Integrated exposomic analysis of lipid phenotypes: leveraging GE.db in environment by environment association studies

ANDRE LUIS GARAO RICO and NICOLE PALMIERO

*Department of Genomics, University of Pennsylvania, 3700 Hamilton Walk  
Philadelphia, PA 19104, USA  
Email: andreluis.rico@pennmedicine.upenn.edu*

MARYLYN D. RITCHIE

*Department of Genomics, University of Pennsylvania, 3700 Hamilton Walk  
Philadelphia, PA 19104, USA  
Email: marylyn@pennmedicine.upenn.edu*

MOLLY A. HALL

*Department of Genomics, University of Pennsylvania, 3700 Hamilton Walk  
Philadelphia, PA 19104, USA  
Email: molly.hall@pennmedicne.upenn.edu*

Gene by environment-wide association studies (GxE) provide insights into genetic-environment interactions but often overlook multiple environmental factors' synergistic effects. This study encompasses the use of environment by environment association studies (ExE) to explore interactions among environmental factors affecting lipid phenotypes (e.g., HDL-C, LDL-C, total cholesterol, triglycerides), crucial for disease risk assessment. Leveraging the GE.db module of the IGEM system, a curated knowledge base integrating genomic and exposomic interactions, we filtered NHANES data (1999-2018) to identify significant ExE interactions. From 101,316 participants and 11,274 variants, filtering identified 382,613 interactions among 217 exposure terms. Quality control and interaction analyses revealed 263 significant interactions (FDR p < 0.1) in discovery and replication datasets, with twenty-one interactions significant for HDL-C (Bonferroni p < 0.05). Notable interactions included docosapentaenoic acid (22:5n-3) (DPA) - arachidic acid (20:0), stearic acid (18:0) - arachidic acid (20:0), and blood 2,5-dimethyfuran - blood benzene impacting HDL-C levels. These findings underscore GE.db's role in enhancing -omics research efficiency and highlight the complex impact of environmental exposures on lipid metabolism, informing future health strategies.

# Introduction

Understanding the intricate interplay between genetics and the environment is pivotal in unraveling the complexities of human traits and diseases. While gene by environment-wide association studies (GxE) have provided valuable insights into how genetic variants interact with environmental factors, they often overlook the synergistic effects of multiple environmental variables[1,2](https://paperpile.com/c/HmCkMR/u5c0+Xicw). This limitation necessitates the need for utilizing environment by environment association studies (ExE), which explore how different environmental factors interact with each other to influence phenotypic outcomes. The outcomes of interest used in this study are lipids such as high-density lipoprotein-cholesterol (HDL-C), low-density lipoprotein-cholesterol (LDL-C), total cholesterol, and triglycerides, as they have relation to important risk factors for a multitude of diseases3–5. These conditions are influenced by a variety of factors, including genetic inheritance, environmental and occupational exposures, medication use, ethnicity, and sex[6,7](https://paperpile.com/c/HmCkMR/mjTc+iAlA). A factor of exposure is any physical, chemical, or biological agent that someone is exposed to and has potential to cause adverse health effects. In the nutritional context, exposure refers to the intake of nutrients that can affect health either positively or negatively.

Moreover, the shift towards knowledge-based filtering in these studies proves more effective than traditional main effect filtering, as it incorporates prior biological knowledge to prioritize genetic variants that are more likely to interact with specific environmental cues, thereby enhancing the precision and relevance of identified associations[8](https://paperpile.com/c/HmCkMR/Ku9H). Thus, ExE coupled with knowledge-based filtering represents a promising approach to comprehensively elucidate the complexities of ExE interactions in health and disease. This paper introduces the Gene x Expsome database (GE.db) module of the Integrative Genome-Exposome Method (IGEM) system[9](https://paperpile.com/c/HmCkMR/YGRT), a comprehensive knowledge base of genomic and exposomic interactions derived from various public databases [see Methods]. The development of GE.db aims to leverage prior knowledge to filter high-volume research datasets, retaining only variables with known interactions. This approach significantly reduces the number of variables for analysis, conserves computational resources and processing time, and minimizes type I errors following multiple testing corrections.

