

1 **iLipidome: hidden biosynthetic interdependencies in the lipidome enhance**
2 **statistical power and interpretability**

3

4 Wen-Jen Lin^{1,2,+}, Austin W.T. Chiang^{1,+,*}, Chenguang Liang³, Wen-Lung Ma^{2,4,5}, Wei-Chung
5 Cheng^{2,6,7}, Nathan E. Lewis^{1,3,*}

¹Department of Pediatrics, University of California, San Diego, CA 92093, USA

²Graduate Institute of Biomedical Science, China Medical University, Taichung 404333, Taiwan

³Department of Bioengineering, University of California, San Diego, CA 92093, USA

⁴Sex Hormone Research Center, China Medical University Hospital, Taichung 40403, Taiwan

⁵Department of Nursing, Asia University, Taichung 413305, Taiwan

⁶Research Center for Cancer Biology, China Medical University, Taichung 404333, Taiwan

⁷Ph.D. program for Cancer Biology and Drug Discovery, China Medical University and Academia Sinica, Taichung 404333, Taiwan

6 ⁺ These authors contributed equally to this work

7

8

*Corresponding authors:

9 Name: Nathan E. Lewis

10 Address: 9500 Gilman Drive MC 0760, La Jolla, CA 92093

E-mail: nlewisres@ucsd.edu

11 [Name: Austin W.T. Chiang](#)

12 [Address: 9500 Gilman Drive MC 0760, La Jolla, CA 92093](#)

[E-mail: austin.chiang@gmail.com](mailto:austin.chiang@gmail.com)

13 **Short running title:** Integrating interdependence into lipidome analysis

14

15

Abstract

Numerous biological processes and diseases are influenced by lipid composition. Advances in lipidomics are elucidating their roles, but analyzing and interpreting lipidomics data at the systems level remain challenging. To address this, we present iLipidome, a method for analyzing lipidomics data in the context of the lipid biosynthetic network, thus accounting for the interdependence of measured lipids. iLipidome enhances statistical power, enables reliable clustering and lipid enrichment analysis, and links lipidomic changes to their genetic origins. We applied iLipidome to investigate mechanisms driving changes in cellular lipidomes following supplementation of docosahexaenoic acid (DHA) and successfully identified the genetic causes of alterations. We further demonstrated how iLipidome can disclose enzyme-substrate specificity and pinpoint prospective glioblastoma therapeutic targets. Finally, iLipidome enabled us to explore underlying mechanisms of cardiovascular disease and could guide the discovery of early lipid biomarkers. Thus, iLipidome can assist researchers studying the essence of lipidomic data and advance the field of lipid biology.

Main Text

Introduction

Lipids are an essential energy reservoir, important modulators of signaling, and fundamental building blocks of cellular membranes. Advances in mass spectrometry (MS) can link lipids to phenotypes¹⁻⁴. Lipidomics analyses often focus on a handful of lipids showing the largest changes in abundance for biological insight⁵⁻⁷. However, there are several limitations to this approach. First, after correcting for multiple hypothesis testing, individual lipids might not meet statistical significance due to the biological variability across samples or the noise generated by lipidomic technologies⁸. Second, in other cases, one may obtain a long list of statistically significant lipids without a clear unifying biological theme; thus *ad hoc* interpretation is required. Third, focusing on individual lipids may miss crucial concerted changes shared across whole pathways. Fourth, since most conventional lipidomics analyses overlook the interconnections among measured lipids, it remains challenging to identify the genetic source of lipidomic changes. For these difficulties, analysis of the data in the context of the lipid molecular network could be a promising solution.

Essentially, a lipid molecular network encoding biosynthetic information could provide a comprehensive understanding of the complex metabolic processes and enable us to explore the perturbed pathways and reactions behind lipid profiles. Furthermore, lipids sharing biosynthetic pathways usually participate in similar biological processes as unique structural subgroups. For instance, it was found that persistent EGFR signaling increased saturated phosphatidylcholine (PC) while concomitantly reducing saturated lysophosphatidylcholine (LPC) in glioblastoma (GBM) cells². Indeed, lysophosphatidylcholine acyltransferase 1 (LPCAT1), which converts LPC to PC, is critical to lipid remodeling and supports oncogenic signaling. Another study employed lipidomics to reveal striking changes to the linoleic acid (LA) to arachidonic acid (AA) metabolism axis upon coronavirus infection⁹. Thus, modulation of the LA-AA axis could help combat emerging coronaviruses through exogenous supplement of LA or AA or inhibition of cytosolic phospholipase A2 α (cPLA2 α). These studies highlight the value of a lipid network accounting for the interconnections and interdependence of lipidomics data.

While many lipidomics analysis tools exist, most do not account for the pathway context nor interdependencies of measured lipids. Web servers such as LipidSig⁵ and LipidSuite⁶ allow researchers to interrogate the associations between diverse lipid characteristics and biological phenotypes. Yet, analysis of single lipids is usually limited by low statistical power after false discovery rate (FDR) correction. Moreover, while lipid characteristics encode structural similarity among lipids, they do not account for their interconnections in the lipid biosynthetic pathways. BioPAN introduces a lipid network analysis approach^{10,11} and enables interpretation of altered pathways. Nevertheless, this method cannot address the interdependence in raw lipidomics data, making the following computation influenced by highly variable or sparsely measured lipids and therefore impacting the accuracy of analyses. Thus, the interconnections and non-independence in lipidomics data require a new approach to more precisely analyze the altered pathways and the genetic source of a lipid profile.

To address these challenges, we present iLipidome, a novel systems biology-based approach to analyze lipidomics data. Importantly, iLipidome accounts for biosynthetic steps and interdependences of experimentally measured lipids. It decomposes each lipid into intermediate substructures that can aggregate slight changes in lipid profiles based on their shared biosynthetic network¹². This method significantly improves statistical power, reduces data sparsity, and increases biosynthetic network coverage in lipidomics data analysis. We also show its value in hierarchical clustering and lipid enrichment analysis. Furthermore, the substructure network encoding biosynthetic information can explore the mechanisms behind changes in lipid abundance. Here, we pinpointed the genetic sources of lipidomic changes following docosahexaenoic acid (DHA) supplementation. The hidden lipid interactions revealed by iLipidome also enabled us to identify the substrate preferences of LPCAT1 and the therapeutic targets for GBM. Lastly, we demonstrated such analyses could reveal the potential mechanisms and guide the early biomarker development for cardiovascular disease (CVD). Collectively, iLipidome facilitates systems-level comparison of lipid profiles with lipid biosynthetic information efficiently and provides a deeper insight into complex lipidomic alterations across samples.

Result

High variability in lipid measurements introduces challenges in conventional lipidomics analysis

Current lipidomics analysis strategies mainly focus on individual lipids or lipid characteristics. However, since many lipids share biosynthetic steps and enzymes, changes in one lipid may impact the abundance of many others. To understand this dependency and how it impacts lipidomics analysis, we first applied conventional lipidomics analysis methods on a dataset studying the effect of DHA (22:6 omega-3) treatment on rat basophilic leukemia (RBL) cells³.

The general differential abundance analysis showed the number of significant lipid species drastically decreases after correcting for multiple hypothesis testing (**Fig. 1a**). This stems from the high variability and sparsity⁸ in detected lipids across samples, wherein few lipids (90/444=20.3%) were measured in all samples (**Fig. 1b**). Many studies have confirmed that the *de novo* synthesis of unsaturated fatty acids (FAs) in cells will be inhibited by DHA supplementation^{3,13,14}. Nonetheless, the high variability impeded the discovery of this important FA remodeling signature in the enrichment analysis after FDR correction (**Fig. 1c**). Practically, one can categorize lipids into different subgroups in terms of specific lipid characteristics (e.g., FA chain length and double bond number) to improve the statistical power^{5,6}. Thus, we employed a similar analysis and found the expected rise in the categories of chain length 22 and double bond 6 (p-value < 0.001), corresponding to the feature of DHA (**Fig. 1d,e**). Further, the chain length-double bond combined analysis unveiled a more precise pattern of FA change (e.g., increase in 16:0 and 22:6 FAs (p-value < 0.01) or decrease in 18:1 and 20:2 FAs (p-value < 0.05)) (**Fig. 1f**). These methods provide insights into shared lipid features and increase statistical power by aggregating lipids with similar structures. However, the aggregation does not consider the sequential connections of the lipids or FAs determined by biosynthetic pathways. It's also difficult to identify genetic source of change, despite noticing a consistent decrease in abundance of a subgroup of unsaturated FAs (18:1, 20:1, and 20:2 FAs) (**Fig. 1f**). Thus, the conventional methods fail to balance the statistical power and the integrity of individual lipids at the same time, and they further pose a challenge in identifying the genetic basis of a global reprogramming of the lipidome. To address these challenges, we developed iLipidome, a substructure-based approach that uses the lipid biosynthetic pathways for data analysis to solve the non-independence issue in lipidomics data, provide interpretable biosynthetic information, and explore impacted pathways and enzyme catalyzed reactions.

iLipidome decomposes lipid profiles into biosynthetic precursors to account for interdependencies during analyses

iLipidome is a systems biology-based method to analyze lipidomics data by decomposing lipids into substructures, reconstructing their biosynthetic network, and identifying critical pathways and genes/enzymes.

First, we manually curated the FA and lipid class pathways from KEGG¹⁵, Reactome¹⁶, and recent lipid studies and reviews (Fig. 2a)¹⁷⁻²². Further, the iLipidome algorithm can combine the resulting lipid biosynthetic network with the LIPID MAPS database²³ and the FA information encoded in lipids to decompose the lipids into three types of substructures: lipid class substructures, FA substructures or lipid species substructures, based on their biosynthetic routes (Fig. 2b). Specifically, the abundance of each measured lipid can be backpropagated to their biosynthetic intermediates, and summed when more than one lipid shares a common substructure. iLipidome, therefore, accounts for interdependencies and interconnections between lipids by searching for substructures (biosynthetic intermediates) of the experimentally measured lipids in lipidomics data. A FA or lipid can be synthesized from the *de novo* (endogenous) pathway using a basic structural unit, such as acetyl-CoA or glycerol-3-phosphate (G3P), or from an exogenous intermediate absorbed by a cell²⁴. Additionally, changes in abundance of the measured lipids could emerge from increased/decreased uptake of related precursors or alterations of enzyme activities. Theoretically, absorbed lipid intermediates or activated enzymes will increase the products in their downstream pathways, while inhibiting upstream pathways. We therefore use the fold change of each substructure between two experimental groups to identify the potential source of altered lipids and further extract the substructures in one biosynthetic path (Fig. 2c). Following the rules above, each lipid in a lipidomics dataset can be transformed to their substructures, all of which will be further used to produce a substructure matrix recording the frequency of occurrence of each substructure (Fig. 2d). The substructure abundance can then be calculated using the substructure matrix and the lipid profiles. The summation of each substructure is not simply the amount of an intermediate but reflects the activity of biosynthetic steps required to form the current lipid profiles. Therefore, the non-independence challenge is addressed through the analysis of lipid substructures. Moreover, since the substructures from biosynthetic pathways are interconnected and can aggregate lipid signals, there is less concern of data sparsity (i.e., unmeasured lipid intermediates) to reconstruct the network from different lipid profiles (Fig. 2e). We further introduce a scoring system to identify essential pathways and reactions that are significantly activated or suppressed across conditions (Fig. 2f)²⁵. More detailed information on the reference pathway collection, substructure decomposition and extraction, and pathway and reaction scoring methods can be found in Materials and Methods.

