-induced obesity. *Nat. Commun.*, **10**, 5291.

38. Turpin,S.M., Nicholls,H.T., Willmes,D.M., Mourier,A., Brodesser,S., Wunderlich,C.M., Mauer,J., Xu,E., Hammerschmidt,P., Brönneke,H.S. *et al.* (2014) Obesity-induced CerS6-dependent C16:0 ceramide production promotes weight gain and glucose intolerance. *Cell Metab.*, **20**, 678–686.

39. Yamashita,A., Hayashi,Y., Nemoto-Sasaki,Y., Ito,M., Oka,S., Tanikawa,T., Waku,K. and Sugiura,T. (2014) Acyltransferases and transacylases that determine the fatty acid composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian cells and model organisms. *Prog. Lipid Res.*, **53**, 18–81.

40. Hashidate-Yoshida,T., Harayama,T., Hishikawa,D., Morimoto,R., Hamano,F., Tokuoka,S.M., Eto,M., Tamura-Nakano,M., Yanobu-Takanashi,R., Mukumoto,Y. *et al.* (2015) Fatty acid remodeling by LPCAT3 enriches arachidonate in phospholipid membranes and regulates triglyceride transport. *eLife*, **4**, e06328

41. Rajakumari,S. and Srivastava,S. (2021) Aging and β3-adrenergic stimulation alter mitochondrial lipidome of adipose tissue. *Biochim. Biophys. Acta. Mol. Cell Biol. Lipids*, **1866**, 158922.

42. Wishart,D.S., Feunang,Y.D., Marcu,A., Guo,A.C., Liang,K., Vázquez-Fresno,R., Sajed,T., Johnson,D., Li,C., Karu,N. *et al.* (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic. Acids. Res.*, **46**, D608–D617.

43. Gorden,D.L., Myers,D.S. Ivanova,P.T., Fahy,E., Maurya,M.R., Gupta,S., Min,J., Spann,N.J., McDonald,J.G., Kelly,S.L. *et al.* (2015) Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J. Lipid Res.*, **56**, 722–736.

44. Moreau,R., Clària,J., Aguilar,F., Fenaille,F., Lozano,J.J., Junot,C., Colsch,B., Caraceni,P., Trebicka,J., Pavesi,M. *et al.* (2020) Blood metabolomics uncovers inflammation-associated mitochondrial dysfunction as a potential mechanism underlying ACLF. *J. Hepatol.*, **72**, 688–701.

45. Gaud,C., B,C.S., Nguyen,A., Fedorova,M., Ni,Z., O'Donnell,V.B., Wakelam,M.J.O., Andrews,S. and Lopez-Clavijo,A.F. (2021)

BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS. *F1000Research*, **10**, 4.

46. Vantaku,V., Dong,J., Ambati,C.R., Perera,D., Donepudi,S.R., Amara,C.S., Putluri,V., Ravi,S.S., Robertson,M.J., Piyarathna,D.W.B. *et al.* (2019) Multi-omics integration analysis robustly predicts high-grade patient survival and identifies CPT1B effect on fatty acid metabolism in bladder cancer. *Clin. Cancer Res.*, **25**, 3689–3701.

47. Pollard,A.K., Ingram,T.L., Ortori,C.A., Shephard,F., Brown,M., Liddell,S., Barrett,D.A. and Chakrabarti,L. (2019) A comparison of the mitochondrial proteome and lipidome in the mouse and long-lived Pipistrelle bats. *Aging*, **11**, 1664–1685.

48. Andreyev,A.Y., Fahy,E., Guan,Z., Kelly,S., Li,X., McDonald,J.G., Milne,S., Myers,D., Park,H., Ryan,A. *et al.* (2010) Subcellular organelle lipidomics in TLR-4-activated macrophages. *J. Lipid Res.*, **51**, 2785–2797.