HHS Public Access

Author manuscript

Environ Int. Author manuscript; available in PMC 2022 June 14.

Published in final edited form as:

Environ Int. 2022 June ; 164: 107240. doi:10.1016/j.envint.2022.107240.

CCDB: A database for exploring inter-chemical correlations in metabolomics and exposomics datasets

Dinesh Kumar Barupal[*], **Priyanka Mahajan**, **Sadjad Fakouri-Baygi**, **Robert O. Wright**, **Manish Arora**, **Susan L. Teitelbaum**

Department of Environmental Medicine and Public Health, Institute for Exposomic Research, Icahn School of Medicine at Mount Sinai, 17 E 102nd St, CAM Building, New York 10029, USA

Abstract

Inter-chemical correlations in metabolomics and exposomics datasets provide valuable information for studying relationships among chemicals reported for human specimens. With an increase in the number of compounds for these datasets, a network graph analysis and visualization of the correlation structure is difficult to interpret. We have developed the Chemical Correlation Database (CCDB), as a systematic catalogue of inter-chemical correlation in publicly available metabolomics and exposomics studies. The database has been provided via an online interface to create single compound-centric views. We have demonstrated various applications of the database to explore: 1) the chemicals from a chemical class such as Per- and Polyfluoroalkyl Substances (PFAS), polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), phthalates and tobacco smoke related metabolites; 2) xenobiotic metabolites such as caffeine and acetaminophen; 3) endogenous metabolites (acyl-carnitines); and 4) unannotated peaks for PFAS. The database has a rich collection of 35 human studies, including the National Health and Nutrition Examination Survey (NHANES) and high-quality untargeted metabolomics datasets. CCDB is supported by a simple, interactive and user-friendly web-interface to retrieve and visualize the inter-chemical correlation data. The CCDB has the potential to be a key computational resource in metabolomics and exposomics facilitating the expansion of our understanding about biological and chemical relationships among metabolites and chemical exposures in the human body. The database is available at www.ccdb.idsl.me site.

[*] Corresponding author at: CAM Building, 3rd floor, 17 E 102nd St, New York, NY 10029, USA. dinesh.kumar@mssm.edu (D.K. Barupal).

Author contributions

DKB and ST conceptualize the study. DKB and SFB prepare the data and conducted data analysis. PM and DKB designed the web-interfaces and database architecture. All authors contributed to the manuscript.

CRediT authorship contribution statement

Dinesh Kumar Barupal: Conceptualization, Methodology, Data curation, Software, Visualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Priyanka Mahajan:** Data curation, Software. **Sadjad Fakouri-Baygi:** Software, Methodology, Data curation. **Robert O. Wright:** Writing – original draft, Writing – review & editing. **Manish Arora:** Writing – original draft, Writing – review & editing. **Susan L. Teitelbaum:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2022.107240.

Keywords

Metabolomics; Inter-chemical correlation; Exposomics; Biomonitoring; Metabolic pathways; NHANES; Database; Software

1. Introduction

Combined exposures to millions of different chemicals and its impact on the health and development of human body is a major component of the exposome (Vermeulen et al., 2020). The chemical exposome is made up of nutrients and environmental non-food chemicals, consisting of natural and synthetic exogenous compounds (Barupal and Fiehn, 2019; Matta et al., 2020; Rappaport et al., 2014). After entering the body, through biotransformation they also become part of the metabolome, which includes metabolic end products of the host and its commensal microbiota. This chemical space (e.g. industrial chemicals, nutrients, drugs, and bioactive internal molecules such as hormones and oxylipins) has significant influence on health trajectories and chronic health outcomes and is implicated in all diseases, including cancer as well as neurological, cardiovascular, and respiratory diseases (Drouin-Chartier et al., 2021; Jobard et al., 2021; Loftfield et al., 2021; Needham et al., 2021; Nemet et al., 2020; Nymand Ennis et al., 2019; Peters et al., 2021; Petrick et al., 2020; Schillemans et al., 2021; Tahir et al., 2021; Vangipurapu et al., 2020). Emerging evidence demonstrates that the scale, magnitude, and structural diversity (Guha et al., 2016; Rappaport et al., 2014) of the internal chemical space is vast and that many chemicals could be classified together because they are structurally and functionally related to each other (Paul-Friedman et al., 2019; Richard et al., 2021; Zimmermann et al., 2019). A systematic understanding and cataloging of targeted and untargeted analyses of small molecules measured in biospecimens is needed, as such datasets are critical to translate the information gathered from exposomics and metabolomics projects (Hendrix et al., 2015). These key datasets include: 1) population-scale biomonitoring surveys; 2) targeted analysis of multiple analytes in hypothesis-driven studies (typically 10–100); and 3) untargeted analysis of thousands of chemicals using a high-resolution mass spectrometry instrument (Barupal et al., 2021a; David et al., 2021). They cover key high priority exposome chemicals (Barupal et al., 2021b) including carcinogens (Hecht et al., 2016; Park et al., 2021), endocrine disrupters (Kassotis et al., 2020) and industry chemicals (Shearer et al., 2021). These core datasets support different statistical and bioinformatics analyses to reveal novel risk factors, hidden metabolic pathways, detrimental exposures and biomarkers for disease.

Computing the correlation coefficient using intensities of two chemicals is a fundamental statistical approach classically used to study enzyme kinetics (Frieden et al., 1976) and biotransformation (Hoffman et al., 1990). For modern multi-analyte targeted and untargeted assays, a pair-wise correlation matrix among detected chemicals is computed for almost every study because this matrix can be used to assess chemical clustering (Barupal et al., 2019a), peak annotation (DeFelice et al., 2017), heatmaps (Shen et al., 2020), and correlation network visualization (Barupal et al., 2019a). Correlation among gene expression data is often interpreted as evidence of a co-regulatory pathway such as a common transcription factor that controls expression of a group of genes (Obayashi et al., 2019;

Yin et al., 2021). As a corollary, with chemicals, correlation can reflect common exposure origins (Edmands et al., 2015) as well as chemical disposition, such as absorption pathways, biotransformation (Frederiksen et al., 2010; Saravanabhavan et al., 2013) and elimination as seen in drugs and their metabolic products (Guo et al., 2020; Guthrie et al., 2019). For exposomic projects, the probable interpretation of inter-chemical correlations is summarized in Fig. 1. The biological interpretation covers both kinetics (i.e. the metabolic fate of a chemical (Cohen et al., 2018)) and dynamics (i.e. the toxic effect of chemical exposure). The system connects to key metabolic pathways (Chen et al., 2020), and creates logical groupings of similar exposures in a chemical class (Barupal et al., 2019a). It can also indicate that two chemicals share an exposure source, such as occupation, consumer products (Stanfield et al., 2021), or food (McKillop et al., 2021). Despite the utility and application of inter-chemical correlation data, a database of these inter-chemical correlations has not yet been developed.

Metabolomic correlation network analyses show that chemically similar compounds and compounds belonging to the same pathway tend to show a higher correlation coefficient (Li et al., 2017; Liang et al., 2020; Toledo et al., 2017). However, creating and analyzing those networks for large and comprehensive metabolomics datasets that often have over ten thousand reported peaks is computationally challenging. It is even more difficult to create and analyze such network graphs for metabolomics datasets that are generated using multiple LC/GC assays (e.g. reverse phase (RP) and hydrophilic interaction liquid chromatography (HILIC) modes) for hundreds of samples (Barupal et al., 2019b). There is a need to catalogue these correlations in a systematic database for mining them in various interpretational contexts.

Herein, we describe a new database, CCDB, which catalogues pairwise inter-chemical correlations from publicly available metabolomics and exposomics studies. It is the largest database of pairwise correlations to date and provides new opportunities for interpreting metabolomics datasets for structural and biological relationships. The database is publicly available at www.ccdb.idsl.me.

2. Methods

2.1. Selection of studies

Table 1 provides the list of 35 studies and the details about the number of compounds and samples. For the development of the database, we constrained our approach to human specimen studies having at least 50 samples. To include a study in the CCDB, the data were reformatted into CCDB Excel template (SI File 1). The template requires three sheets 1) “data_dictionary” which contains the metadata for annotated and unannotated compounds 2) “data_matrix” which contains the intensity data for all peaks and 3) “sample_metadata” which contains the information about each sample. If data from different chromatography and ionization modes were available, data were stacked in the “data_dictionary” and “data_matrix” sheets. If data were not scaled or normalized, we applied a log2 transformation before computing the correlation.

2.2. Processing of untargeted metabolomics studies

Only untargeted liquid chromatography high resolution mass spectrometry studies were selected. For each selected untargeted study (Table 1), we searched for a set of data types in the EBI-MetaboLights and Metabolomics WorkBench repositories. The set included 1) intensity values for annotated peaks 2) intensity values for un-annotated peaks 3) sample metadata and 4) metadata for the annotated peaks. For each reported peak, information about the analysis mode (reverse phase or hydrophilic interaction liquid chromatography) mass to charge ratio and retention time were collected in the "data_dictionary" tab in the CCDB template (https://github.com/idslme/chemcordb/blob/main/MTBSL204_INPUT.xlsx).

2.3. Processing of the National health and Nutrition Examination Survey (NHANES) data

Laboratory data for continuous variables were downloaded from the NHANES website (https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory) in the SAS export format (.XPT). Variables that reflected a chemical entity were used for calculating the inter-chemical correlation data (Table S1). Data files were imported in the R programming language and merged using the NHANES SEQN number as the linking identifier. NHANES data were used for computing correlation statistics without any transformation, normalization and scaling. Survey design weights do not affect the inter-chemical correlations, so they were not taken into account.

2.4. Processing of datasets generated by Metabolon Inc. platform

Metabolomics datasets generated by the Metabolon Inc. company available in the supplementary section of a published article (Germain et al., 2020) or via metabolomics repositories were included in CCDB. The company provides datasets with up to 2,000 high-confidence chemicals reported for blood and urine specimens. If these data were not scaled or normalized, we applied a log2 transformation before computing the correlation. For the CCDB input format, only metabolite names reported in the table were used in the "data dictionary" tab of the CCDB format.

2.5. Correlation calculation

The Pearson correlation coefficient was used for computing a pairwise correlation among reported peaks within each study using the cor function available in the WGCNA R package (Langfelder and Horvath, 2008). A correlation between two intensity vectors was computed only if they had at least 10% non-zero values. We did not compute any p-values for the correlation statistics given that our goal was to create a database of inter-chemical correlations, not to find a biomarker of phenotype. Therefore, the application of a false discovery rate correction was not required. If p-values were computed, they would be expected to be extremely small considering the large sample sizes of the selected studies.