To demonstrate the utility of GE.db, we conducted an exposure-exposure (ExE) interaction analysis with lipid outcomes using the National Health and Nutritional Examination Survey (NHANES)10 data from 1999-2018. By aggregating data from all NHANES cycles, information from 101,316 participants and 11,274 fields, including survey questions and laboratory results, was obtained. Among these fields, 3,619 were associated with GE.db terms, and of these, 1,107 were identified as exposure-related factors, revealing 344,143 interactions between these fields. After quality control and ExE interaction analysis, we identified twenty-one significant interactions with a Bonferroni-adjusted p < 0.05. These interactions include, for example, blood 2,5-dimethylfuran and blood benzene with an adjusted p value of 4.48E-12 for HDL-C.

By focusing on a comprehensive exposome analysis approach and utilizing a well-established knowledge base, this research can provide important insights for the prevention and management of lipid-based health risk factors. As well as highlight the potential of GE.db to enhance the efficiency and accuracy of -omics research by filtering datasets based on known interactions, thereby facilitating more focused and reliable analyses.

# Methods

***2.1 NHANES Dataset***

The National Health and Nutrition Examination Survey (NHANES) is an ongoing initiative conducted by the Centers for Disease Control and Prevention (CDC) aimed at evaluating the health and nutritional status of the U.S. population. Its primary objectives include identifying risk factors for prevalent diseases and informing the development of public health policies. Data collection encompasses a wide range of participant information, including demographics, dietary recalls, health examinations, toxin exposures, and laboratory measurements, all obtained through structured interviews and physical examinations conducted either at participants' homes or mobile testing centers.

Datasets were extracted from the NHANES website[11](https://paperpile.com/c/HmCkMR/ugJ6), covering the cycles from 1999 to 2018. These datasets were integrated into a comprehensive table, where each row corresponds to a participant and each column represents a specific NHANES variant. The resulting dataset comprises 101,316 participants and 11,274 variants. It is noteworthy that NHANES fields are not consistently maintained across cycles; fields may be modified or discontinued over time, posing challenges for longitudinal analyses.

***2.2 GE.db***

The GE.db module is an integral component of the IGEM system, designed as a comprehensive knowledge base of genomic and exposomic interactions. This module aggregates data from various public databases, providing a curated repository of interactions that can be leveraged to filter high-volume research datasets effectively. The primary purpose of GE.db is to utilize prior knowledge of gene and exposure interactions to filter datasets, thereby retaining only the variables with known interactions. This strategic filtering significantly reduces the number of variables requiring analysis, which, in turn, conserves computational resources, reduces processing time, and minimizes the occurrence of type I errors after multiple testing corrections.

*2.2.1 Data Sources*

GE.db derives its data from multiple reputable public databases that are frequently updated and maintained. For this analysis, the following databases were considered: National Institutes of Health (NIH)[12](https://paperpile.com/c/HmCkMR/rbae), which include extensive repositories of genetic and environmental factors influencing health; Human Metabolome Database (HMDB)[13](https://paperpile.com/c/HmCkMR/3iOq), a detailed resource containing information on small molecule metabolites found in the human body, crucial for understanding metabolic interactions and pathways; Comparative Toxicogenomics Database (CTD)[14](https://paperpile.com/c/HmCkMR/Qyua), which integrates information on chemical-gene/protein interactions, chemical-disease, and gene-disease relationships, facilitating insights into the molecular mechanisms of environmental diseases; and Kyoto Encyclopedia of Genes and Genomes (KEGG)[15](https://paperpile.com/c/HmCkMR/AK9q), which provides comprehensive data on gene functions, biological pathways, diseases, drugs, and chemical substances, supporting the integration of genomic and metabolic information.

At the time of analysis, the GE.db contained 1,057,827 terms grouped into categories such as anatomy, chemicals, diseases, chromosomes, genes, metabolites, pathways, and SNPs, along with 15,667,807 interactions among these terms.

The GE.db module is designed with a flexible architecture that allows for the seamless integration of new data sources. It includes several key components: Term Table, which contains key terms and concepts essential for the analysis, organized into groups and categories for efficient retrieval; Interaction Table, which stores documented interactions between various genomic and exposomic variables, providing a robust foundation for filtering datasets; and Mapping Algorithms, which utilize advanced algorithms to match external data terms to internal GE.db terms, ensuring consistency and reliability in the filtering process.