iLipidome increases the statistical power and reveals genetic source responsible for DHA-induced FA remodeling

We first applied iLipidome to the previous DHA dataset and compared the results from the original and substructure transformed data. By encoding biosynthetic context, iLipidome provides richer information (Fig. 3a,b) and significantly ($p = 6.9e-04$) enhanced statistical power than using raw FAs alone (Fig. 3c). Besides, it gives us the opportunity to examine altered pathways and enzymes based on lipidomics data only. In the FA substructure network, three significantly altered pathways can be clearly observed: (1) activation of omega-3 FA synthesis, (2) accumulation in saturated FAs, and (3) reduction in unsaturated FAs in the non-essential FA synthesis (Fig. 3b,d). Intriguingly, we identified several key desaturases and elongases, such as stearoyl-CoA desaturase 1 (SCD1), ELOVL1/2/3/5/7 (ELOVL Fatty Acid Elongase 1/2/3/5/7), and fatty acid desaturase 1/2 (FADS1/2), that may engage in the FA remodeling induced by DHA supplementation (Fig. 3e). Among them, SCD1 catalyzed reactions are decreased and found to locate exactly at the boundary between active and suppressed pathways, highlighting their central roles in overall non-essential FA remodeling (Fig. 3b,e). In fact, down-regulation of SCD1 by polyunsaturated fatty acids (PUFAs) containing DHA has been validated in many studies^{13,26}. It suppresses the crucial steps in FA unsaturation and counteracts DHA perturbations³. On the contrary, analysis of unprocessed lipidomics data using a conventional approach showed a lower statistical power, discovered only partially impacted pathways, and did not successfully identify critical SCD1 (Fig. 1a,d and Extended Data Fig. 1a,b).

Another interesting finding is the increased activity of eicosapentaenoic acid (EPA, 20:5 omega-3) synthesis (Fig. 3b), which iLipidome suggests can be attributed to the inhibition of ELOVL2/5 (Fig. 3e). This result is consistent with many other studies, which suggest that DHA supplementation may slow down EPA metabolism and lead to an accumulation of EPA synthesized from alpha-linolenic acid (ALA, 18:3 omega-3) (Fig. 3b)^{17,27-29}. Indeed, DHA feeding lowers the expression of ELOVL2/5 in mice^{28,29}. Many other elongase involved reactions including ELOVL1/3/7 were also detected changed by DHA treatment (Fig. 3e). Generally,

they are responsible for long chain saturated and monounsaturated FA elongation^{30,31}. However, how they are reversely modulated by PUFAs is poorly studied in comparison to ELOVL2/5³².

Since the substructures summarize the biosynthetic context and highlight the difference across conditions, we further tested if the hidden biosynthetic similarities between lipids could benefit sample clustering. Indeed, FA substructures displayed a higher adjusted Rand index value (0.435 to 0.692) and a more consistent clustering on the DHA samples (**Fig. 3f-g**). Further, in the row clustering, the highly unsaturated FAs (3 or more double bonds) and saturated FAs were clearly separated from the other unsaturated FAs (1 and 2 double bonds), consistent with the previous conclusion of DHA-induced FA remodeling (**Fig. 3b,g**). We also observed DHA4 has less DHA incorporation, which can explain why it is an outlier and closer to control group when clustering (**Extended Data Fig. 1c,d**). Taken together, iLipidome successfully increases the statistical power and reliability for distinguishing samples and FA features correctly. More importantly, it accurately predicts the pivotal pathways and reactions that are hardly visible using conventional methods, providing more comprehensive insights into overall FA remodeling. These advantages can also be observed when we decomposed lipids to species substructures (**Fig. 4 and Extended Data Fig. 2 and Supplementary Text 1**). The additional statistical power relies on improvement of data sparsity and enhance following enrichment analysis. Finally, iLipidome exhibited a higher network coverage and greatly increased the interpretability of lipidomics data by reconstructing its biosynthesis.

iLipidome reveals enzyme FA preference and suggests potential therapeutic targets for GBM

The Lands cycle is crucial for FA remodeling in membrane structural lipids, which can further impact membrane properties and functions^{2,33,34}. Here, we applied iLipidome to uncover the hidden interactions by LPCAT1 in the Lands cycle and to explore promising therapeutic targets with a GBM dataset² (**Fig. 5a**). In the dataset, their LC/MS lipidomics analysis could not differentiate each FA in phospholipids (e.g., PC 32:0 instead of PC 16:0-16:0). We found saturated PCs were most suppressed after LPCAT1 knockout but it is difficult to deduce their relationships (**Extended Data Fig. 3a**). We accounted for the interdependence among lipids and were able to increase statistical power and build the connections between lysophospholipids (one FA) and phospholipids (two FAs) (**Fig. 5b and Extended Data Fig. 3b**). LPCAT1 has specificity to synthesize saturated PC from saturated FAs and LPC^{33,35}. As expected, the pathways involving saturated PC were suppressed in the LPCAT1 knockout group (**Fig. 5c**). Simultaneously, we found three saturated LPC-PC reactions were most inhibited in the Lands cycle in the substructure network (**Fig. 5d**). In contrast, no pathways showed significant and only two LPCAT1 reactions were identified when we analyzed the raw lipid species data (**Extended Data Fig. 3c,d**). The interactions unraveled by iLipidome allowed us to understand the exact chain length and double bond number of PCs (12:0-16:0, 14:0-16:0, and 16:0-16:0), along with free FAs involved (12:0, 14:0, and 16:0) (**Fig. 5b**). Our findings matched the enzyme activity experiment showing that these saturated FAs were three of the most preferred substrates for LPCAT1³⁵. As lauric acid (12:0), myristic acid (14:0), and palmitic acid (16:0) are the major components of palm oil and coconut oil, the FA information exposed by iLipidome also raises their dietary source as having potential relevance to GBM, alongside treatments modulating LPCAT1. Finally, a corresponding accumulation of free 16:0 FA was validated in the LPCAT1 knockout groups (**Fig. 5e**). Altogether, the unique function in iLipidome gives us an opportunity to dissect exact FA composition, interrogate enzyme substrate specificity, and identify dietary lipids associated with disease and candidate treatments.

iLipidome explores molecular mechanisms for cardiovascular diseases (CVDs) and facilitates early biomarker discovery

Disturbances in lipid homeostasis have been linked to many disorders, making lipidomics analysis a powerful tool to investigate disease mechanisms and to search for biomarkers³⁶⁻³⁸. Normally, most significant lipids will be selected for further analysis, but they can be limited to high variability in lipidomics data and barely explain the fluctuations of the entire lipidome. To expand the applicability of iLipidome for exploring mechanisms and biomarker [candidates](#), we applied our method to a dataset composed of 146 plasma lipid profiles from 85 controls and 61 potential CVD patients (**Fig. 6a**)⁴. The 61 individuals went to the hospital with chest pain and

displayed a similar lipidomic pattern with those having more severe CVD symptoms. This suggests that it may have both therapeutic and diagnostic value to analyze lipid alterations in these patients.

First, iLipidome significantly improves the statistical power in this dataset (**Extended Data Fig. 4a**). The pathway analysis revealed that lipids with 18:2 FA and 20:4 FA were enriched in the suppressed and active pathways, respectively (**Fig. 6b and Extended Data Fig. 4b,c**). The major 18:2 FA in human body is LA, which can normalize blood cholesterol and triglyceride level through proprotein convertase subtilisin/kexin type-9 (PCSK9)^{39,40} and negatively correlate with CVD risk⁴¹. Otherwise, AA is the main form of 20:4 FA and an important precursor of many inflammatory mediators, such as hydroxyeicosatetraenoic acid (HETE) and leukotrienes⁴². Their critical roles in inflammation and coagulation may explain their positive association with CVD risk in our and the other studies⁴³⁻⁴⁶. Moreover, growing evidence indicates arachidonoyl phospholipids (PLs), especially PE, can act as substrates for lipid peroxidation that result in ferroptosis and cell death^{47,48}. Our iLipidome analysis also showed the most altered reactions tend to accumulate AA in PC and PE in the CVD group (**Extended Data Fig. 4d,e**). Besides, most ferroptosis related reactions, such as the Lands cycle or LPAAT reactions involving AA addition, were consistently up-regulated (**Extended Data Fig. 4f,g**). As our substructure method can capture upstream intermediates LPA, it allows a unique chance to survey both types of reactions and their associated enzymes (Extended Data Fig. 4g). Since our lipidomics data came from plasma and these reactions usually occur in cells, it is difficult to determine where these AA-rich PLs were synthesized. However, their increase in plasma as free PLs or lipoprotein conjugates might have a higher chance to be incorporated into tissues through lipid transporters or lipoprotein receptors and exert related biological functions including inflammation and ferroptosis, both of which have been linked to CVDs⁴⁹⁻⁵¹.

Our substructure method reveals potential mechanisms and summarizes most impacted lipid pathways, allowing us to find robust biomarkers for CVDs. We compared the p-values from the top 5 significant raw lipid species and substructures (**Fig. 6c**). Although the most significant feature is LPC 16:0, lysophospholipids usually locate at the end of a pathway, so theoretically their statistics will not be affected by the substructure abundance of other lipids. On the contrary, we found the statistical power increases in three out of four phospholipids after transformation (**Fig. 6c**). Interestingly, these lipids all carry 18:2 FA, corresponding to the most suppressed pathways we identified (**Extended Data Fig. 4c**). A further paired analysis demonstrated that the substructures could aggregate lipid signals in a biosynthetic scenario and thus were more stable and showed a better discriminatory power than using raw lipid species alone (**Fig. 6d**). Moreover, iLipidome clearly points out the composite lipid units for each substructure, providing a more systems-based platform to optimize biomarkers. In summary, iLipidome allows intensive analysis of underlying mechanisms behind complicated lipidomics data and simultaneously guides robust biomarker discovery that can be translated into clinical application.

Discussion

Lipidomic data are typically studied as if each lipid is independent^{48,52,53}. iLipidome, in contrast, uses a data-driven approach combining FA information to search substructures for a lipid profile based on their biosynthetic pathways. These innovative features of iLipidome enrich small changes in shared lipid intermediates, resulting in statistically significant substructures and pathways. Firstly, we showed here with three example datasets that iLipidome elucidated significant increases in statistical power (**Fig. 3c and Extended Data Fig. 2a, 3b and 4a**) and better sample coverage (**Fig. 4c and Extended Data Fig. 1e**). Accounting for interdependence and interconnections enabled us to explore altered biosynthetic fluxes in a unified biological theme (**Fig. 3b, 4f, 5b and 6b**). Moreover, the advantages are also apparent in the further analysis, such as hierarchical clustering and enrichment. Substructure clustering displays a more biologically meaningful grouping of both the experimental conditions and the lipid features (**Fig. 3f,g and Extended Data Fig. 2e,f**). FA enrichment analysis for substructures also reveals a significant DHA-induced FA remodeling that may be missed in conventional analysis of lipidomics data (**Fig. 4b**).