Overall average detection rate across all studies were 60% of above (Table S2). However, it is common for human biospecimen studies that several compounds, especially exposure-related are detected only in a fraction of samples in a study. For example, Fluoro-phenoxybenzoic acid was found only in 170/2694 (6.3%) samples (https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/UPHOPM_E.htm#URD4FPLC). Therefore, for NHANES we have

used a criteria that a compound must be detected in at least 100 samples to be included in the computation of inter-chemical correlations.

2.6. CCDB indexing

For each selected study, a unique name directory was created in a webserver's filesystem and the pairwise correlation data were saved inside the corresponding directory. For each compound, a vector of correlation against all other chemicals in the study were computed and then stored in the file system. For the naming convention, a distinct study-specific identifier was assigned to each reported chemical. Linux operating system Ubuntu 20.04 was used for the webserver.

2.7. Online interface and querying the CCDB

The online front interface was developed using the AngularJS 1.5 javascript framework and bootstrap. On the backend, a nginx proxy server was used to route the web requests to the data indexed in the CCDB. The opencpu framework (https://www.opencpu.org/) in R was used as a middleware to process each web request. For biomonitoring (NHANES), Metabolon Inc's datasets and untargeted full-scan datasets, three separate types of web-interfaces were developed. For visualizing the correlation data online, Vis.JS javascript library was utilized. If there are more than 100 hits that pass the correlation threshold only the first hundred hits are visualized in the compound centric network and full data were provided as Cytoscape network file.

For each study, a specific web-address was created (Table S3). For NHANES data, the query parameter is a variable identifier provided in the Table S1. For Metabolon Inc's datasets, chemical names were utilized. For full-scan untargeted datasets, m/z with a mass tolerance was used to retrieve the matched peaks in the database. To obtain putative annotation hits, m/z values were matched against a list of compounds that have been associated with a published paper.

2.8. Chemical similarity enrichment (ChemRICH) analysis

ChemRICH is a database independent and p-value distribution-based approach to rank the chemical sets that are associated with an exposure (Barupal and Fiehn, 2017). As an example, cor.test function in R was used to obtain p-values and estimates for the correlation between Perfluorooctanoic acid (PFOA) intensities and other chemicals from the study IDSLCCDB0001 (Needham et al., 2021). These results and the subpathway information made available by the Metabolon Inc's report were used as an input for the chemical similarity enrichment analysis using the ChemRICH software (Barupal and Fiehn, 2017).

2.9. Data and code availability

All data and resources are available at www.ccdb.idsl.me site. Core scripts to compute the inter-chemical correlation data from biomonitoring and metabolomics studies have been provided at https://github.com/idslme/CCDB.

3. Results

3.1. CCDB is a comprehensive database of inter-chemical correlations for human biospecimens

To build a comprehensive database of inter-chemical correlations in human biospecimens, we found three types of chemical analyses that should be covered. These included 1) biomonitoring surveys that have used a targeted analysis for chemical panels 2) metabolomics datasets having structurally annotated peaks 3) untargeted LC/GC-HRMS datasets having primarily unannotated peaks. In the first version of the CCDB, 35 studies were included (Table 1). The coverage for specimen types was 28 (blood), 3 (urine), 4 (stool). The number of individual participants was 107,258 for NHANES with 607 laboratory measurement variables. For 18 datasets that were generated by Metabolon Inc, the sample size ranged from 52 to 1,336 with the reported peak count ranging between 517 and 1989. For 16 full-scan untargeted LC-HRMS studies, the sample size ranged between 51 and 781 with a reported peak count of 459 to 81867, and 8 studies had reported only unannotated peaks that were referenced using m/z and retention time values. To update the database, we plan to regularly screen publicly available datasets in the Metabolomics Workbench, EBI MetaboLights, GNPS-Massive and consortium/cohort specific repositories and supplementary tables for published papers and include the relevant studies in the CCDB database. By covering three types of chemical measurement datasets, CCDB can provide unique opportunities to not only learn about the biological relationships among metabolites, but also prioritize chemicals that are yet to be annotated in untargeted LC/HRMS datasets.

3.2. A large number of inter-chemical correlations were observed in the catalogued studies

To populate the database, pair-wise correlations among reported chemicals were computed for each selected study. A computational pipeline has been established for an efficient indexing of a new dataset in the database. For that, a minimal level of manual curation was needed to prepare the dataset in the required format (See methods). We investigated the prevalence of strong inter-chemical correlations across the catalogued studies. A total of 121.4 million inter-chemical correlations across the studies passed a threshold of 0.6 Pearson coefficient, indicating the large-scale and magnitude of strong correlation patterns that exists among chemical compounds measured for human biospecimens (Fig. 2). More of these correlations were observed for untargeted datasets which had thousands of mostly unannotated peaks. We noticed that endogenous compounds tend to show a higher number of significant correlations in comparison to exogenous and xenobiotic compounds (Fig. S1). This suggested that at a lower correlation threshold level, we can capture new relationships among chemicals that would otherwise be missed if the correlation data is visualized as a network graph created using a stringent threshold.

For example, by a Pearson coefficient cutoff of 0.4, we have noticed a relationship among blood glucose and acyl-choline lipids (Fig. S2) in the study MTBL136 (Stevens et al., 2018) which will be missed on a cutoff of 0.6. This association has been linked with energy disturbance and implicated in diabetes and chronic fatigue disease related studies. This underscores the need to access the correlation data in a flexible and interactive approach so

we can capture both the known and novel types of functional and biological relationships among reported chemicals.

3.3. Dataset type specific web-interfaces provided access to correlation data for both annotated and unannotated compounds

Because a large number of inter-chemical correlations were observed in the selected studies, it was not practical to visualize them all as a global network in Cytoscape network visualization (Shannon et al., 2003) or any other network graph visualization software unless the network graph is created using very stringent correlation thresholds, which will likely miss biological insights. Therefore, we stored all the correlation data for each compound from each study in a web-server's file system. This allows us to readily load the correlation vector in the computer memory without the need to re-calculate them and enabled a faster response time for the online visualization. A network-based visualization highlighted a compound centric view of inter-chemical correlations, which can be updated by different correlation thresholds. A compound centric view was found to be a cleaner, readable and meaningful visualization than creating a network graph of all compounds reported in a study. It enables a focused investigation of a single compound and its chemical and biochemical relationships with other chemicals in a study. Three types of web interfaces were developed to provide a tailored access inter-chemical correlation data for biomonitoring, annotated peaks and unannotated data in metabolomics and exposomics assays (Fig. S3–5). These interfaces enabled queries by chemical names, CAS numbers, NHANES identifiers and mass to charge (m/z) ratio. For untargeted assays, data from different analysis modes were stacked which allowed to find peaks from the same compound in two analysis modes such as an ESI positive and negative or HILIC (+) or RP (+) (Fig. S6). Network data were also provided as Cytoscape network files to enable additional visualization strategies. These simple and flexible web-interfaces allowed a seamless and interactive access to the inter-chemical correlation data for a chemical from a study.

3.4. Compounds from a chemical class correlated strongly with each other in the NHANES biomonitoring dataset

First, we asked if compounds from a known chemical class correlate with each other and can be retrieved by querying a single chemical. We have observed that chemicals from well-recognized environmental exposures PCB, PFC and PAH groups indeed correlated with a representative chemical from these classes (Fig. 3). This probably suggested a common source of exposure for these chemicals. When cotinine, a biomarker of tobacco smoke was queried, it retrieved many other tobacco smoke related chemicals, providing a quick overview of biomarkers of smoke exposures.

This compound-centric retrieval of inter-chemical correlations in the NHANES biomonitoring dataset suggested that chemical exposures with similar structure and origin correlates strongly with each other.

3.5. Stronger correlations among compounds belonging to a chemical class in metabolomics datasets

Next, we investigated if endogenous metabolites from a chemical class correlated with each other in a metabolomics dataset. We queried a ubiquitous endogenous blood metabolite, C-16 carnitine and retrieved its neighbors in the ST002089 study. At the Pearson correlation cutoff of 0.5, we retrieve mostly other saturated and unsaturated carnitines (Fig. 4). However, at the 0.4 Pearson correlation cutoff, we found that carnitines have biochemical relationships with fatty acids and acylcarnitines.

We learned that structurally similar compounds from an endogenous chemical class can have a high correlation coefficient among them, suggesting an enzyme activity that can react on any member of a chemical class, for instance the carnitine palmitoyltransferase I enzyme. As the Pearson correlation cutoff was lowered, we found long-distance biochemical relationships suggesting different chemical classes that may belong to a metabolic pathway, for instance, fatty acids and acylcarnitines. It also highlighted the unidentified metabolites that correlated strongly with C16-carnitine in the Metabolon Inc's report may belong to the acyl-carnitine chemical class. In summary, by modifying the correlation cutoff, the CCDB interface enables retrieval of short and long-distance biochemical relationships in a metabolic network around a single chemical. This can be used for hypothesizing novel biochemical relationships in untargeted metabolomics datasets.

3.6. Products of xenobiotic metabolism

Next, we checked if metabolites of a xenobiotic compound correlate with the parent compound's levels. In the NHANES biomonitoring survey, several metabolites of caffeine strongly correlated with caffeine levels (Fig. 5, upper panel). The same pattern was found in a metabolomics dataset (Fig. S7). Similarly, metabolites of mono-n-butyl phthalate (MnBP), a commonly used plasticizer correlated with structurally and metabolically related chemicals. MnBP also correlated with other phthalate molecules (Fig. 5 lower panel), indicating common exposure sources. It was expected that people exposed to dibutyl phthalates will excrete MnBP and mono-isobutyl phthalate in their urine (Qian et al., 2015). For acetaminophen, a commonly used over the counter pain-reliever drug, its sulfate metabolite was found to be correlating with other acetaminophen metabolites (Fig. S8).

3.7. Putative annotation of peaks in untargeted data by correlation patterns

So far, we have learned from the NHANES and other high quality metabolomics dataset that chemicals within a chemical class or having the same origin or similar pathway tends to show strong correlations. Relying on this information, we explore the untargeted metabolomics datasets to test if m/z values for chemicals from a chemical class show inter-chemical correlations. To test this, we have queried the m/z value 498.9291 for the M–H adduct of perfluorooctanesulfonic acid (PFOS) in reverse phase chromatography data for the ST001430 study. It retrieved three other chemicals on in the correlation cutoff of 0.3, which matched to the M–H adducts for other common PFCs - PFOA and PFHxS (Fig. 6). In another untargeted study ST001231, we found that PFOS correlated with many more PFCs compounds (Fig. 6).