To maintain the GE.db, the IGEM system employs version control routines and layers of data ingestion and data transformation to fetch data from their sources and transform them into term links (Figure 1). The GE.filter is another component of IGEM that enables various operations on the GE.db knowledge base, including term matching, interaction identification, and data reduction. The IGEM system, along with its modules GE.db and GE.filter, is deployed in a Python environment on an institutional Low-Performance Computing (LPC) cluster. The database utilized is SQLite, which currently has a size of 2.7 GB.

A screenshot of a computer screen

Description automatically generated

Figure 1. Visualization of GE.db workflow from database to interaction term identification.

***2.3 Filter Exposures x Exposures Interactions***

To align the NHANES variables with GE.db, all NHANES variable descriptions (excluding lipid phenotypes and confounders) were processed through the GE.filter function. GE.filter utilizes an internal NLP (Natural Language Processing) engine to identify corresponding GE.db terms based on textual descriptions. This process identified 3,619 NHANES variables related to 534 GE.db terms.

A subsequent review of these related NHANES variables identified 1,136 exposure factors, corresponding to 217 unique terms. These 217 terms were then used as filter parameters for another GE.filter function run, which searched the GE.db knowledge base for all interactions among these terms, resulting in the identification of 382,613 Exposure x Exposure interactions.

***2.3 Phenotypes and Confounders***

Within the NHANES dataset, specific variables were identified as phenotypes and confounders for this analysis. The selected phenotypes included are observed in Table 1. For HDL-C, NHANES altered the calculation method for this indicator over different cycles. Consequently, these three fields were maintained separately, creating three distinct cohorts.The selected confounders included Gender (RIAGENDR), Age (RIDAGEYR), BMI (BMXBMI), Race/Ethnicity (RIDRETH1), and Survey Cycle (SDDSRVYR).

***2.4 Adjusting for Cholesterol Medications***

To account for the influence of cholesterol-lowering medications on lipid measurements, we adjusted the LDL-C and Total Cholesterol (TC) values for participants who reported using statins (Figures S-1,2). This adjustment is crucial for accurately assessing lipid levels and their associations with various exposures, as statins significantly alter cholesterol levels. We utilized the NHANES dataset RXQ\_RX to identify participants who reported using at least one of the following statin components: ATORVASTATIN CALCIUM, SIMVASTATIN, PRAVASTATIN SODIUM, and FLUVASTATIN SODIUM.

For these participants, we adjusted the LDL and TC values as follows: LDL-cholesterol (LBDLDL) values were divided by 0.7 to account for the reduction effect of statins, and Total Cholesterol (LBXTC) values were divided by 0.8 to adjust for statin usage[16](https://paperpile.com/c/HmCkMR/ebFg). By incorporating these adjustments, we enhanced the precision of our lipid measurements, ensuring that our analysis of exposure-lipid interactions was both accurate and reliable.

***2.5 Quality Control (QC)***

Quality Control is a critical step to ensure the integrity, reliability, and validity of the dataset used in the analysis. The IGEM system includes specialized functions that accelerate and assist in the application of QC procedures to OMICS data analyses. The following procedures were applied to the NHANES dataset after filtering and modifications from previous steps.

For continuous data type QC, all variables with more than 90% missing values were removed. The distribution of phenotypes was calculated using the skewness (3(mean-median)/standard deviation.) and all phenotypes were log-transformed to normalize the distribution (Figure S-3).

Participants were then separated into discovery and replication groups for the six cohorts of phenotypes, resulting in twelve datasets. For each dataset, a minimum of 200 observations for categorical and binary exposures was maintained. Only variables present in both discovery and replication datasets for each phenotype were retained to ensure consistency and reliability (Table 1).