As lipidomics data lack information about the lipid biosynthesis and interdependence, it remains difficult to pinpoint genetic causes of alterations in lipid profiles. Through iLipidome, we can now more easily find the enzymes and perturbed reactions that drive lipidome reprogramming across conditions. For example, the

inhibitory effect of DHA or PUFAs on SCD1 has been extensively studied^{3,13,26,54}. Our method successfully elucidated this relationship with the FA substructure data (**Fig. 3e and Extended Data Fig. 1b**). Moreover, iLipidome successfully attributed the increase of EPA following DHA to the inhibition of ELOVL2/5 and EPA elongation to docosapentaenoic acid (DPA, 22:5 omega-3) (**Fig. 3b**). Our results are further supported by the experiments showing that ELOVL2/5 are down-regulated by DHA and dietary ALA is necessary for the EPA accumulation with DHA feeding^{28,29,55}. Interestingly, numerous desaturases and elongases were also identified to be affected in the DHA-treated samples (**Fig. 3e**). More studies, however, are required to clarify these relationships and whether the common regulators of lipid metabolism, such as sterol regulatory element-binding protein (SREBP) or proliferator-activated receptor (PPAR), have a role to play^{32,56}. We found the genetic source to induce FA reprogramming after DHA supplementation and saw the effect can be profoundly passed to all lipid species (**Fig. 3b and 4f**). Also, our substructure decomposition can capture lipid intermediates and enables one to explore the upstream reactions (e.g., LPAAT) that could induce wide-ranging phospholipid remodeling (**Extended Data Fig. 2j**). In the other dataset, we identified the reactions to accumulate arachidonic PE and PC increased in the CVD patients (**Extended Data Fig. 4c**). These plasma lipids could potentially induce inflammation or ferroptosis and contribute to the development of CVD^{43,45,49,50}. Since the reactions involving AA addition primarily occur in cells, it might be valuable to investigate where these reactions and associated enzymes (e.g., LPCAT3 and AGPAT3) involving AA transfer are dysregulated in the early stage of CVDs. LPCAT3 has LPEAT and LPCAT activity, while AGPAT3 (LPAAT) catalyzes LPA reactions³³. To further examine the general applicability of our method, a dataset containing multiple gene knockouts in lipid metabolic pathways was also used to assess the performance of iLipidome⁵⁷. Overall, the top 5 perturbed pathways across 11 gene knockouts revealed by iLipidome were highly consistent with the enzyme specificities and the expected lipidomic changes affected by each knockout (Extended Data Table 1). More essentially, five out of six genes that appear in our reference pathway were also ranked in the first place of their respective most suppressed reactions. Taken together, iLipidome sheds light on the genetic source of complex lipid profiles. These insights will guide future research into the mechanisms regulating lipid disorders or interactions between PUFAs and endogenous FA biosynthesis.

Another strength of iLipidome is to reconstruct the active biosynthetic network in measured samples. Indeed, the low concentrations of lipid intermediates make it hard to link reactants and products in a lipid profile^{33,58}. Compared to the BioPAN tool¹⁰ (**Extended Data Fig. 5a-c**), our substructure method forms a network with a high degree of coverage and connectivity (**Fig. 4d and Extended Data Fig. 1f**). It might be because the BioPAN algorithm requires free FA data to connect the reactions involving FA transfer. In addition, iLipidome offers a more comprehensive analysis pipeline that uses the constructed network to identify lipid intermediates/reactions and improve statistical power and data sparsity in contrast to the other lipid network analysis tools (Extended Data Table 2)⁵⁹. In iLipidome, we have two unique systems to decompose lipids according to different structural resolutions. First, FAs should be inherited in lipid species using the same biosynthetic pathways. Second, for those lipids lacking the exact identity of FAs (e.g., PC 30:0)^{2,48}, they will be matched to the LIPID MAPS Structure Database (LMSD)²³ and find possible FA combinations. These methods allow us to maintain the completeness of biosynthetic networks and the important FA information, thus extending the method's applicability in searching potential therapeutic targets and clinical biomarkers. For example, high saturated PC levels produced by LPCAT1 are required to maintain EGFR signaling and GBM cell growth². Our algorithm successfully established clear links between LPC 16:0 and three saturated PC (28:0, 30:0, and 32:0) (**Fig. 5b**) and suggested two compelling therapeutic strategies for GBM: (1) inhibition of LPCAT1 and (2) depletion of free FAs, especially lauric acid (12:0), myristic acid (14:0), and palmitic acid (16:0). iLipidome not only filled the gap of FA information but also correctly explained the substrate specificity of LPCAT1³⁵. In the CVD dataset, the inherited feature of lipids sharing same FAs helped to connect each substructure and can be applied to identify novel lipidomic biomarkers. By decoding substructures, we clearly see how lipid signals aggregate in a biosynthetic pathway (**Fig. 6d**). This process allows iLipidome to explain interdependence among measured lipids, thereby improving statistical power and sparsity and making substructures a more stable and precise biomarker. Altogether, this functionality will be invaluable for interrogating the specificity of the acyltransferase family members and its effect on the membrane protein behavior. Furthermore, it provides an ideal platform for researchers to design reliable lipid biomarkers in a systems biology scenario.

Figure legends

Fig 1: Limitations in conventional lipidomics analysis methods. **a**, The volcano plots show the differential abundance of lipid species based on adjusted p-values in the comparisons between 7 controls and 6 DHA-treated RBL cell samples. **b**, The sample coverage of lipid species is quantified as the frequency of a lipid species detected in 13 samples. Many lipid species show low overlap in the raw lipidomics data. The sparsity and non-independence in lipid profiles are not corrected in differential expression analysis, which contributes to the low statistical power. **c**, FA enrichment analysis for the statistically significant lipid species before and after false discovery rate (FDR) correction. The significantly enriched and depleted FAs were outlined in red and blue, respectively. **d-f**, We calculated (**d**) FA chain length profile, (**e**) lipid unsaturation (double bond) profile, and (**f**) the two-characteristics combined heatmap in membrane glycerophospholipids (GPLs) and diacylglycerols (DAGs) in the DHA dataset. DHA supplementation not only leads to a dramatic increase in its corresponding 22:6 FA but also causes obvious remodeling in the other FAs. However, these methods cannot explain the interconnected nature of altered lipids and the genetic source of changes. Two-tailed Student's t-tests were used to calculate the p-values in (**a**) and (**d-f**). Benjamini-Hochberg p-value correction was used for all analyses. The dashed lines in (**a**) and (**c**) indicate p-value=0.05. *p<0.05, **p<0.01, ***p<0.001.

Fig 2: Schematic overview of the analysis workflow for iLipidome. **a**, The reference FA and lipid class biosynthetic pathways were manually curated from KEGG¹⁵, Reactome¹⁶, and recent lipid studies and reviews¹⁷⁻²². **b**, Lipids are decomposed to substructures (lipid class substructures, FA substructures or lipid species substructures) based on the reference biosynthetic pathways and the LIPID MAPS database²³. **c**, To identify the potential source of altered lipids, substructure units in each biosynthetic route are further reduced through backpropagation and use of the fold changes between control and experimental groups (see Methods). For example, two lipid species (yellow nodes) were decomposed into their substructures. In the left pathway, the propagation stops at the third node due to the opposite fold change with the lipid species. Thus, only two substructures are kept (red rectangle). Likewise, four substructures are retained in the right pathway (blue rectangle). **d**, The presence frequency of each substructure for each lipid is recorded in a substructure matrix, which is further weighted by the lipid profiles and summed into the final substructure abundance matrix. **e**, The substructure profiles produced by each sample retain biosynthetic information, and these are combined to reconstruct a unified lipid biosynthetic network. **f**, iLipidome enables one to identify the essential pathways and reactions in the biosynthetic network. The significant representative pathways (color paths) and the highly perturbed reactions (crosses) are calculated and labeled in the network, with red and blue indicating activation and suppression, respectively. Edge colors and weights can indicate pathway importance.

Fig 3: Analysis of FA substructures using iLipidome increases statistical power and elucidates FA remodeling and its genetic sources induced by DHA. **a,b**, The FA biosynthetic network derived from (**a**) the raw FA or (**b**) the iLipidome FA substructure data in RBL cells supplemented with DHA. Two-tailed Student's t-tests with Benjamini-Hochberg correction were used to calculate the adjusted p-values. Nodes are filled according to log₂ (fold change) and their sizes denote -log₁₀ (adjusted p-value). If one node shows significant changes in abundance, its border will be marked as purple. The significant representative pathways and the highly perturbed edges are also labeled in the network, with red and blue indicating activation and suppression, respectively. Line width and color depth marks pathway importance. **c**, Boxplots reveal a significant improvement in the adjusted p-values of the substructures. **d**, Significantly activated and suppressed representative pathways in the network from panel (**b**). Pathways are simplified and represented by the first and the last metabolite. If different pathways have a same name, an additional metabolite is used to differentiate them. **e**, Top 5 significantly activated and suppressed edges (reactions) in the network from panel (**b**). Here, we removed reactions including DHA to avoid artifacts due to supplementation. One edge (reaction) is composed of two nodes (substrate and product), whose colors are defined as the fold changes of their abundance. Edge types including increase, decrease, and no change reflect the changes of the two associated nodes. Bars are outlined yellow if the edge type is same as the perturbation of reaction. **f,g**, Substructure profiles enable more precise clustering of biological samples and lipid characteristics. Heatmap of hierarchical clustering for (**f**) unprocessed FAs and (**g**) FA substructures in 13 samples. We divided samples into two groups and measured the clustering similarity with the original labels. FA variables on the rows are colored according to the double bond number. Black, blue, and red denote FAs without a double bond, with 1 and 2 double bonds, and with 3

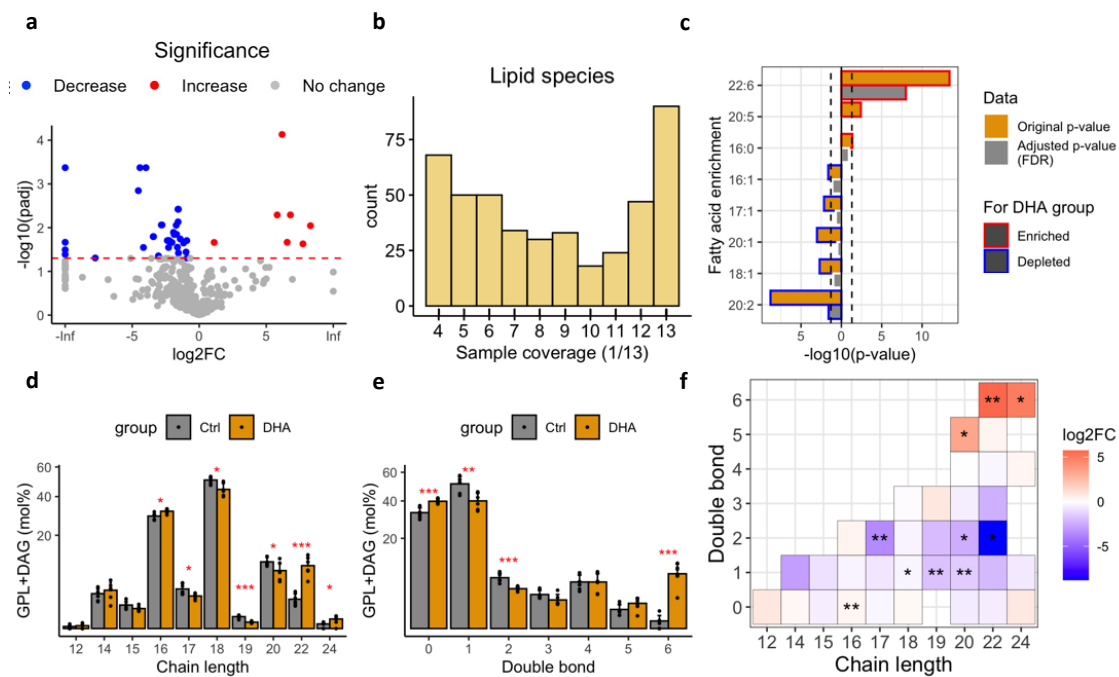
or more double bonds, respectively. FA variables are also separated into two groups with black and red combined into one category and calculated the clustering similarity. EPA, eicosapentaenoic acid; DHA, docosahexaenoic acid; w3, omega-3; w6, omega-6; w7, omega-7; w9, omega-9; endo, endogenous; exo, exogenous; SCD1, stearoyl-CoA desaturase-1; FADS1/2, fatty acid desaturase 1/2; ELOVL1/2/3/5/7, ELOVL Fatty Acid Elongase 1/2/3/5/7; ARI, adjusted Rand index.