3.8. Metabolic effect of a hazardous chemical - PFOA

Finally, we asked if we could utilize the inter-chemical correlation data to understand metabolic effects of a chemical exposure. Perfluorochemicals (PFCs) are concerning chemicals for public health. They are exclusively synthetic and accumulate in human body overtime. The ubiquitous exposures to them have been under high priority investigations since they may have contributed to the etiology of a range of chronic diseases. Endogenous metabolites that correlate with PFCs exposures may reflect the biological response to these hazardous chemicals. In several of Metabolon Inc's reports, Perfluorooctanoic acid (PFOA) peak was annotated and found to be correlated with many chemicals when we indexed these reports in the CCDB.

Many metabolites that correlated with PFOA levels may belonged to the same pathway or chemical class. Identifying these chemical sets can assist in understanding the systematic metabolic effect of PFOA exposure which can span over multiple metabolic pathways (Fig. 7). Therefore, we have utilized ChemRICH analysis (Barupal and Fiehn, 2017) to identify the PFOA associated chemical sets, which suggested that PFOA exposure has a negative effect on most of the lipid sets except triglycerides (Sen et al., 2022; Sinisalu et al., 2020). PFOA exposure may have also induced the amino acid and tocopherol metabolism pathways. This analysis highlighted that CCDB correlation data can also be used for investigating the metabolic hazardous effect of a chemical exposure of public health concern using a chemical set analysis approach.

4. Discussion

Inter-chemical correlations in biomonitoring, metabolomics and exposomics datasets is a useful source of information to expand our understanding about the relationships between different metabolites, metabolic pathways and the chemical exposures. There is a need to systematicaly catalogue and preserve these correlation patterns in a database to support useful queries. In this paper, we have presented the CCDB database which aims to build a catalogue of inter-chemical correlation in chemical measurement datasets and then provide users access to the correlation data using a web-interface. As of March 2022, the database includes data from from 35 studies covered. We plan to regularly screen literature as well as metabolomics and exposomics repositories to identify additional studies that can be catalogued in the CCDB. The database currently only hosts studies related to human specimens, however given the generic nature of the catalogued data and indexing pipelines, it will be able to incorporate studies of other species or sources as long as data are provided in the required format. We foresee a regular use of the database in the field of metabolomics and exposomics to explore about the biochemical and chemical relationships around a chemical that has been prioritized by a researcher using statistical or by text mining approaches. We believe the CCDB will be a core database resource in these fields where the interpretation of multi-analyte datasets remains a major challenge.

A large number of significant inter-chemical correlations are ubiquitously observed in these core datasets. An obvious question is "what are the reasons behind these correlations"? At present this is a challenging question because these correlations can be interpreted only in the context of known exposure sources, biochemical absorption pathways and

transformation reactions (Barupal et al., 2018). With time, and additional cataloguing of the exposome, the reasons behind these correlations will become more evident. Pathway-centric approaches do not cover many high-priority exposome chemicals, including their chemical classes, source origin and transformation products. The CCDB is designed to address this issue by curating and interpreting inter-chemical correlations in exposomics and metabolomics core datasets while integrating information about functional and structural relationships among chemicals.

In the transcriptomics field, gene correlation or co-expression databases (Lee et al., 2020; Obayashi et al., 2019) have been developed for multiple species and disease conditions. These databases allow the identification of gene function(s) based on the similarity between two gene's expression levels. They have shown that the similarity in expression levels reflect a shared function or regulation in the genetic networks. CCDB is in line with these databases to provide similar resource for chemicals. For the first time, we developed an inter-chemical correlation database to be used for metabolomics and exposomics hypothesis generation and characterization.

Due to the large number of analytes in targeted and untargeted assays, a traditional correlation network graph of all analytes (Kitagawa et al., 2019; Lau et al., 2018) using Cytoscape (Shannon et al., 2003) or similar software would not be meaningful to explore inter-chemical correlation data because the network graphs would be over-crowded requiring a stringent correlation threshold to draw the edges. Instead, we propose to use a single-compound centric network to generate clear and readable networks that are easy to deploy in online interfaces. We suggest that investigators can explore correlation data in this interactive, compound centric way so that novel relationships among chemicals can be readily explored. In this way, CCDB fullfills critical gaps in the mining of metabolomics correlation data.

CCDB can play a role in peak annotation in untargeted metabolomics, because compounds belonging to the same class, metabolic pathway or source origin tends to correlate with each other. By querying a single compound's m/z, we will be able to estimate the chemical class or in some cases the exact identity of a peak, although it will be only based on the MS match against a priority list of chemicals from a database. There is a need to develop further tools to utilize the isotope patterns, MS2 spectra to refine the annotation patterns. For full-scan untargeted datasets, m/z with a mass tolerance will be used to retrieve matched peaks in the database. In untargeted chemical analyses, many inter-chemical correlations are often observed due to non-biological causes. They are useful in annotating peaks in the untargeted dataset with isotope information (Semente et al., 2021), chemical fragments (J Guo et al., 2021), and errors during data processing, such as duplicate peaks. These annotations can be transferred to other untargeted studies with many unidentified peaks, with the logic that pairs of the same compound will show similar inter-chemical correlation irrespective of analysis platform. It was shown in the example for PFCs and caffeine metabolites (Figs. 7 and S8). It is expected that some of the inter-chemical correlations may not be found across multiple studies or may not have the same strength, which can suggest that the underlying regulatory or source mechanisms are operating differently in two studies. These differences can be considered high priority hypothesis.

In the future version of the CCDB, we may include with more tissue types and clinical outcome datasets (open-access) from the HHEAR program and other NIH supported consortiums. This may enable us to highlight the biomedical relevance of a compound-centric correlation network that is created for a phenotype or outcome. Applications of text mining (Barupal et al., 2021b), chemoinformatics and other bioinformatics resources (Barupal et al., 2018) can also be explored to aid in the interpretation of inter-chemical correlations.

5. Conclusions

We describe CCDB, a new key database in the field of metabolomics and exposomics that provides access to fundamental information on the inter-chemical correlations among chemical signals derived from human specimens. The database has a potential to accelerate learning about the chemical and biochemical relationships among reported chemicals. It can be used for prioritizing chemicals, identifying new hypotheses, interpreting metabolomics datasets, annotating peaks in untargeted metabolomics datasets, and for investigating the metabolic effects of a known chemical exposures. Overall, CCDB will start a new wave of database types in the metabolomics and exposomics field that are more interpretive than just a catalogue of information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by NIH grant U2CES026561, P30ES023515, U2CES030859, U2CES026555, R01ES032831, R35ES030435 and UL1TR001433.

References

Abdrabou W, Dieng MM, Diawara A, Sermé SS, Almojil D, Sombié S, Henry NB, Kargougou D, Manikandan V, Soulama I, Idaghdour Y, 2021. Metabolome modulation of the host adaptive immunity in human malaria. Nat. Metab. 3 (7), 1001–1016. [PubMed: 34113019]

Barupal DK, Fiehn O, 2017. Chemical similarity enrichment analysis (chemrich) as alternative to biochemical pathway mapping for metabolomic datasets. Sci. Rep. 7 (1), 14567. 10.1038/s41598-017-15231-w. [PubMed: 29109515]

Barupal DK, Fan S, Fiehn O, 2018. Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets. Curr. Opin. Biotechnol. 54, 1–9. 10.1016/j.copbio.2018.01.010. [PubMed: 29413745]

Barupal DK, Baillie R, Fan S, Saykin AJ, Meikle PJ, Arnold M, Nho K, Fiehn O, Kaddurah-Daouk R, Zetterberg H, 2019a. Sets of coregulated serum lipids are associated with alzheimer's disease pathophysiology. Alzheimers Dement (Amst) 11 (1), 619–627. [PubMed: 31517024]

Barupal DK, Fiehn O, 2019. Generating the blood exposome database using a comprehensive text mining and database fusion approach. Environ. Health Perspect 127 (9), 97008. 10.1289/EHP4713. [PubMed: 31557052]

Barupal DK, Zhang Y, Shen T, Fan S, Roberts BS, Fitzgerald P, et al. , 2019b. A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the international mouse phenotyping consortium. Metabolites 9. 10.3390/metabo9050101.

Barupal DK, Baygi SF, Wright RO, Arora M, 2021a. Data processing thresholds for abundance and sparsity and missed biological insights in an untargeted chemical analysis of blood specimens for exposomics. Front Public Health 9, 653599. 10.3389/fpubh.2021.653599. [PubMed: 34178917]

Barupal DK, Schubauer-Berigan MK, Korenjak M, Zavadil J, Guyton KZ, 2021b. Prioritizing cancer hazard assessments for iarc monographs using an integrated approach of database fusion and text mining. Environ. Int. 156, 106624. 10.1016/j.envint.2021.106624. [PubMed: 33984576]

Bifarin OO, Gaul DA, Sah S, Arnold RS, Ogan K, Master VA, Roberts DL, Bergquist SH, Petros JA, Fernández FM, Edison AS, 2021. Machine learning-enabled renal cell carcinoma status prediction using multiplatform urine-based metabolomics. J. Proteome Res. 20 (7), 3629–3641. [PubMed: 34161092]

Chen Z, Yang T, Walker DI, Thomas DC, Qiu C, Chatzi L, Alderete TL, Kim JS, Conti DV, Breton CV, Liang D, Hauser ER, Jones DP, Gilliland FD, 2020. Dysregulated lipid and fatty acid metabolism link perfluoroalkyl substances exposure and impaired glucose metabolism in young adults. Environ. Int. 145, 106091. 10.1016/j.envint.2020.106091. [PubMed: 32892005]

Clendinen CS, Gaul DA, Monge María.E., Arnold RS, Edison AS, Petros JA, Fernández FM, 2019. Preoperative metabolic signatures of prostate cancer recurrence following radical prostatectomy. J. Proteome Res. 18 (3), 1316–1327. [PubMed: 30758971]

Cohen IV, Cirulli ET, Mitchell MW, Jonsson TJ, Yu J, Shah N, Spector TD, Guo L, Venter JC, Telenti A, 2018. Acetaminophen (paracetamol) use modifies the sulfation of sex hormones. EBioMedicine 28, 316–323. [PubMed: 29398597]

Colicino E, Ferrari F, Cowell W, Niedzwiecki MM, Foppa Pedretti N, Joshi A, Wright RO, Wright RJ, 2021. Non-linear and non-additive associations between the pregnancy metabolome and birthweight. Environ. Int. 156, 106750. 10.1016/j.envint.2021.106750. [PubMed: 34256302]