***A close-up of a list of cholesterol

Description automatically generated***

***2.6 Run Models (Discovery and Replication)***

The IGEM system, inheriting functionalities from the CLARITE system[17](https://paperpile.com/c/HmCkMR/eAf3), performs interaction analyses by calculating the p-value of the Likelihood Ratio Test (LRT) between two models:

Full Model:

*Yphenotype = β0 + β1term1 + β2term2 + β3(term1 x term2) + β4cov1 + … + βn+1covn* (1)

Restricted Model:

*Yphenotype = β0 + β1term1 + β2term2 + β3cov1 + … + βn+1covn* (2)

The LRT is utilized to compare the fit of the two models, with the full model including the interaction term (*β3(term1×term2)*) and the restricted model excluding it. The analysis involves fitting the full model to the data to obtain the log-likelihood (Lfull) and fitting the restricted model to obtain the log-likelihood (Lrestricted​). The LRT statistic is calculated as:

*D=−2(Lrestricted−Lfull)* (3)

The difference in degrees of freedom between the two models is 1, since the full model has one additional parameter (*β3(term1×term2)*). The p-value is derived from the chi-squared (χ2) distribution with 1 degree of freedom:

*p-value = P(χ2 ≥ D | df = 1)* (4)

The LRT p-values were calculated for each interaction identified in the discovery dataset for each phenotype.

However, in some cases, the p-value of the LRT cannot be calculated. The following messages inform the user of the reasons: Too few complete observations (min\_n filter: N < 200): the number of complete observations is insufficient to perform the analysis, as the minimum required is 200; Both models are equivalent in terms of fit: the two models are equivalent in terms of fit, with no significant difference between them; and No Overlap (min\_n filter: 0 < 200): there is insufficient data overlap to perform the analysis, as the minimum required is 200.

Following the interaction model analysis, the IGEM function was applied to adjust the p-values for multiple testing using both Bonferroni correction and False Discovery Rate (FDR) adjustment. From the discovery analysis, interactions with an FDR-adjusted p-value < 0.1 were filtered. These significant interactions were then isolated in the replication dataset. The same interaction analysis was conducted in the replication cohort, applying identical model specifications and LRT. The replication analysis also included multiple testing corrections using Bonferroni and FDR methods, consistent with the discovery phase. This rigorous approach ensures that the identified interactions are robust and not due to random chance.

**3. Results**

In this study, we examined the interactions between various exposure variables and lipid phenotypes using the NHANES dataset. We performed a comprehensive analysis to identify significant exposure-exposure (ExE) interactions that are associated with lipid levels. Below are the key findings from our discovery and replication datasets. Of all the 26,107 interactions tested that passed QC, for each lipid phenotype a total of 263 interactions were significant in the discovery dataset permitting with an FDR p < 0.1 (Table 2). A total of sixty-one interactions were found to be significant in both discovery and replication when allowing for an FDR p < 0.1 (assorted by lipid phenotype) and twenty-one interactions associated with the HDL-cholesterol outcome were significant with a Bonferroni corrected p < 0.05 (Figure 2). Additionally, these interactions demonstrated consistent directions of effect across both discovery and replication datasets.

A table with numbers and a number of objects

Description automatically generated

A screenshot of a graph

Description automatically generated

Figure 2. The twenty-one significant results starting from the left showcasing Bonferroni adjusted p < 0.05 (direction of effect pointing left is negative and right is positive), the interaction beta for both exposures, and the sample sizes[18](https://paperpile.com/c/HmCkMR/cq9X).

## 3.1 Significant Interactions

## The top three results with the most significant interactions associated with HDL-cholesterol include:

## Docosapentaenoic acid (22:5n-3) (DPA) - arachidic acid (20:0) (Discovery: Bonferroni adjusted p-value = 8.43E-13, β = -0.00014; Replication: Bonferroni adjusted p-value = 3.25E-4, β = -0.00012) (Figure 3A)

* Blood 2,5-dimethyfuran - blood benzene (Discovery: Bonferroni adjusted p-value = 2.75E-7, β = 0.97; Replication: Bonferroni adjusted p-value = 4.48E-12, β = 0.78) (Figure 3B)
* Stearic acid (18:0) - arachidic acid (20:0) (Discovery: Bonferroni adjusted p-value = 8.88E-12, β = -7.79E-6; Replication: Bonferroni adjusted p-value = 3.47E-7, β = -1.26E-5) (Figure 3C)

A screenshot of a graph

Description automatically generated

**4. Discussion**

In this study, we leveraged the comprehensive genomic and exposomic knowledge base provided by the GE.db module of the IGEM system to investigate exposure-exposure (ExE) interactions influencing lipid phenotypes. By utilizing data from the NHANES dataset spanning 1999 to 2018, we identified several significant interactions between various exposures and lipid levels. The replication of these findings across independent datasets underscores the robustness of our approach and highlights the potential of GE.db in facilitating large-scale omics research.