Fig 4: Substructure analysis improves the non-independence and sparsity of measured lipids, thus improving the statistical analysis of the differences across conditions and improving interpretability. **a**, A Venn diagram shows the increased number of significant features (lipids and substructures) with iLipidome (adjusted p -value <0.05) in DHA-treated RBL cells. **b**, Enrichment analysis for significant lipid species or substructures based on FA categories. **c**, Proportion of samples containing a lipid or substructure in 13 samples, and the associated probability distribution. **d**, The coverage in entire biosynthetic network for 13 lipid profiles using lipid species or substructures. The improvement index in **(c)** and **(d)** is calculated based on the average sample or network coverage from the raw lipid species and the iLipidome substructure data. **e,f**, The lipid biosynthetic network built from all significant pathways included in the top 3 activated and suppressed representative pathways in **Extended Data Fig. 2d,e**. Red and blue indicate activation and suppression while color depth denotes the importance of pathways. The top 5 perturbed edges in **Extended Data Fig. 2f,g** are also labeled in the network. GPAT, glycerophosphate acyltransferase; LPAAT, lysophosphatidic acid acyltransferase; LPCAT, lysophosphatidylcholine acyltransferase; LPEAT, lysophosphatidylethanolamine acyltransferase; LPIAT, lysophosphatidylinositol acyltransferase; PLA, phospholipase A; PEMT, phosphatidylethanolamine N-methyltransferase; PSS2, phosphatidylserine synthase 2.

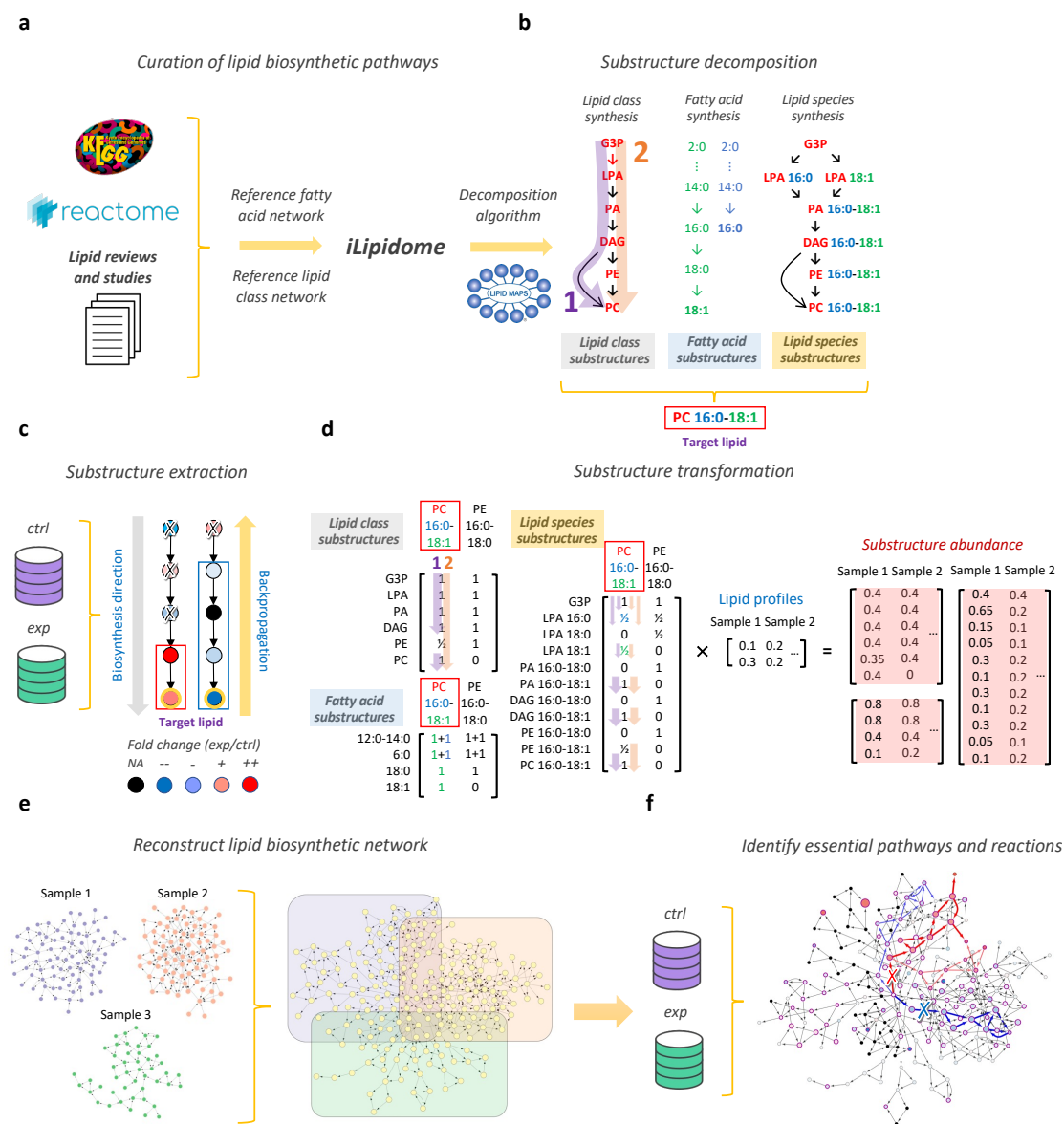
Fig 5: Identification of LPCAT1-associated reactions and its FA specificity by iLipidome lipids associated with GBM and potential treatments. **a**, LPCAT1-mediating membrane lipid remodeling is required for sustained EGFR signaling and promotes GBM cell growth. **b**, When comparing wild type and LPCAT1 knockout U87EGFRvIII GBM cells, the iLipidome substructure network built included three significantly suppressed pathways containing saturated PCs. **c**, Top 5 significantly suppressed pathways in the substructure analysis. Pathways are simplified and represented by the first and the last metabolite. If the last metabolite belongs to lysophospholipids, we add the second to last metabolite. **d**, Top 5 Lands cycle's reactions that are significantly suppressed in the network built from **(c)** We highlighted the pathways or the reactions containing saturated PC with a red frame in **(c)** and **(d)**. **e**, Boxplots display a significant increase in 16:0 free FA in the LPCAT1 knockout groups.

Fig 6: iLipidome unveils the lipid dysregulation and identifies early biomarker for cardiovascular diseases (CVDs). **a**, A schematic of the experiment comparing plasma lipidomics between controls and CVD patients. **b**, The substructure network built from all significant pathways included in the top 5 activated and suppressed representative pathways in **Extended Data Fig. 4d**. The top 5 perturbed edges in **Extended Data Fig. 4e** are also labeled in the network. **c**, Paired comparison of the statistics from top 5 significant raw lipid species and substructures. Only those lipids not missing in raw lipidomics data are considered. If the substructure method improves the statistical power, the lipids will be colored in orange. **d**, Boxplots show species substructures are more stable and have a higher discriminatory power than raw lipid species. LPCAT, lysophosphatidylcholine acyltransferase; LPEAT, lysophosphatidylethanolamine acyltransferase; PLA, phospholipase A; PLC, Phospholipase C.