David A, Chaker J, Price EJ, Bessonneau V, Chetwynd AJ, Vitale CM, Klánová J, Walker DI, Antignac J-P, Barouki R, Miller GW, 2021. Towards a comprehensive characterisation of the human internal chemical exposome: Challenges and perspectives. Environ. Int. 156, 106630. 10.1016/j.envint.2021.106630. [PubMed: 34004450]

DeFelice BC, Mehta SS, Samra S, Čajka T, Wancewicz B, Fahrmann JF, Fiehn O, 2017. Mass spectral feature list optimizer (ms-flo): A tool to minimize false positive peak reports in untargeted liquid chromatography-mass spectroscopy (lcms) data processing. Anal. Chem. 89 (6), 3250–3255. [PubMed: 28225594]

Drouin-Chartier J-P, Hernández-Alonso P, Guasch-Ferré M, Ruiz-Canela M, Li J, Wittenbecher C, Razquin C, Toledo E, Dennis C, Corella D, Estruch R, Fitó M, Eliassen AH, Tobias DK, Ascherio A, Mucci LA, Rexrode KM, Karlson EW, Costenbader KH, Fuchs CS, Liang L, Clish CB, Martínez-González MA, Salas-Salvadó J, Hu FB, 2021. Dairy consumption, plasma metabolites, and risk of type 2 diabetes. Am. J. Clin. Nutr. 114 (1), 163–174. [PubMed: 33742198]

Edmands WMB, Ferrari P, Rothwell JA, Rinaldi S, Slimani N, Barupal DK, Biessy C, Jenab M, Clavel-Chapelon F, Fagherazzi G, Boutron-Ruault M-C, Katzke VA, Kühn T, Boeing H, Trichopoulou A, Lagiou P, Trichopoulos D, Palli D, Grioni S, Tumino R, Vineis P, Mattiello A, Romieu I, Scalbert A, 2015. Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across european countries. Am. J. Clin. Nutr. 102 (4), 905–913. [PubMed: 26269369]

Fitzgerald BL, Molins CR, Islam MN, Graham B, Hove PR, Wormser GP, Hu L, Ashton LV, Belisle JT, 2020. Host metabolic response in early lyme disease. J. Proteome Res. 19 (2), 610–623. [PubMed: 31821002]

Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, Vatanen T, Hall AB, Mallick H, McIver LJ, Sauk JS, Wilson RG, Stevens BW, Scott JM, Pierce K, Deik AA, Bullock K, Imhann F, Porter JA, Zhernakova A, Fu J, Weersma RK, Wijmenga C, Clish CB, Vlamakis H, Huttenhower C, Xavier RJ, 2019. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. 4 (2), 293–305. [PubMed: 30531976]

Frederiksen H, Jorgensen N, Andersson AM, 2010. Correlations between phthalate metabolites in urine, serum, and seminal plasma from young danish men determined by isotope dilution liquid chromatography tandem mass spectrometry. J. Anal. Toxicol. 34, 400–410. 10.1093/jat/34.7.400. [PubMed: 20822678]

Frieden C, Gilbert HR, Bock PE, 1976. Phosphofructokinase. Iii. Correlation of the regulatory kinetic and molecular properties of the rabbit muscle enzyme. J. Biol. Chem. 251 (18), 5644–5647. [PubMed: 9395]

Germain A, Barupal DK, Levine SM, Hanson MR, 2020. Comprehensive circulatory metabolomics in me/cfs reveals disrupted metabolism of acyl lipids and steroids. Metabolites 10 (1), 34.

Gillenwater LA, Pratte KA, Hobbs BD, Cho MH, Zhuang Y, Halper-Stromberg E, Cruickshank-Quinn C, Reisdorph N, Petrache I, Labaki WW, O'Neal WK, Ortega VE, Jones DP, Uppal K, Jacobson S, Michelotti G, Wendt CH, Kechris KJ, Bowler RP, 2020. Plasma metabolomic signatures of chronic obstructive pulmonary disease and the impact of genetic variants on phenotype-driven modules. Netw. Syst. Med. 3 (1), 159–181. [PubMed: 33987620]

Gillenwater LA, Kechris KJ, Pratte KA, Reisdorph N, Petrache I, Labaki WW, O'Neal W, Krishnan JA, Ortega VE, DeMeo DL, Bowler RP, 2021. Metabolomic profiling reveals sex specific associations with chronic obstructive pulmonary disease and emphysema. Metabolites 11 (3), 161. [PubMed: 33799786]

Guha N, Guyton KZ, Loomis D, Barupal DK, 2016. Prioritizing chemicals for risk assessment using chemoinformatics: Examples from the iarc monographs on pesticides. Environ. Health Perspect. 124, 1823–1829. 10.1289/EHP186. [PubMed: 27164621]

Guo H, Yu X, Liu Z, Li J, Ye J, Zha Z, 2020. Deltamethrin transformation by bacillus thuringiensis and the associated metabolic pathways. Environ. Int. 145, 106167. 10.1016/j.envint.2020.106167. [PubMed: 33035892]

Guo J, Shen S, Xing S, Yu H, Huan T, 2021. Isfrag: De novo recognition of in-source fragments for liquid chromatography-mass spectrometry data. Anal. Chem. 93, 10243–10250. 10.1021/acs.analchem.1c01644. [PubMed: 34270210]

Guo K, Savelieff MG, Rumora AE, Alakwaa FM, Callaghan BC, Hur J, Feldman EL, 2022. Plasma metabolomics and lipidomics differentiate obese individuals by peripheral neuropathy status. J. Clin. Endocrinol. Metab. 107 (4), 1091–1109. [PubMed: 34878536]

Guthrie L, Wolfson S, Kelly L, 2019. The human gut chemical landscape predicts microbe-mediated biotransformation of foods and drugs. Elife 8. 10.7554/eLife.42866.

Hecht SS, Stepanov I, Carmella SG, 2016. Exposure and metabolic activation biomarkers of carcinogenic tobacco-specific nitrosamines. Acc. Chem. Res. 49, 106–114. 10.1021/acs.accounts.5b00472. [PubMed: 26678241]

Hendrix JA, Finger B, Weiner MW, Frisoni GB, Iwatsubo T, Rowe CC, Kim SY, Guinjoan SM, Sevlever G, Carrillo MC, 2015. The worldwide alzheimer's disease neuroimaging initiative: An update. Alzheimers Dement 11 (7), 850–859. [PubMed: 26194318]

Hoffman DA, Wallace SM, Verbeeck RK, 1990. Circadian rhythm of serum sulfate levels in man and acetaminophen pharmacokinetics. Eur. J. Clin. Pharmacol. 39, 143–148. 10.1007/BF00280048. [PubMed: 2253663]

Huang H, Shi L-Y, Wei L-L, Han Y-S, Yi W-J, Pan Z-W, Jiang T-T, Chen J, Tu H-H, Li Z-B, Hu Y-T, Li J-C, 2019. Plasma metabolites xanthine, 4-pyridoxate, and d-glutamic acid as novel potential biomarkers for pulmonary tuberculosis. Clin. Chim. Acta 498, 135–142. [PubMed: 31442449]

Hur B, Gupta VK, Huang H, Wright KA, Warrington KJ, Taneja V, Davis JM, Sung J, 2021. Plasma metabolomic profiling in patients with rheumatoid arthritis identifies biochemical features predictive of quantitative disease activity. Arthritis Res. Ther. 23 (1), 164. 10.1186/s13075-021-02537-4. [PubMed: 34103083]

Jobard E, Dossus L, Baglietto L, Fornili M, Lécuyer L, Mancini FR, Gunter MJ, Trédan O, Boutron-Ruault M-C, Elena-Herrmann Bénédicte, Severi G, Rothwell JA, 2021. Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: A case-control study nested within the french e3n cohort. Br. J. Cancer 124 (10), 1734–1743. [PubMed: 33723391]

Kassotis CD, Vandenberg LN, Demeneix BA, Porta M, Slama R, Trasande L, 2020. Endocrine-disrupting chemicals: Economic, regulatory, and policy implications. Lancet Diab. Endocrinol. 8, 719–730. 10.1016/S22138587(20)30128-5.

Kitagawa H, Ohbuchi K, Munekage M, Fujisawa K, Kawanishi Y, Namikawa T, Kushida H, Matsumoto T, Shimobori C, Nishi A, Sadakane C, Watanabe J, Yamamoto M, Hanazaki K, 2019.

Phenotyping analysis of the japanese kampo medicine maoto in healthy human subjects using wide-targeted plasma metabolomics. J. Pharm. Biomed. Anal. 164, 119–127. [PubMed: 30368117]

Krishnan S, Nordqvist H, Ambikan AT, Gupta S, Sperk M, Svensson-Akusjärvi S, Mikaeloff F, Benfeitas R, Saccon E, Ponnan SM, Rodriguez JE, Nikouyan N, Odeh A, Ahlén G, Asghar M, Sällberg M, Vesterbacka J, Nowak P, Végvári Ákos, Sönnerborg A, Treutiger CJ, Neogi U, 2021. Metabolic perturbation associated with covid-19 disease severity and sars-cov-2 replication. Mol. Cell Proteomics 20, 100159. 10.1016/j.mcpro.2021.100159. [PubMed: 34619366]

Langfelder P, Horvath S, 2008. Wgcna: An r package for weighted correlation network analysis. BMC Bioinform. 9 (1), 559. 10.1186/1471-2105-9-559.

Lau CE, Siskos AP, Maitre L, Robinson O, Athersuch TJ, Want EJ, et al. , 2018. Determinants of the urinary and serum metabolome in children from six european populations. BMC Med. 16, 202. 10.1186/s12916-018-1190-8. [PubMed: 30404627]

Lee J, Shah M, Ballouz S, Crow M, Gillis J, 2020. Cococonet: Conserved and comparative co-expression across a diverse set of species. Nucleic Acids Res. 48 10.1093/nar/gkaa348.