***4.1 Clinical and Public Health Implications***

Our analysis revealed several key interactions, notably DPA and stearic acid with arachidic acid associated with HDL-C. DPA is a known essential omega-3 fatty acid, and stearic and arachidic acid are saturated fatty acids19–21.  These results suggest that specific combinations of environmental exposures may have synergistic effects on lipid metabolism, though most research only touches on their individual effects on lipid profiles. For instance, omega-3 fatty acids, such as DPA, are generally linked with increased HDL cholesterol levels22, while high consumption of saturated fatty acids like arachidic acid may unfavorably affect lipid profiles, potentially leading to elevated LDL-C levels23. Our findings indicate a negative impact on HDL-C when arachidic acid interacts with fatty acids typically associated with positive HDL-C effects, suggesting that arachidic acid could potentially diminish the benefits of HDL-C promoting fatty acids. Other research suggests that stearic acid may have a neutral or even beneficial effect on cholesterol levels, possibly not adversely affecting HDL-C on its own24. However, as seen in our results, when combined with arachidic acid, this interaction could overall have a negative impact, counteracting any neutral or positive effects on HDL-C.

Additionally, the interaction between blood 2,5-dimethylfuran and blood benzene highlights the potential combined impact of exposure to volatile organic compounds (VOCs) on HDL-C levels. Benzene has been observed to increase LDL-C levels which would naturally displace or plateau HDL-C levels procuring a negative effect25,26. Measures of 2,5-dimethyfuran though, have limited research indicating influence on lipids, but may pose health risks similar to other VOCs. These risks can include respiratory irritation, and potential systemic effects that could indirectly affect lipid metabolism and cardiovascular health27–29. Conversely, our results demonstrate a positive interaction effect on HDL-C with benzene and 2,5-dimethylfuran. Therefore, further study of this interaction is warranted, especially considering the known detrimental impact of VOCs on public health. In summary, all these findings have important implications for public health, as they point to the need for considering multiple concurrent exposures in dietary and environmental risk assessments. Public health strategies could be developed to mitigate the combined effects of specific dietary and environmental exposures on lipid metabolism.

***4.2 Methodological Strengths, Limitations, and Future Directions***

A major strength of this study is the use of the GE.db knowledge base, which allowed us to filter high-volume research datasets effectively, focusing only on variables with known interactions. This approach significantly reduced the computational burden and enhanced the reliability of our findings by minimizing type I errors through multiple testing corrections.

The rigorous quality control (QC) procedures, including the categorization of variables, data cleaning, and adjustment for confounders such as statin use, enhanced the integrity and accuracy of our analysis. The split of data into discovery and replication datasets based on NHANES cycles further increased the validity of our results, as significant interactions identified in the discovery phase were consistently replicated.

Despite the robustness of our findings, several limitations warrant consideration. First, the observational nature of the NHANES data limits the ability to infer causal relationships between exposures and lipid levels. Interaction effects, as we have noted, may have opposite signs of effect when compared to the main effect betas, which complicates the interpretability of the results. Other datasets with repeated measures of QC and analysis as we have specified with the NHANES data, can help with cross checking all the betas, refining the elucidation of significant interactions. Future studies could also incorporate longitudinal data and more sophisticated causal inference methods to address this limitation.

Moreover, while our analysis accounted for several covariates, there may be other unmeasured factors that could influence the observed interactions. Further research should aim to include a broader range of potential confounders and explore the underlying biological mechanisms driving these interactions.

Another limitation is the reliance on self-reported data for certain exposures, which may introduce reporting biases. The integration of more objective measures of exposure, such as well-established biomarkers, could enhance the reliability of future analyses.

***4.3 Conclusion***

In conclusion, this study demonstrates the utility of the GE.db module in identifying significant ExE interactions influencing lipid phenotypes. The consistent replication of key interactions across independent variables highlights the robustness of our approach and its potential to uncover future novel insights into the complex interplay between environmental exposures and lipid metabolism. These findings pave the way for future research aimed at understanding and mitigating the multifactorial nature of dyslipidemias, ultimately contributing to improved public health outcomes.