1 **Figures**
2
3 **Fig. 1**



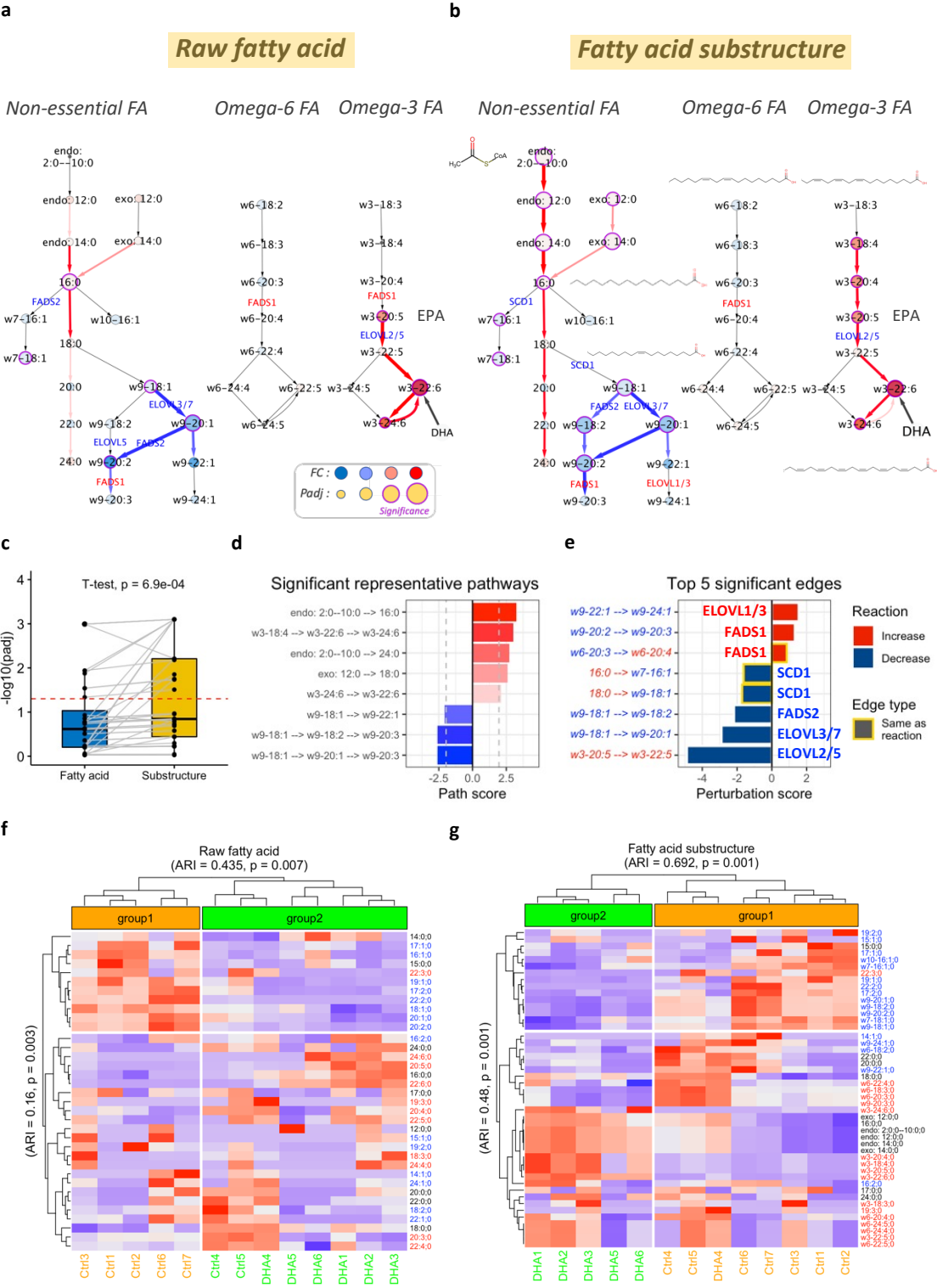
1 Fig. 2



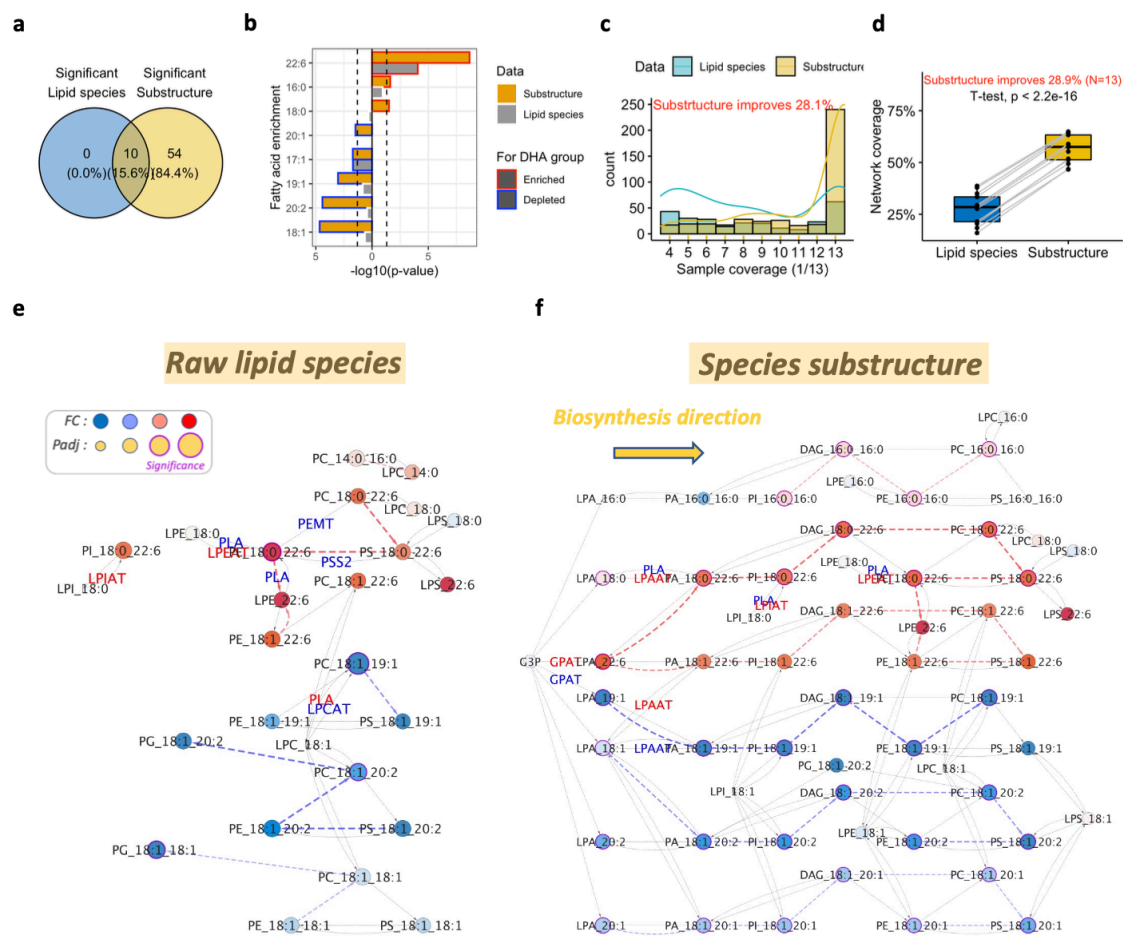
2

3

1 Fig. 3

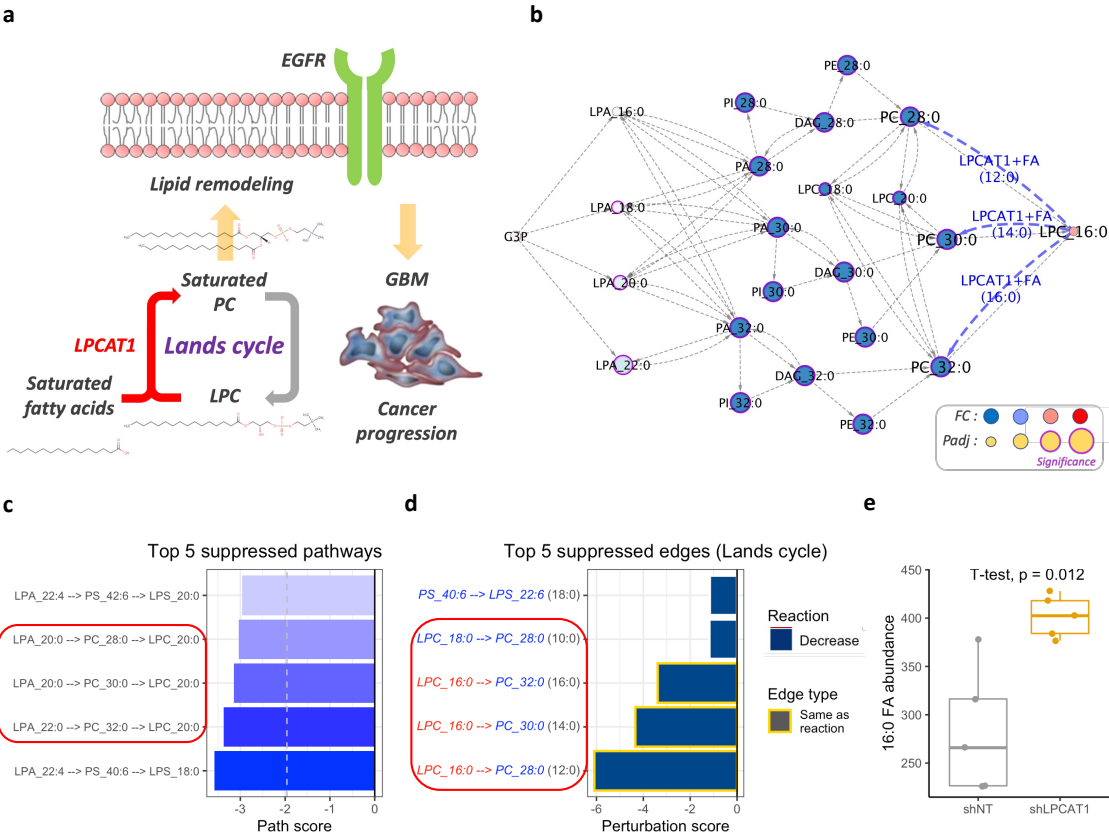


1 Fig. 4

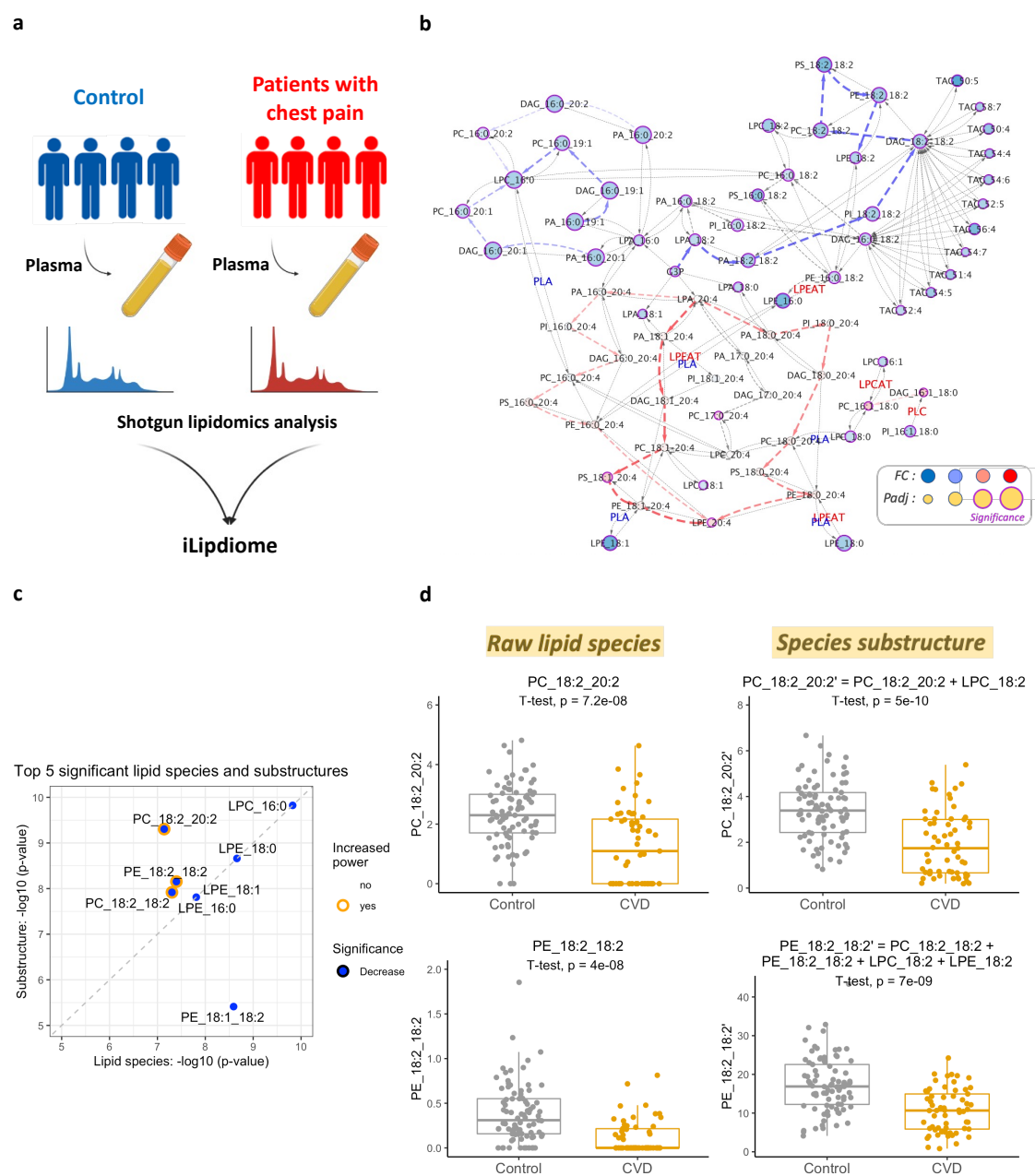


2
3

1 Fig. 5



2



Materials and Methods

Demonstrating lipidomics datasets

We demonstrate the developed functionalities of iLipidome on four publicly available lipidomics datasets^{2-4,57}. Except the CVD dataset, we adopted the same data-processing approach. All of our datasets used internal standards to quantify measured lipids and obtained their picomole values (absolute quantification). First, we excluded the lipid species with missing values in over 70% samples to reduce the noise. The default parameter for this step in MetaboAnalyst⁶⁰ and LipidSig⁵ is 50%, which we used in the larger CVD dataset (146 samples). On the contrary, since the other three datasets have fewer samples (13, 10, and 6 or 9 samples for each comparison), we relaxed the threshold (70%) to include more lipid species and to better reflect comprehensive lipidomic changes. Secondly, the values were transformed into mol% to obtain a final lipid profile. The first dataset explored the lipidomic and biophysical homeostasis of mammalian membranes as influenced by dietary lipids, which contains 13 membrane lipid profiles including 7 controls and 6 DHA treated rat basophilic leukemia (RBL) cells³. To investigate membrane lipid metabolism, 444 individual lipid species across 16 lipid classes in GPLs and diglycerol (DAGs) were selected for the following analysis (**Supplementary Table 1**). We also computed the FA profiles through the decomposition of lipid species into FAs and the summation of their values (**Supplementary Table 2**). Two different kinds of profiles were used to calculate FA or species substructures, respectively. We removed ether lipids when computing species substructures since we could not differentiate alkyl and alkenyl ether lipids. It might introduce a bias if we directly map them to our reference pathways, where two kinds of ether lipids are separated. Further, the biosynthetic pathways of ether and ester lipids are bifurcated at glycerol-3-phosphate, the start point of glycerophospholipid biosynthesis. Hence, the backpropagation process of ether lipids would have a minor effect on the results we obtained from ester lipids only. The second dataset studied the role of LPCAT1 in the transduction of oncogenic signals through shaping membrane lipid composition, which contains 10 lipid profiles of human glioblastoma cell line (U87EGFRvIII), expressing LPCAT1 shRNA or a non-targeting control². There are 82 lipid species across 6 lipid classes and 4 free FAs (16:0, 18:1, 18:0, and 20:4) in this dataset (**Supplementary Table 3**). The abundances of lipid species were pre-processed using the common method above while free FAs were analyzed with raw values. For the CVD dataset, we collected 146 plasma lipid profiles including 85 controls and 61 potential CVD patients to explore the mechanisms and early biomarkers for CVDs⁴. Those patients went to the hospital with chest pain but without indicators for stable angina pectoris, unstable angina pectoris, or myocardial infarction. The lipid species missing in over 50% samples were excluded, resulting in an expression table composed of 197 lipid species across 9 lipid classes (**Supplementary Table 4**). To reflect the actual concentration of lipids in the plasma, we adopted the raw lipid values to analyze this dataset and design the biomarkers. The final dataset investigated multiple gene knockouts in lipid metabolic pathways and explored how they affected the lipidome in the human colorectal carcinoma (HCT116) cell line⁵⁷. We selected the lipid profiles of 2 control groups based on the different knockout method (gene trap or deletion) and of 11 gene knockout groups whose effects on the lipidome have been clearly described in the paper. Each group contains at least three replicates. In total, 30 lipid profiles (3 controls and 27 knockouts across 8 genes) from the gene trap method and 15 lipid profiles (3 controls and 12 knockouts across 3 genes) from the deletion method were collected for the following analysis. Two methods also result in 1871 and 1945 lipid species across 24 lipid classes, respectively (**Supplementary Table 5**).

Substructure decomposition based on biosynthetic pathways

Curation of the reference lipid biosynthetic pathways. To decompose lipids into substructures, FA and lipid biosynthetic pathways for human, mouse, and rat were obtained from two widely used pathway databases (KEGG¹⁵ and Reactome¹⁶). In this study, we collected 6 KEGG pathways (Fatty acid biosynthesis, Glycerolipid metabolism, Glycerophospholipid metabolism, Ether lipid metabolism, Sphingolipid metabolism, and Biosynthesis of unsaturated fatty acids) and 6 Reactome pathways (Fatty acyl-CoA biosynthesis, alpha-linolenic (omega3) and linoleic (omega6) acid metabolism, Triglyceride metabolism, Phospholipid metabolism, Sphingolipid metabolism, and Wax and plasmalogen biosynthesis) to cover most common FAs and lipid classes in lipidomics analysis¹⁻⁴. Together with recent studies and reviews on the metabolism of polyunsaturated FAs¹⁷⁻

¹⁹, ether lipids^{20,21}, and deoxysphingolipids²², we manually curated the resulting reference lipid biosynthetic pathways, which compose of 64 lipid class (**Supplementary Table 6**) and 49 FA related nodes (**Supplementary Table 7**). Currently, the reference pathways in iLipidome cover generic biosynthetic routes in human, mouse, and rat and not tissue-specific or cell type-specific lipid biosynthetic routes. For each target lipid, we search all possible routes to synthesize it with the *all_simple_paths()* function of R igraph package. The metabolites in the routes are considered as substructure candidates. Note that the source vertex differs according to the FA or lipid class metabolism pathway each lipid belongs to (e.g., G3P for GPLs, glycerolipids, and ether lipids; serine and palmitoyl-CoA for sphingolipids; alanine and palmitoyl-CoA for deoxysphingolipids; acetyl-CoA, linoleic acid (omega-6), and alpha-linolenic acid (omega-3) for all FAs).