Levi I, Gurevich M, Perlman G, Magalashvili D, Menascu S, Bar N, Godneva A, Zahavi L, Chermon D, Kosower N, Wolf BC, Malka G, Lotan-Pompan M, Weinberger A, Yirmiya E, Rothschild D, Leviatan S, Tsur A, Didkin M, Dreyer S, Eizikovitz H, Titngi Y, Mayost S, Sonis P, Dolev M, Stern Y, Achiron A, Segal E, 2021. Potential role of indolelactate and butyrate in multiple sclerosis revealed by integrated microbiome-metabolome analysis. Cell Rep. Med. 2 (4), 100246. 10.1016/j.xcrm.2021.100246. [PubMed: 33948576]

Li S, Sullivan NL, Rouphael N, Yu T, Banton S, Maddur MS, McCausland M, Chiu C, Canniff J, Dubey S, Liu K, Tran V, Hagan T, Duraisingham S, Wieland A, Mehta AK, Whitaker JA, Subramaniam S, Jones DP, Sette A, Vora K, Weinberg A, Mulligan MJ, Nakaya HI, Levin M, Ahmed R, Pulendran B, 2017. Metabolic phenotypes of response to vaccination in humans. Cell 169 (5).

Liang L, Rasmussen M-L, Piening B, Shen X, Chen S, Röst H, Snyder JK, Tibshirani R, Skotte L, Lee NCY, Contrepois Kévin, Feenstra B, Zackriah H, Snyder M, Melbye M, 2020. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. Cell 181 (7).

Liu H, Garrett TJ, Su Z, Khoo C, Zhao S, Gu L, 2020. Modifications of the urinary metabolome in young women after cranberry juice consumption were revealed using the uhplc-q-orbitrap-hrms-based metabolomics approach. Food Funct. 11, 2466–2476. 10.1039/c9fo02266j. [PubMed: 32133462]

Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. , 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662. 10.1038/s41586-019-1237-9. [PubMed: 31142855]

Loftfield E, Stepien M, Viallon V, Trijsburg L, Rothwell JA, Robinot N, Biessy C, Bergdahl IA, Bodén S, Schulze MB, Bergman M, Weiderpass E, Schmidt JA, Zamora-Ros R, Nøst TH, Sandanger TM, Sonestedt E, Ohlsson B, Katzke V, Kaaks R, Ricceri F, Tjønneland A, Dahm CC, Sánchez M-J, Trichopoulou A, Tumino R, Chirlaque M-D, Masala G, Ardanaz E, Vermeulen R, Brennan P, Albanes D, Weinstein SJ, Scalbert A, Freedman ND, Gunter MJ, Jenab M, Sinha R, Keski-Rahkonen P, Ferrari P, 2021. Novel biomarkers of habitual alcohol intake and associations with risk of pancreatic and liver cancers and liver disease mortality. J. Natl Cancer Inst. 113 (11), 1542–1550. [PubMed: 34010397]

Matta MK, Florian J, Zusterzeel R, Pilli NR, Patel V, Volpe DA, Yang Y, Oh L, Bashaw E, Zineh I, Sanabria C, Kemp S, Godfrey A, Adah S, Coelho S, Wang J, Furlong L-A, Ganley C, Michele T, Strauss DG, 2020. Effect of sunscreen application on plasma concentration of sunscreen active ingredients: A randomized clinical trial. JAMA 323 (3), 256. [PubMed: 31961417]

McKillop K, Harnly J, Pehrsson P, Fukagawa N, Finley J, 2021. Fooddata central, usda's updated approach to food composition data systems. Curr. Develop. Nutrit. 5, 596–596.

Meister I, Zhang P, Sinha A, Sköld CM, Wheelock Åsa.M., Izumi T, Chaleckis R, Wheelock CE, 2021. High-precision automated workflow for urinary untargeted metabolomic epidemiology. Anal. Chem. 93 (12), 5248–5258. [PubMed: 33739820]

Michonneau D, Latis E, Curis E, Dubouchet L, Ramamoorthy S, Ingram B, de Latour Régis.P., Robin M, de Fontbrune FS, Chevret S, Rogge L, Socié G, 2019. Metabolomics analysis of human

acute graft-versus-host disease reveals changes in host and microbiota-derived metabolites. Nat. Commun. 10 (1), 5695. 10.1038/s41467-019-13498-3. [PubMed: 31836702]

Needham BD, Adame MD, Serena G, Rose DR, Preston GM, Conrad MC, Campbell AS, Donabedian DH, Fasano A, Ashwood P, Mazmanian SK, 2021. Plasma and fecal metabolite profiles in autism spectrum disorder. Biol. Psychiatry 89 (5), 451–462. [PubMed: 33342544]

Nemet I, Saha PP, Gupta N, Zhu W, Romano KA, Skye SM, Cajka T, Mohan ML, Li L, Wu Y, Funabashi M, Ramer-Tait AE, Naga Prasad SV, Fiehn O, Rey FE, Tang WHW, Fischbach MA, DiDonato JA, Hazen SL, 2020. A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors. Cell 180 (5).

Nymand Ennis Z, Arnspang Pedersen S, Rix Hansen M, Pottegård A, Patrick Ahern T, Hallas J, Damkier P, 2019. Use of phthalate-containing prescription drugs and the risk of gastric cancer: A danish nationwide case-control study. Acta Oncol. 58 (6), 852–858. [PubMed: 30882263]

Obayashi T, Kagaya Y, Aoki Y, Tadaka S, Kinoshita K, 2019. Coxpresdb v7: A gene coexpression database for 11 animal species supported by 23 coexpression platforms for technical evaluation and evolutionary inference. Nucleic Acids Res. 47, D55–D62. 10.1093/nar/gky1155. [PubMed: 30462320]

Park EY, Kim J, Park E, Oh J-K, Kim B, Lim MK, 2021. Serum concentrations of persistent organic pollutants and colorectal cancer risk: A case-cohort study within korean national cancer center community (knccc) cohort. Chemosphere 271, 129596. 10.1016/j.chemosphere.2021.129596. [PubMed: 33460900]

Paul-Friedman K, Martin M, Crofton KM, Hsu C-W, Sakamuru S, Zhao J, Xia M, Huang R, Stavreva DA, Soni V, Varticovski L, Raziuddin R, Hager GL, Houck KA, 2019. Limited chemical structural diversity found to modulate thyroid hormone receptor in the tox21 chemical library. Environ. Health Perspect 127 (9), 97009. 10.1289/EHP5314. [PubMed: 31566444]

Peters A, Nawrot TS, Baccarelli AA, 2021. Hallmarks of environmental insults. Cell 184, 1455–1468. 10.1016/j.cell.2021.01.043. [PubMed: 33657411]

Petrick JL, Florio AA, Koshiol J, Pfeiffer RM, Yang B, Yu K, et al. , 2020. Prediagnostic concentrations of circulating bile acids and hepatocellular carcinoma risk: Reveal-hbv and hcv studies. Int. J. Cancer 147, 2743–2753. 10.1002/ijc.33051. [PubMed: 32406072]

Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, et al. , 2019. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. Nat. Med. 25, 1442–1452. 10.1038/s41591-019-0559-3. [PubMed: 31477907]

Qian H, Chen M, Kransler KM, Zaleski RT, 2015. Assessment of chemical coexposure patterns based upon phthalate biomonitoring data within the 2007/2008 national health and nutrition examination survey. J. Expo. Sci. Environ. Epidemiol. 25, 249–255. 10.1038/jes.2014.24. [PubMed: 24756100]

Rago D, Rasmussen MA, Lee-Sarwar KA, Weiss ST, Lasky-Su J, Stokholm J, Bønnelykke K, Chawes BL, Bisgaard H, 2019. Fish-oil supplementation in pregnancy, child metabolomics and asthma risk. EBioMedicine 46, 399–410. [PubMed: 31399385]

Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A, 2014. The blood exposome and its role in discovering causes of disease. Environ. Health Perspect. 122, 769–774. 10.1289/ehp.1308015. [PubMed: 24659601]

Richard AM, Huang R, Waidyanatha S, Shinn P, Collins BJ, Thillainadarajah I, Grulke CM, Williams AJ, Lougee RR, Judson RS, Houck KA, Shobair M, Yang C, Rathman JF, Yasgar A, Fitzpatrick SC, Simeonov A, Thomas RS, Crofton KM, Paules RS, Bucher JR, Austin CP, Kavlock RJ, Tice RR, 2021. The tox21 10k compound library: Collaborative chemistry advancing toxicology. Chem. Res. Toxicol. 34 (2), 189–216. [PubMed: 33140634]

Rumora AE, Guo K, Alakwaa FM, Andersen ST, Reynolds EL, Jørgensen ME, Witte DR, Tankisi H, Charles M, Savelieff MG, Callaghan BC, Jensen TS, Feldman EL, 2021. Plasma lipid metabolites associate with diabetic polyneuropathy in a cohort with type 2 diabetes. Ann. Clin. Transl. Neurol. 8 (6), 1292–1307. [PubMed: 33955722]

Saravanabhavan G, Guay M, Langlois E, Giroux S, Murray J, Haines D, 2013. Biomonitoring of phthalate metabolites in the canadian population through the canadian health measures survey (2007–2009). Int. J. Hyg. Environ. Health 216, 652–661. 10.1016/j.ijheh.2012.12.009. [PubMed: 23419587]

Schillemans T, Shi L, Donat-Vargas C, Hanhineva K, Tornevi A, Johansson I, Koponen J, Kiviranta H, Rolandsson O, Bergdahl IA, Landberg R, Åkesson A, Brunius C, 2021. Plasma metabolites associated with exposure to perfluoroalkyl substances and risk of type 2 diabetes - a nested case-control study. Environ. Int. 146, 106180. 10.1016/j.envint.2020.106180. [PubMed: 33113464]

Semente L, Baquer G, Garcia-Altares M, Correig-Blanchar X, Rafols P, 2021. Rmsiannotation: A peak annotation tool for mass spectrometry imaging based on the analysis of isotopic intensity ratios. Anal. Chim. Acta 1171, 338669. 10.1016/j.aca.2021.338669. [PubMed: 34112434]

Sen P, Qadri S, Luukkonen PK, Ragnarsdottir O, McGlinchey A, Jäntti S, Juuti A, Arola J, Schlezinger JJ, Webster TF, Orešič M, Yki-Järvinen H, Hyötyläinen T, 2022. Exposure to environmental contaminants is associated with altered hepatic lipid metabolism in non-alcoholic fatty liver disease. J. Hepatol. 76 (2), 283–293. [PubMed: 34627976]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T, 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13 (11), 2498–2504. [PubMed: 14597658]

Shearer JJ, Callahan CL, Calafat AM, Huang W-Y, Jones RR, Sabbisetti VS, Freedman ND, Sampson JN, Silverman DT, Purdue MP, Hofmann JN, 2021. Serum concentrations of per- and polyfluoroalkyl substances and risk of renal cell carcinoma. J. Natl. Cancer Inst. 113 (5), 580–587. [PubMed: 32944748]

Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, Quan S, Zhang F, Sun R, Qian L, Ge W, Liu W, Liang S, Chen H, Zhang Y, Li J, Xu J, He Z, Chen B, Wang J, Yan H, Zheng Y, Wang D, Zhu J, Kong Z, Kang Z, Liang X, Ding X, Ruan G, Xiang N, Cai X, Gao H, Li L, Li S, Xiao Q, Lu T, Zhu Y, Liu H, Chen H, Guo T, 2020. Proteomic and metabolomic characterization of covid-19 patient sera. Cell 182 (1).