Fatty acid substructure decomposition. FA decomposition generally follows three major biosynthetic pathways: non-essential FA biosynthesis, omega-6 FA biosynthesis, and omega-3 FA biosynthesis. Since some of our lipidomics datasets don't provide exact double bond position, a FA could be mapped to multiple targets (e.g., 18:3 for omega-6 and omega-3 FAs). For these targets, the expression is generally equally distributed and the pathway search will be done separately. Among them, 18:2 FA and 20:4 FA cross different biosynthetic pathways and account for a large proportion in total FA as LA (18:2 omega-6) and AA (20:4 omega-6). Since they will have a huge impact in substructure computation, we identify them as their major form instead of equal distribution. Besides, palmitate (16:0) is the first FA produced in de novo FA biosynthesis⁶¹. The FAs shorter than it may be the intermediates or come from exogenous sources. Therefore, we separate them to endogenous and exogenous pathways in the beginning. Only those in the endogenous pathways will be further decomposed. Finally, in the DHA dataset, we did not process DHA and its neighbor FAs (22:5 and 24:6 omega-3) to control the artifacts owing to supplementation.

Lipid class and species substructure decomposition. The iLipidome algorithm can decompose lipid species into lipid class and species substructures based on the lipid pathways we collected. For the lipid class substructure, our reference lipid biosynthetic network composed of lipid classes allows us to directly decompose them. The lipid species substructures, on the other hand, further require the FA information encoded in lipid species and the LIPID MAPS database²³ to decompose. Thus, substructures can represent precursors or intermediates in potential biosynthetic routes to synthesize certain lipid species and be used to reconstruct the following lipid molecular network. There are four major steps for processing FA information of the lipid species substructures from lipid biosynthetic pathways. Firstly, if the substructures in one pathway have the same FA number as the target species (e.g., DAG in PC pathway and both of them have two FAs), we assigned the same FA composition of the target to all these substructures (e.g., DAG 16:0-18:1 for PC 16:0-18:1). Secondly, for those substructures S with less FA number (e.g., LPA in PC pathway but it only has one FA), we find the lipid class k closest to the target species and list all possible FA combinations of it S_{k_1, k_2, \dots, k_n} based on the target (LPA 16:0 and LPA 18:1 for PC 16:0-18:1). Only the substructures S_{k_1, k_2, \dots, k_n} appearing in lipidomics data will be kept, which allows us to control the number of potential substructures. If none of the substructures match, we retain all of them, each of which represents one biosynthetic route. Besides, the FA combinations defined above will apply to the other substructures in S with the same FA number. Third, for the substructures with more FA number (e.g., PC in LPC pathway), we also find the closest lipid class and select those with the same FA as the target species in lipidomics data (PC 16:0-18:1 for LPC 16:0). If there is no match, we will remove the FA information of the substructure (PC for LPC 16:0). Fourth, some datasets might not provide the exact FA identity for their lipids (e.g., PC 34:1 instead of PC 16:0-18:1), which will cause difficulties in processing the substructures with different FA number. To address this problem, target lipid species with unknown FA composition are matched to the LIPID MAPS Structure Database (LMSD)²³ to find all possible FA combinations of them (PC 16:0-18:1, PC 16:1-18:0, etc.). With the FA information, we can decompose these lipids via the approach mentioned above. We note that the imputation of fatty acyl compositions for the target species is only performed to identify FA information for the substructures with different FA number (e.g., LPA 16:0 and LPA 18:1 for PC 34:1). The substructures with same FA number will still remain the original FA form as PC 34:1 and DAG 34:1. Furthermore, if the FA composition of the closest substructure is unknown in lipidomics data (e.g., how to find PC 34:1 in LPC 16:0 pathway), we will search the candidate species with the same lipid class as the substructures (PC) and the same FA as the target species (16:0) using the LMSD database. The FA information of candidate species in the database can be summed (e.g., PC 16:0-16:1 and PC 16:0-18:1 to PC 32:1 and PC 34:1) and used to screen in the lipid profiles so that we have more confidence to build the connections between LPC 16:0 and PC 32:1 or PC 34:1. Finally, we do not consider the exact position of FAs

in relation to the glycerol backbone (sn1 or sn2) except for the ether lipids. To summarize, these processes enable us to precisely identify [the substructures](#) and connect biosynthetic pathways using available information in a lipid profile and pave the way for subsequent analyses.

Substructure extraction and computing substructure profiles

To identify the potential source of synthesized lipids and correct the impact of exogenous lipid uptake, the substructures in each route were mapped with their fold change from raw lipidomics data and were extracted through a backpropagated process. From the target FA or lipid species, the checking process along the biosynthetic route would not stop until it meets a substructure with an opposite fold change. One exception is the endogenous biosynthesis pathway for FAs in the upstream of palmitate. Because they are synthesized as a group, we do not check their fold change and keep all substructures. Only those passing the examination are selected with their weights initialized as 1. If there is more than one biosynthetic route for a FA or lipid species, the weight is equally divided and distributed to each path. Because we allow the substructures that did not appear in our dataset to pass the backpropagation, a threshold to control the percentage of missing data in one pathway is set to reduce artifacts. Typically, the threshold from 70% to 100% can be adjusted according to total lipid class number in each dataset. If the proportion of missing substructures exceeds the threshold in one biosynthetic route, the target FA or lipid species will not be decomposed to substructures. [The checking process of fold changes and the threshold of missing substructures can prevent unlimited backpropagation of substructures and over-weighting certain pathways.](#) To construct a substructure matrix, the substructures with their weights from each FA or lipid species are matched to a list composed of all substructures extracted from a dataset. A substructure matrix encoding the frequency of each substructure is then multiplied by the lipid profile to produce the resulting substructure profile, which records the abundance of each substructure and can be used to reconstruct the biosynthetic network. [We note that substructure profiles are calculated using a direct add-up function for selected substructures regardless of data source. Although similar method has been widely used to analyze lipid features, such as double bond and chain length, in many studies^{3,34,62}, users should consider whether their data are absolute quantification and suitable for this process given that different lipid classes have different ionization tendencies depending on their headgroup.](#)