Sinisalu L, Sen P, Salihović S, Virtanen SM, Hyöty H, Ilonen J, Toppari J, Veijola R, Orešič M, Knip M, Hyötyläinen T, 2020. Early-life exposure to perfluorinated alkyl substances modulates lipid metabolism in progression to celiac disease. Environ. Res. 188, 109864. 10.1016/j.envres.2020.109864. [PubMed: 32846648]

Stanfield Z, Addington CK, Dionisio KL, Lyons D, Tornero-Velez R, Phillips KA, Buckley TJ, Isaacs KK, 2021. Mining of consumer product ingredient and purchasing data to identify potential chemical coexposures. Environ. Health Perspect. 129 (6), 67006. 10.1289/EHP8610. [PubMed: 34160298]

Stevens VL, Wang Y, Carter BD, Gaudet MM, Gapstur SM, 2018. Serum metabolomic profiles associated with postmenopausal hormone use. Metabolomics 14 (7), 97. 10.1007/s11306-018-1393-1. [PubMed: 30830410]

Tahir UA, Katz DH, Zhao T, Ngo D, Cruz DE, Robbins JM, Chen Z-Z, Peterson B, Benson MD, Shi X, Dailey L, Andersson C, Vasan RS, Gao Y, Shen C, Correa A, Hall ME, Wang TJ, Clish CB, Wilson JG, Gerszten RE, 2021. Metabolomic profiles and heart failure risk in black adults: Insights from the jackson heart study. Circ. Heart Fail 14 (1). 10.1161/CIRCHEARTFAILURE.120.007275.

Tanes C, Bittinger K, Gao Y, Friedman ES, Nessel L, Paladhi UR, Chau L, Panfen E, Fischbach MA, Braun J, Xavier RJ, Clish CB, Li H, Bushman FD, Lewis JD, Wu GD, 2021. Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. Cell Host Microbe 29 (3).

Tang Y, El-Chemaly S, Taveira-Dasilva A, Goldberg HJ, Bagwe S, Rosas IO, Moss J, Priolo C, Henske EP, 2019. Alterations in polyamine metabolism in patients with lymphangioleiomyomatosis and tuberous sclerosis complex 2-deficient cells. Chest 156 (6), 1137–1148. [PubMed: 31299246]

Toledo JB, Arnold M, Kastenmuller G, Chang R, Baillie RA, Han X, et al. , 2017. Metabolic network failures in alzheimer's disease: A biochemical road map. Alzheimers Dement 13, 965–984. 10.1016/j.jalz.2017.01.020. [PubMed: 28341160]

Vangipurapu J, Fernandes Silva L, Kuulasmaa T, Smith U, Laakso M, 2020. Microbiota-related metabolites and the risk of type 2 diabetes. Diab. Care 43, 1319–1325. 10.2337/dc19-2533.

Vega RB, Whytock KL, Gassenhuber J, Goebel B, Tillner J, Agueusop I, Truax AD, Yu G, Carnero E, Kapoor N, Gardell S, Sparks LM, Smith SR, 2021. A metabolomic signature of glucagon action in healthy individuals with overweight/obesity. J. Endocr. Soc. 5 (9) 10.1210/jendso/bvab118.

Vermeulen R, Schymanski EL, Barabasi AL, Miller GW, 2020. The exposome and health: Where chemistry meets biology. Science 367, 392–396. 10.1126/science.aay3164. [PubMed: 31974245]

Yin W, Mendoza L, Monzon-Sandoval J, Urrutia AO, Gutierrez H, Provero P, 2021. Emergence of co-expression in gene regulatory networks. PLoS ONE 16 (4), e0247671. 10.1371/journal.pone.0247671. [PubMed: 33793561]

Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R, Goodman AL, 2019. Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature 570, 462–467. 10.1038/s41586-019-1291-3. [PubMed: 31158845]

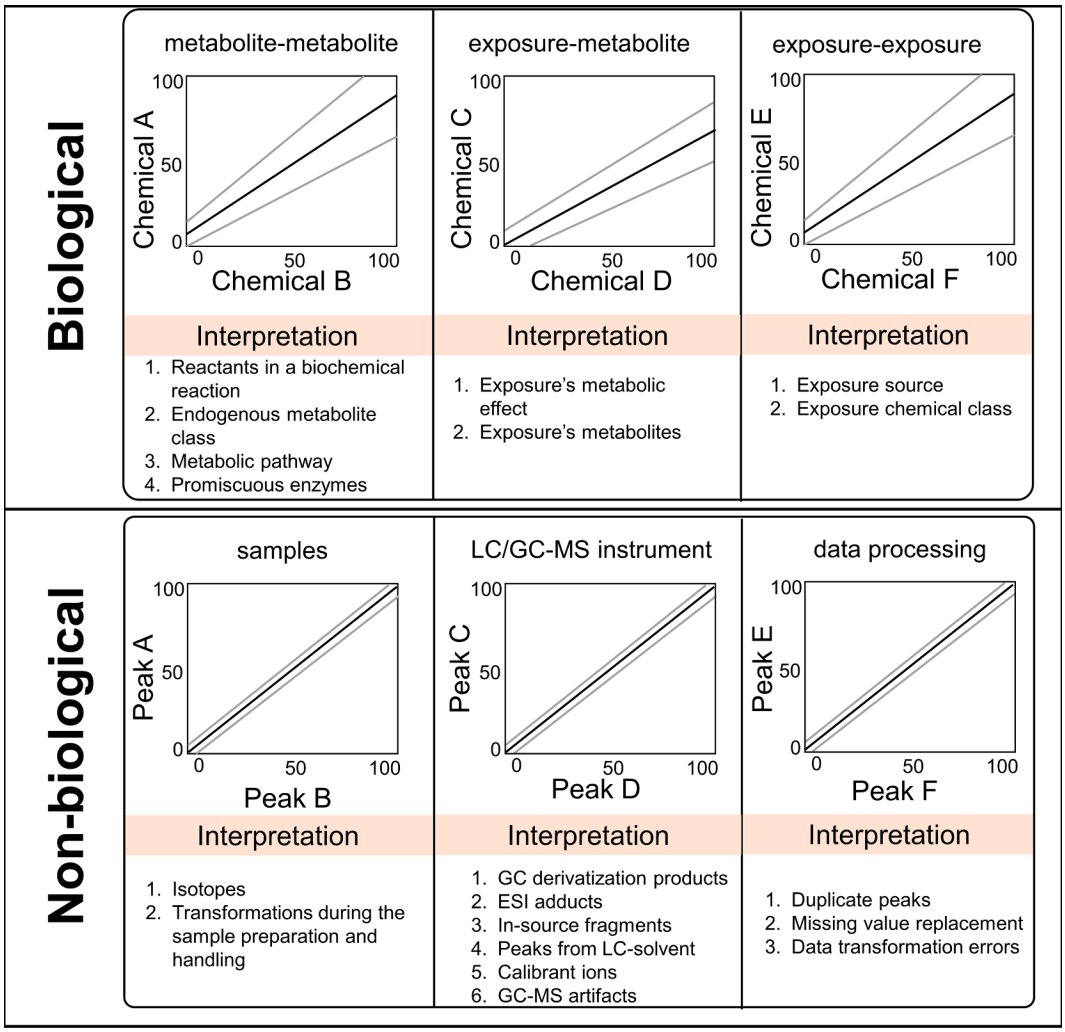

Fig. 1.

Probable interpretations of correlation in targeted and untargeted GC/LC-HRMS datasets.

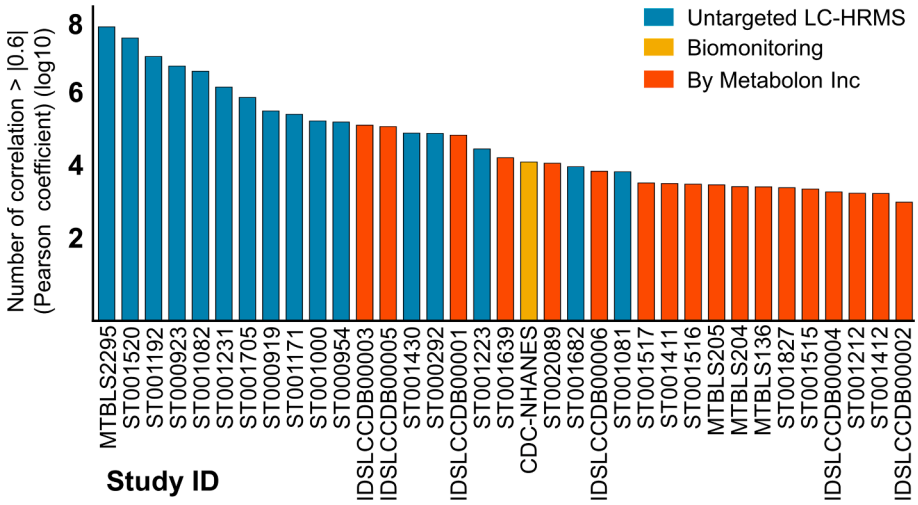

Fig. 2.

Prevalence of strong inter-chemical correlations across 35 studies in the CCDB. These are unique correlations. See the Table 1 for the description of each study and number of compounds. Table S3 shows the chemical detection rate across the indexed studies.

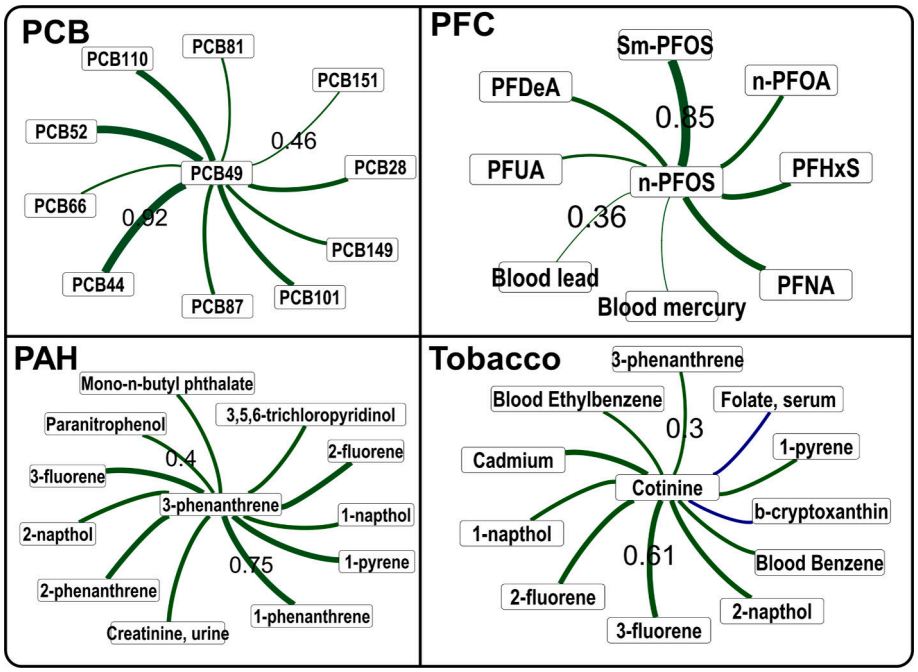

Fig. 3.

Correlations among chemicals within a class or having same source origin in the NHANES dataset. The correlation cutoff was 0.3 for PCB, PFC and Tobacco compounds, and 0.4 for PAHs. Online network can be accessed at the site - https://chemcor.idsl.site/originaldata/biomonitoring/#?studyid=NHANES. Edge thickness shows the correlation strength, by only the minimum and maximum correlation values are labelled on the edges for clarity. Thickness of edges are not comparable in two network figures. Abbreviations: Perfluorodecanoic acid (PFDeA), Perfluorohexane sulfonic acid (PFHxS), Perfluorononanoic acid (PFNA), Perfluoroundecanoic acid (PFUA), n-perfluorooctanoic acid (n-PFOA), n-perfluorooctane sulfonic acid (n-PFOS), Perfluoromethylheptane sulfonic acid isomers (SmPFOS), Polychlorinated Biphenyls (PCB); polyaromatic hydrocarbons (PAH), Perfluorinated compounds (PFC).

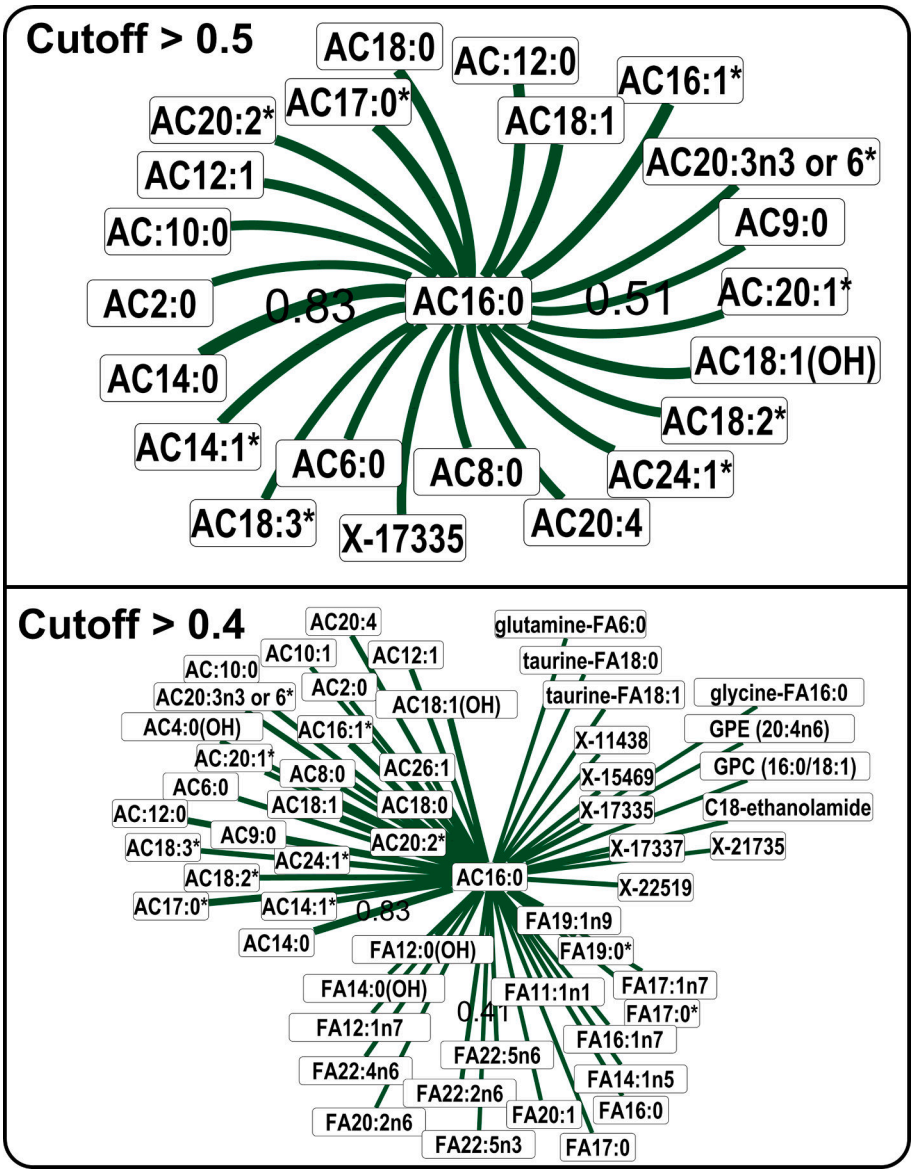

Fig. 4.

Compounds correlation with acylcarnitine 16:0 in the study ST002089. Edge thickness shows the correlation strength, by only the minimum and maximum correlation values are labelled on the edges for clarity. Thickness of edges are not comparable in two network figures. Abbreviations: acyl-carnitines (AC). Fatty acid (FA), glycerophosphoethanolamine (GPE), glycerophosphocholine (GPC).

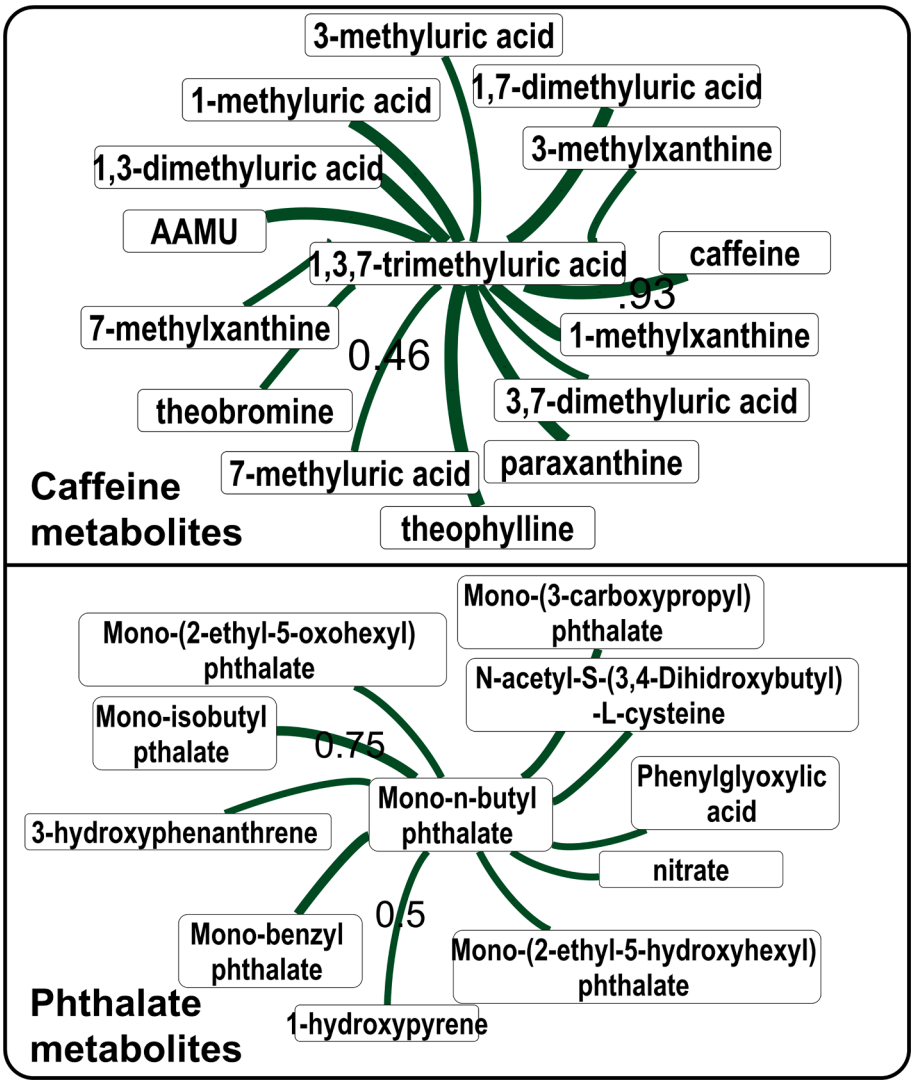

Fig. 5.

Caffeine and phthalate metabolites in the NHANES survey data. Variable id URXMBP_PHTHTE_D (year 2005–2006) was used for mono-n-butyl phthalate (MnBP). Variable id URXMX7_CAFE_H (year 2013–2014) was used for caffeine. Label on the edges show the Pearson coefficient. Edge thickness shows the correlation strength, by only the minimum and maximum correlation values are labelled on the edges for clarity. Thickness of edges are not comparable in two network figures. Abbreviations: acetylamino-6-formylamino-3-methyluracil(AAMU).