Pathway scoring method

For any two nodes in the biosynthetic network, we search all possible pathways using the *all_simple_paths()* function from the R igraph package. We then explore the active and suppressed ones in the pathway set using the method adapted from⁶³. Specifically, we transform the p-value p_i of each node i in a pathway into z-score, which will be used to calculate the pathway score. More specifically, given that t-distribution can be approximated by a normal distribution, each p-value is converted to a z-score through $z_i = CDF^{-1}\left(1 - \frac{1}{2} \times p_i\right)$, where CDF is the cumulative distribution function. If the fold change between two experimental conditions is lower than 1, we add a minus sign to z_i . For a random data, z_i will follow a standard normal distribution. To calculate the score of a pathway A of n nodes, we summed all z_i in the pathway by taking $z_A = \frac{1}{\sqrt{n}} \sum_{i \in A} z_i$. Because the variance of a sum is the sum of the variances for independent random variables, z_A will also follow a standard normal distribution if the z_i are independently drawn from a standard normal distribution. A high z_A indicates an active pathway while a low z_A corresponds to a suppressed pathway. [In addition, the length \(n\) of pathway is considered in the function helps us not over-weight certain pathways.](#) Further, to calibrate z_A against the background distribution, we adopt a Monte Carlo approach to randomly select n nodes and compute their scores z_A' . The values are used to derive the mean $\mu_{A'}$ and standard deviation $\sigma_{A'}$ for each n . z_A is corrected by these estimates to produce a final path score $s_A = \frac{z_A - \mu_{A'}}{\sigma_{A'}}$. With this calibration, we can reduce noise and guarantee to have $\mu = 0$ and $\sigma = 1$ for a randomized dataset. To determine if a pathway is significantly activated or suppressed, we choose the critical value 1.96 as the threshold based on the significant level (p-value<0.05) from a two-tailed test. Consequently, $s_A > 1.96$ and $s_A < -1.96$ represent the significantly activated and suppressed pathways, respectively.

Identify representative pathways

For all pathways found in the biosynthetic network, we use two methods according to FA or lipid species to classify them into different representative pathways. At first, pathways are separated to up-regulation or down-regulation based on the signs of their scores. We then rank their importance using the absolute value of path scores, followed by a pathway similarity search from top to bottom. The pathway with the largest score is assigned as the first representative pathway. In the FA analysis, we sequentially compute the proportion of overlapping edges between the candidate pathway and the representative pathway set. If the overlapping proportion of a pathway is over 50% with one representative pathway, it will be classified into that one. If there are no representative pathways over 50%, the candidate pathway will be a new representative pathway. On the other hand, because the FA information is inherited in the lipid species through biosynthetic connections, we can find the dominate FA composition with highest frequency (e.g., 16:0-18:1 or 32:0) in each lipid species pathway. Similar to the FA analysis, pathways with the same dominate FA will be categorized to the same representative pathway. If the dominate FA is the first time appear in this process, the pathway will be considered as a new representative pathway. In this study, we typically collected all significant pathways belonging to the top 3 or top 5 representative pathways for constructing the species biosynthetic network.

Reaction scoring

To determine the potential driver reactions or enzymes in the biosynthetic network, we developed a novel algorithm to compute the perturbation score for each reaction (edge). The concept is to compare the ratio of product over reactant and evaluate its statistical significance magnitude across conditions. Firstly, we can compute this ratio in a certain reaction for each sample. Then, the fold change is obtained by the averaged ratio from control group divided by the averaged ratio from experimental group. Next, we use two-tailed Student's t-tests to compute the p-value of ratios between different conditions. With these values, the final perturbation score of reaction is calculated by: $-\log_{10}(p\text{-value}) \times \log_2(\text{fold change})$. The score is considered significant when the p-value lower than 0.05. Genes/Enzymes involved in reactions are labelled based on the KEGG¹⁵ and LIPID MAPS²³ database and other recent studies^{17-19,22}. Note that since these databases and studies only provide information at the lipid class level, the corresponding genes/enzymes in lipid molecular reactions highlighted by iLipidome do not consider lipid species specificities. In our figures, we only chose one of them to represent the reactions.

Hierarchical clustering

Clustering analysis and heatmap visualization was done in R with the package ComplexHeatmap⁶⁴. Each lipid variable was scaled to have mean 0 and standard deviation 1 before clustering. The distance matrix, defined as 1 minus the Spearman correlation, was processed by the complete method. Further, the *cutree()* function was applied to split the samples and FA variables into two groups. To compare the similarity between the clustering results and the original labels, we calculated the adjusted Rand index using the R package GeometricMorphometricsMix^{65,66}, and the statistical significance is further tested by permutation test. FA variables were categorized according to the double bond number with saturated FAs and highly unsaturated FAs (3 or more double bonds) in one group and less unsaturated FAs (1 and 2 double bonds) in the other.

Enrichment analysis

We performed two different enrichment methods to explore and compare fatty acid signature between substructure and unprocessed data. First, lipid species or substructures were separated into increased and decreased groups according to fold change in advance. For each group, we further categorized them by significance (P-value < 0.05) and FA composition. Finally, enrichment analysis for each FA category was performed using over representation analysis with Fisher's exact test. On the other hand, all lipid species or substructures were only ranked according to their statistics and grouped by FA composition. An enrichment analysis was further performed using gene set enrichment analysis (GSEA) in the 'fgsea' package⁶⁷.

Statistical Analysis

All statistical analyses were performed on R (v 4.1.0) platform. Two-tail Student's t tests or moderated t-tests in the limma package⁶⁸ were used for comparing two experimental groups. We applied Benjamini–Hochberg correction to compute the adjusted p-values for a multiple testing. Statistical methods, numbers of samples, and p-values for each section are also indicated in the figure legends or methods.

Data availability

All lipidome data used in this study were downloaded from the publicly available datasets and have been deposited with their code in GitHub (<https://github.com/LewisLabUCSD/iLipidome>). Source data for each figure/table are also provided with this paper.

Code availability

R files and functions to perform iLipidome analysis and generate all figures in this study are available through GitHub (<https://github.com/LewisLabUCSD/iLipidome>). In addition, we also provide documentation and the example code for users.

Acknowledgements

This work was supported by NIGMS (R35 GM119850) and generous support from the Novo Nordisk Foundation provided to the National Biologics Facility at the Technical University of Denmark (NNF20SA0066621). It was also supported by National Health Research Institute: grant # NHRI-EX111-11110BI, National Science and Technology Council: grant # MOST 111-2320-B-039-011, and China Medical University Hospital: grant # DMR-110-231, and # DMR-111-118. WJL was supported by the Elite Cultivation Scholarship Program Sponsored by China Medical University.

References

- 1 Surma, M. A. *et al.* An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. *European journal of lipid science and technology : EJLST* **117**, 1540-1549, doi:10.1002/ejlt.201500145 (2015).
- 2 Bi, J. *et al.* Oncogene Amplification in Growth Factor Signaling Pathways Renders Cancers Dependent on Membrane Lipid Remodeling. *Cell Metabolism* **30**, 525-538.e528, doi:<https://doi.org/10.1016/j.cmet.2019.06.014> (2019).
- 3 Levental, K. R. *et al.* Lipidomic and biophysical homeostasis of mammalian membranes counteracts dietary lipid perturbations to maintain cellular fitness. *Nat Commun* **11**, 1339, doi:10.1038/s41467-020-15203-1 (2020).
- 4 Matthiesen, R. *et al.* Shotgun mass spectrometry-based lipid profiling identifies and

distinguishes between chronic inflammatory diseases. *EBioMedicine* **70**, 103504, doi:10.1016/j.ebiom.2021.103504 (2021).

Lin, W.-J. *et al.* LipidSig: a web-based tool for lipidomic data analysis. *Nucleic Acids Research* **49**, W336-W345, doi:10.1093/nar/gkab419 (2021).

Mohamed, A. & Hill, M. M. LipidSuite: interactive web server for lipidomics differential and enrichment analysis. *Nucleic Acids Research* **49**, W346-W351, doi:10.1093/nar/gkab327 (2021).

Pellegrino, R. M. *et al.* LipidOne: user-friendly lipidomic data analysis tool for a deeper interpretation in a systems biology scenario. *Bioinformatics*, btab867, doi:10.1093/bioinformatics/btab867 (2021).

Hyötyläinen, T. & Orešič, M. Optimizing the lipidomics workflow for clinical studies--practical considerations. *Anal Bioanal Chem* **407**, 4973-4993, doi:10.1007/s00216-015-8633-2 (2015).

Yan, B. *et al.* Characterization of the Lipidomic Profile of Human Coronavirus-Infected Cells: Implications for Lipid Metabolism Remodeling upon Coronavirus Replication. **11**, 73 (2019).

Gaud, C. *et al.* BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS. *F1000Res* **10**, 4, doi:10.12688/f1000research.28022.1 (2021).

Nguyen, A., Rudge, S. A., Zhang, Q. & Wakelam, M. J. Using lipidomics analysis to determine signalling and metabolic changes in cells. *Current opinion in biotechnology* **43**, 96-103, doi:10.1016/j.copbio.2016.10.003 (2017).

Bao, B. *et al.* Correcting for sparsity and interdependence in glycomics by accounting for glycan biosynthesis. *Nature communications* **12**, 4988, doi:10.1038/s41467-021-25183-5 (2021).

Sekiya, M. *et al.* Polyunsaturated fatty acids ameliorate hepatic steatosis in obese mice by SREBP-1 suppression. *Hepatology* **38**, 1529-1539, doi:10.1016/j.hep.2003.09.028 (2003).

Mauvoisin, D. & Mounier, C. Hormonal and nutritional regulation of SCD1 gene expression. *Biochimie* **93**, 78-86, doi:<https://doi.org/10.1016/j.biochi.2010.08.001> (2011).

Kanehisa, M. & Goto, S. J. N. a. r. KEGG: kyoto encyclopedia of genes and genomes. **28**, 27-30 (2000).

Fabregat, A. *et al.* The reactome pathway knowledgebase. **46**, D649-D655 (2018).

Metherel, A. H. & Bazinet, R. P. Updates to the n-3 polyunsaturated fatty acid biosynthesis pathway: DHA synthesis rates, tetracosahexaenoic acid and (minimal) retroconversion. *Prog Lipid Res* **76**, 101008, doi:10.1016/j.plipres.2019.101008 (2019).

Park, H. G., Park, W. J., Kothapalli, K. S. & Brenna, J. T. The fatty acid desaturase 2 (FADS2) gene product catalyzes $\Delta 4$ desaturation to yield n-3 docosahexaenoic acid and n-6 docosapentaenoic acid in human cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **29**, 3911-3919, doi:10.1096/fj.15-271783

(2015).

Saini, R. K. & Keum, Y.-S. Omega-3 and omega-6 polyunsaturated fatty acids: Dietary sources, metabolism, and significance — A review. *Life Sciences* **203**, 255-267, doi:<https://doi.org/10.1016/j.lfs.2018.04.049> (2018).

Schooneveldt, Y. L., Paul, S., Calkin, A. C. & Meikle, P. J. Ether Lipids in Obesity: From Cells to Population Studies. *Front Physiol* **13**, 841278, doi:10.3389/fphys.2022.841278 (2022).