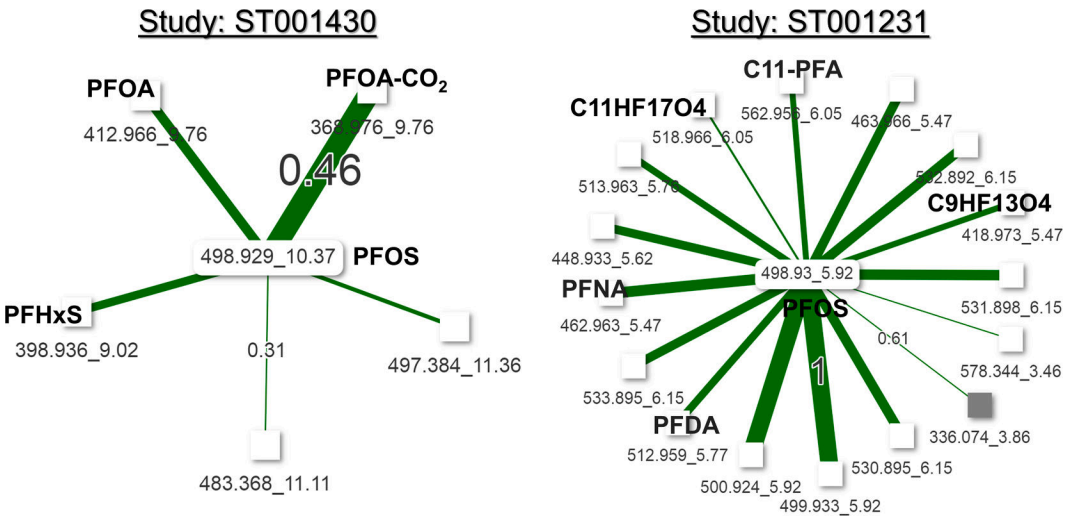

Fig. 6.

Inter-chemical correlation among PFCs in the untargeted metabolomics datasets. Correlation threshold for ST001430 was 0.3 and for 0.6 for ST001231. White color node mean it was detected in by the reverse phase ESI (−) mode and a grey node means it was detected by a reverse phrase ESI (+) mode. Edge thickness shows the correlation strength, by only the minimum and maximum correlation values are labelled on the edges for clarity. Thickness of edges are not comparable in two network figures.

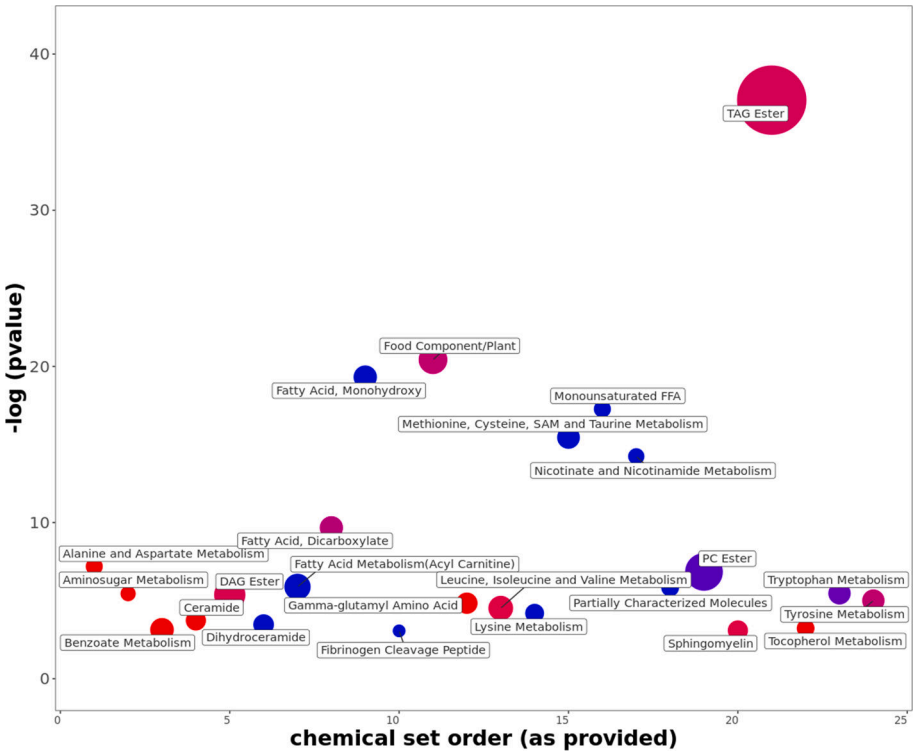

Fig. 7.

Chemical similarity enrichment analysis of PFOA and its correlation with other metabolites in the IDSLCCDB00001 study.

Table 1

Covered studies in the CCDB on March 2022.

Database/Source	Accession ID	Title	Number of Samples	Number of Peaks	Specimen
Metabolomics WorkBench	ST000923	Longitudinal Metabolomics of the Human Microbiome in Inflammatory Bowel Disease (Lloyd-Price et al., 2019)	546	81,867	Stool
Metabolomics WorkBench	ST001000	Gut microbiome structure and metabolic activity in inflammatory bowel disease (Franzosa et al., 2019)	220	8847	Stool
Metabolomics WorkBench	ST001192	A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research (Poyet et al., 2019)	180	54,402	Stool
Metabolomics WorkBench	ST001520	Stool unknowns profiled using hybrid nontargeted methods (part-II)(Tanes et al., 2021)	166	54,014	Stool
CDC-NHANES	NHANES	National Health and Nutrition Examination Survey – USA. Continuous NHANES data from 1999 to 2020 period. https://www.cdc.gov/nchs/nhanes/index.htm	107,258 (SEQN IDs)	607 (variables)	Blood/Urine
Metabolomics WorkBench	ST001223	Host Metabolic Response in Early Lyme Disease (Fitzgerald et al., 2020)	518	2193	Blood
Metabolomics WorkBench	ST001081	Combined NMR and MS Analysis of PC patient serum (part-I)(Clendinen et al., 2019)	168	459	Blood
Metabolomics WorkBench	ST001082	Combined NMR and MS Analysis of PC patien serum (part-II)(Clendinen et al., 2019)	265	24,928	Blood
Metabolomics WorkBench	ST001682	Untargeted urine LC-HRMS metabolomics profiling for bladder cancer binary outcome classification	311	982	Urine
EBI MetaboLights	MTBLS136	Serum metabolomic profiles associated with postmenopausal hormone use (Stevens et al., 2018)	1336	1385	Blood
EBI MetaboLights	MTBLS204	Metabolomics analysis of human acute graft-versus-host disease reveals changes in host and microbiota-derived metabolites (Michonneau et al., 2019)	86	801	Blood
EBI MetaboLights	MTBLS205	Metabolomics analysis of human acute graft-versus-host disease reveals changes in host and microbiota-derived metabolites (Michonneau et al., 2019)	112	929	Blood
Metabolomics WorkBench	ST001516	Identification of distinct metabolic perturbations and associated immunomodulatory events during intra-erythrocytic development stage of pediatric Plasmodium falciparum malaria (Abdrabou et al., 2021)	199	668	Blood
Metabolomics WorkBench	ST001517	Identification of distinct metabolic perturbations and associated immunomodulatory events during intra-erythrocytic development stage of pediatric Plasmodium falciparum malaria (Abdrabou et al., 2021)	106	652	Blood
Metabolomics WorkBench	ST001639	Plasma Metabolomic signatures of COPD in a SPIROMICS cohort (Gillenwater et al., 2020)	649	1174	Blood
Metabolomics WorkBench	ST001212	Fish-oil supplementation in pregnancy, child metabolomics and asthma risk (Rago et al., 2019)	577	656	Blood
Metabolomics WorkBench	ST001827	The pregnancy metabolome from a multi-ethnic pregnancy cohort (Colicino et al., 2021)	410	1110	Blood
PMC Open Access	IDSLCCDB00001	Plasma and Fecal Metabolite Profiles in Autism Spectrum Disorder (Needham et al., 2021)	222	1611	Blood

Database/Source	Accession ID	Title	Number of Samples	Number of Peaks	Specimen
PMC Open Access	IDSLCCDB00002	Potential role of indolelactate and butyrate in multiple sclerosis revealed by integrated microbiome-metabolome analysis (Levi et al., 2021)	180	517	Blood
PMC Open Access	IDSLCCDB00003	Comprehensive Circulatory Metabolomics in ME/CFS Reveals Disrupted Metabolism of Acyl Lipids and Steroids (Germain et al., 2020)	52	1790	Blood
PMC Open Access	IDSLCCDB00004	Plasma Metabolomic Profiling in Patients with Rheumatoid Arthritis Identifies Biochemical Features Indicative of Quantitative Disease Activity (Hur et al., 2021)	128	686	Blood
PMC Open Access	IDSLCCDB00005	Alterations in Polyamine Metabolism in Patients With Lymphangioleiomyomatosis and Tuberous Sclerosis Complex 2-Deficient Cells (Tang et al., 2019)	78	1989	Blood
PMC Open Access	IDSLCCDB00006	Metabolic perturbation associated with COVID-19 disease severity and SARS-CoV-2 replication (Krishnan et al., 2021)	72	1086	Blood
Metabolomics WorkBench	ST002089	Plasma metabolomic signatures of COPD: A metabolomic severity score for airflow obstruction and emphysema (Gillenwater et al., 2021)	1120	1394	Blood
Metabolomics WorkBench	ST001411	Plasma metabolites of lipid metabolism associate with diabetic polyneuropathy in a cohort with screen-tested type 2 diabetes: ADDITION-Denmark (Rumora et al., 2021)	106	991	Blood
Metabolomics WorkBench	ST001412	Metabolomics study in Plasma of Obese Patients with Neuropathy Identifies Potential Metabolomics Signatures (K Guo et al., 2021)	131	842	Blood
Metabolomics WorkBench	ST001515	A Metabolomic Signature of Glucagon Action in Healthy Individuals with Overweight/Obesity Humans (Vega et al., 2021)	187	649	Blood
Metabolomics WorkBench	ST001171	Metabolomics of World Trade Center Exposed New York City Firefighters	248	2504	Blood
Metabolomics WorkBench	ST001430	Metabolic dynamics and prediction of gestational ange and time to delivery in pregnant women (Liang et al., 2020)	781	9651	Blood
Metabolomics WorkBench	ST001705	Machine learning-enabled renal cell carcinoma status prediction using multi-platform urine-based metabolomics (part-I)(Bifarin et al., 2021)	256	7097	Urine
Metabolomics WorkBench	ST000292	LC-MS Based Approaches to Investigate Metabolomic Differences in the Plasma of Young Women after Drinking Cranberry Juice or Apple Juice (Liu et al., 2020)	51	3395	Blood
Metabolomics WorkBench	ST000919	Investigating Eicosanoids Implications on the Blood Pressure Response to Thiazide Diuretics	140	10,322	Blood
Metabolomics WorkBench	ST000954	Explore Metabolites and Pathways Associated Increased Airway Hyperresponsiveness in Asthma	55	7930	Blood
Metabolomics WorkBench	ST001231	Plasma untargeted metabolomics study of pulmonary tuberculosis (Huang et al., 2019)	159	17,146	Blood
EBI MetaboLights	MTBLS2295	High-Precision Automated Workflow for Urinary Untargeted Metabolomic Epidemiology (Meister et al., 2021)	87	655	Urine