Paul, S., Lancaster, G. I. & Meikle, P. J. Plasmalogens: A potential therapeutic target for neurodegenerative and cardiometabolic disease. *Prog Lipid Res* **74**, 186-195, doi:10.1016/j.plipres.2019.04.003 (2019).

Duan, J. & Merrill, A. H., Jr. 1-Deoxysphingolipids Encountered Exogenously and Made de Novo: Dangerous Mysteries inside an Enigma. *The Journal of biological chemistry* **290**, 15380-15389, doi:10.1074/jbc.R115.658823 (2015).

Liebesch, G. *et al.* Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *J Lipid Res* **61**, 1539-1555 (2020).

Beloribi-Djefafli, S., Vasseur, S. & Guillaumond, F. Lipid metabolic reprogramming in cancer cells. *Oncogenesis* **5**, e189, doi:10.1038/onsis.2015.49 (2016).

Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics (Oxford, England)* **18**, S233-S240, doi:10.1093/bioinformatics/18.suppl_1.S233 %J Bioinformatics (2002).

Yahagi, N. *et al.* A crucial role of sterol regulatory element-binding protein-1 in the regulation of lipogenic gene expression by polyunsaturated fatty acids. *The Journal of biological chemistry* **274**, 35840-35844, doi:10.1074/jbc.274.50.35840 (1999).

Metherel, A. H., Irfan, M., Klingel, S. L., Mutch, D. M. & Bazinet, R. P. Compound-specific isotope analysis reveals no retroconversion of DHA to EPA but substantial conversion of EPA to DHA following supplementation: a randomized control trial. *Am J Clin Nutr* **110**, 823-831, doi:10.1093/ajcn/nqz097 (2019).

Suzuki-Kemuriyama, N. *et al.* Different Effects of Eicosapentaenoic and Docosahexaenoic Acids on Atherogenic High-Fat Diet-Induced Non-Alcoholic Fatty Liver Disease in Mice. *PLoS One* **11**, e0157580-e0157580, doi:10.1371/journal.pone.0157580 (2016).

Metherel, A. H. *et al.* Two weeks of docosahexaenoic acid (DHA) supplementation increases synthesis-secretion kinetics of n-3 polyunsaturated fatty acids compared to 8 weeks of DHA supplementation. *The Journal of Nutritional Biochemistry* **60**, 24-34, doi:<https://doi.org/10.1016/j.jnutbio.2018.07.002> (2018).

Sassa, T. & Kihara, A. Metabolism of very long-chain Fatty acids: genes and pathophysiology. *Biomol Ther (Seoul)* **22**, 83-92, doi:10.4062/biomolther.2014.017 (2014).

Guillou, H., Zadavec, D., Martin, P. G. P. & Jacobsson, A. The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Progress in*

Lipid Research **49**, 186-199, doi:<https://doi.org/10.1016/j.plipres.2009.12.002> (2010).

32 Wang, Y. *et al.* Tissue-specific, nutritional, and developmental regulation of rat fatty acid
 33 elongases. *J Lipid Res* **46**, 706-715, doi:10.1194/jlr.M400335-JLR200 (2005).

33 Yamashita, A. *et al.* Acyltransferases and transacylases that determine the fatty acid
 34 composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian
 35 cells and model organisms. *Prog Lipid Res* **53**, 18-81, doi:10.1016/j.plipres.2013.10.001
 36 (2014).

34 Levental, K. R. *et al.* ω -3 polyunsaturated fatty acids direct differentiation of the membrane
 35 phenotype in mesenchymal stem cells to potentiate osteogenesis. *Sci Adv* **3**, eaao1193,
 36 doi:10.1126/sciadv.aao1193 (2017).

35 Nakanishi, H. *et al.* Cloning and characterization of mouse lung-type acyl-
 36 CoA:lysophosphatidylcholine acyltransferase 1 (LPCAT1). Expression in alveolar type II
 37 cells and possible involvement in surfactant production. *The Journal of biological chemistry*
 38 **281**, 20140-20147, doi:10.1074/jbc.M600225200 (2006).

36 Poss, A. M. *et al.* Machine learning reveals serum sphingolipids as cholesterol-independent
 37 biomarkers of coronary artery disease. *The Journal of clinical investigation* **130**, 1363-1376,
 38 doi:10.1172/JCI131838 (2020).

37 Gerl, M. J. *et al.* Machine learning of human plasma lipidomes for obesity estimation in a
 38 large population cohort. *PLoS Biol* **17**, e3000443, doi:10.1371/journal.pbio.3000443 (2019).

38 Corn, K. C., Windham, M. A. & Rafat, M. Lipids in the tumor microenvironment: From
 39 cancer progression to treatment. *Prog Lipid Res* **80**, 101055,
 40 doi:10.1016/j.plipres.2020.101055 (2020).

39 Bjermo, H. *et al.* Effects of n-6 PUFAs compared with SFAs on liver fat, lipoproteins, and
 40 inflammation in abdominal obesity: a randomized controlled trial. *Am J Clin Nutr* **95**, 1003-
 41 1012, doi:10.3945/ajcn.111.030114 (2012).

40 Imamura, F. *et al.* Effects of Saturated Fat, Polyunsaturated Fat, Monounsaturated Fat, and
 41 Carbohydrate on Glucose-Insulin Homeostasis: A Systematic Review and Meta-analysis of
 42 Randomised Controlled Feeding Trials. *PLoS Med* **13**, e1002087,
 43 doi:10.1371/journal.pmed.1002087 (2016).

41 Marklund, M. *et al.* Biomarkers of Dietary Omega-6 Fatty Acids and Incident Cardiovascular
 42 Disease and Mortality. *Circulation* **139**, 2422-2436, doi:10.1161/circulationaha.118.038908
 43 (2019).

42 Wang, B. *et al.* Metabolism pathways of arachidonic acids: mechanisms and potential
 43 therapeutic targets. *Signal Transduction and Targeted Therapy* **6**, 94, doi:10.1038/s41392-
 44 020-00443-w (2021).

43 Zhang, T., Zhao, J. V. & Schooling, C. M. The associations of plasma phospholipid
 44 arachidonic acid with cardiovascular diseases: A Mendelian randomization study.
 45 *EBioMedicine* **63**, 103189, doi:10.1016/j.ebiom.2020.103189 (2021).

Zu, L., Guo, G., Zhou, B. & Gao, W. Relationship between metabolites of arachidonic acid and prognosis in patients with acute coronary syndrome. *Thrombosis Research* **144**, 192-201, doi:<https://doi.org/10.1016/j.thromres.2016.06.031> (2016).

Chung, H.-K. *et al.* Plasma phospholipid arachidonic acid and lignoceric acid are associated with the risk of cardioembolic stroke. *Nutrition Research* **35**, 1001-1008, doi:<https://doi.org/10.1016/j.nutres.2015.09.007> (2015).

Sun, Y. *et al.* Plasma fatty acids, oxylipins, and risk of myocardial infarction: the Singapore Chinese Health Study. *J Lipid Res* **57**, 1300-1307, doi:10.1194/jlr.P066423 (2016).

Kagan, V. E. *et al.* Oxidized arachidonic and adrenic PEs navigate cells to ferroptosis. *Nature chemical biology* **13**, 81-90, doi:10.1038/nchembio.2238 (2017).

Zou, Y. *et al.* Plasticity of ether lipids promotes ferroptosis susceptibility and evasion. *Nature* **585**, 603-608, doi:10.1038/s41586-020-2732-8 (2020).

Fang, X. *et al.* Ferroptosis as a target for protection against cardiomyopathy. *Proceedings of the National Academy of Sciences* **116**, 2672-2680, doi:10.1073/pnas.1821022116 (2019).

Wu, X., Li, Y., Zhang, S. & Zhou, X. Ferroptosis as a novel therapeutic target for cardiovascular disease. *Theranostics* **11**, 3052-3059, doi:10.7150/thno.54113 (2021).

Chen, X. *et al.* Ferroptosis and cardiovascular disease: role of free radical-induced lipid peroxidation. *Free Radic Res* **55**, 405-415, doi:10.1080/10715762.2021.1876856 (2021).

Wang, G. *et al.* Lung cancer scRNA-seq and lipidomics reveal aberrant lipid metabolism for early-stage diagnosis. *Sci Transl Med* **14**, eabk2756, doi:10.1126/scitranslmed.abk2756 (2022).

Ecker, J. *et al.* The Colorectal Cancer Lipidome: Identification of a Robust Tumor-Specific Lipid Species Signature. *Gastroenterology* **161**, 910-923.e919, doi:10.1053/j.gastro.2021.05.009 (2021).

Kim, H.-J., Takahashi, M. & Ezaki, O. Fish Oil Feeding Decreases Mature Sterol Regulatory Element-binding Protein 1 (SREBP-1) by Down-regulation of SREBP-1c mRNA in Mouse Liver: A POSSIBLE MECHANISM FOR DOWN-REGULATION OF LIPOGENIC ENZYME mRNAs*. *Journal of Biological Chemistry* **274**, 25892-25898, doi:<https://doi.org/10.1074/jbc.274.36.25892> (1999).

Domenichiello, A. F., Chen, C. T., Trepanier, M.-O., Stavro, P. M. & Bazinet, R. P. Whole body synthesis rates of DHA from α -linolenic acid are greater than brain DHA accretion and uptake rates in adult rats. *J Lipid Res* **55**, 62-74, doi:10.1194/jlr.M042275 (2014).

Wang, Y. *et al.* Regulation of hepatic fatty acid elongase and desaturase expression in diabetes and obesity. *J Lipid Res* **47**, 2028-2041, doi:10.1194/jlr.M600177-JLR200 (2006).

Spiegel, A. *et al.* A set of gene knockouts as a resource for global lipidomic changes. *Scientific Reports* **12**, 10533, doi:10.1038/s41598-022-14690-0 (2022).

Hu, C., Duan, Q. & Han, X. Strategies to Improve/Eliminate the Limitations in Shotgun Lipidomics. *Proteomics* **20**, e1900070-e1900070, doi:10.1002/pmic.201900070 (2020).

Rose, T. D. *et al.* Lipid network and moiety analysis for revealing enzymatic dysregulation and mechanistic alterations from lipidomics data. *Brief Bioinform*, doi:10.1093/bib/bbac572 (2023).

Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* **46**, W486-W494, doi:10.1093/nar/gky310 (2018).

Smith, S., Witkowski, A. & Joshi, A. K. Structural and functional organization of the animal fatty acid synthase. *Progress in lipid research* **42**, 289-317 (2003).

Purdy, J. G., Shenk, T. & Rabinowitz, J. D. Fatty acid elongase 7 catalyzes lipidome remodeling essential for human cytomegalovirus replication. *Cell Rep* **10**, 1375-1385, doi:10.1016/j.celrep.2015.02.003 (2015).

Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18 Suppl 1**, S233-240, doi:10.1093/bioinformatics/18.suppl_1.s233 (2002).

Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849, doi:10.1093/bioinformatics/btw313 (2016).

Qannari, E. M., Courcoux, P. & Faye, P. Significance test of the adjusted Rand index. Application to the free sorting task. *Food Quality and Preference* **32**, 93-97, doi:<https://doi.org/10.1016/j.foodqual.2013.05.005> (2014).

Rand, W. M. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association* **66**, 846-850, doi:10.2307/2284239 (1971).

Korotkevich, G. *et al.* Fast gene set enrichment analysis. *BioRxiv*, 060012 (2021).

Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47-e47 (2015).