Supported in part by ….[ask Marylyn for funding information]

Code for GE.db, GE.db filter, quality control steps, and supplemental table and figures S-1, S-2, and S-3 are available at <https://github.com/HallLab/pbs_igem/tree/main>.

**References**

1. [Virolainen, S. J., VonHandorf, A., Viel, K. C. M. F., Weirauch, M. T. & Kottyan, L. C. Gene-environment interactions and their impact on human health. *Genes Immun.* **24**, 1–11 (2023).](http://paperpile.com/b/HmCkMR/u5c0)

2. [Ottman, R. Gene-environment interaction: definitions and study designs. *Prev. Med.* **25**, 764–770 (1996).](http://paperpile.com/b/HmCkMR/Xicw)

3. [Castelli, W. P. Lipids, risk factors and ischaemic heart disease. *Atherosclerosis* **124 Suppl**, S1–9 (1996).](http://paperpile.com/b/HmCkMR/gPHd)

4. [Emerging Risk Factors Collaboration *et al.* Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **302**, 1993–2000 (2009).](http://paperpile.com/b/HmCkMR/qQ3p)

5. [Dayimu, A. *et al.* Trajectories of Lipids Profile and Incident Cardiovascular Disease Risk: A Longitudinal Cohort Study. *J. Am. Heart Assoc.* **8**, e013479 (2019).](http://paperpile.com/b/HmCkMR/dTJH)

6. [Amin, K. A., Homeida, A. M., El Mazoudy, R. H., Hashim, K. S. & Garelnabi, M. Dietary Lipids in Health and Disease. *J. Lipids* **2019**, 5729498 (2019).](http://paperpile.com/b/HmCkMR/mjTc)

7. [Hornburg, D. *et al.* Dynamic lipidome alterations associated with human health, disease and ageing. *Nat Metab* **5**, 1578–1594 (2023).](http://paperpile.com/b/HmCkMR/iAlA)

8. [Ritchie, M. D. *et al.* Incorporation of Biological Knowledge Into the Study of Gene-Environment Interactions. *Am. J. Epidemiol.* **186**, 771–777 (2017).](http://paperpile.com/b/HmCkMR/Ku9H)

9. [Term — IGEM 0.1.0 documentation.](http://paperpile.com/b/HmCkMR/YGRT) <https://igem.readthedocs.io/en/latest/ge/md/term.html>[.](http://paperpile.com/b/HmCkMR/YGRT)

10. [National health and nutrition examination survey.](http://paperpile.com/b/HmCkMR/m1QR) <https://www.cdc.gov/nchs/nhanes/index.htm> [(2024).](http://paperpile.com/b/HmCkMR/m1QR)

11. [CDC. Centers for disease control and prevention.](http://paperpile.com/b/HmCkMR/ugJ6) <https://www.cdc.gov/> [(2024).](http://paperpile.com/b/HmCkMR/ugJ6)

12. [Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).](http://paperpile.com/b/HmCkMR/rbae)

13. [Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **35**, D521–6 (2007).](http://paperpile.com/b/HmCkMR/3iOq)

14. [Davis, A. P. *et al.* Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res.* **51**, D1257–D1262 (2023).](http://paperpile.com/b/HmCkMR/Qyua)

15. [Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).](http://paperpile.com/b/HmCkMR/AK9q)

16. [Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758–1766 (2017).](http://paperpile.com/b/HmCkMR/ebFg)

17. [Lucas, A. M. *et al.* CLARITE Facilitates the Quality Control and Analysis Process for EWAS of Metabolic-Related Traits. *Front. Genet.* **10**, 1240 (2019).](http://paperpile.com/b/HmCkMR/eAf3)

18. [Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.* **6**, 18 (2013).](http://paperpile.com/b/HmCkMR/cq9X)

19. [Human Metabolome Database: Showing metabocard for Stearic acid (HMDB0000827).](http://paperpile.com/b/HmCkMR/A9Ad) <https://hmdb.ca/metabolites/HMDB0000827>[.](http://paperpile.com/b/HmCkMR/A9Ad)

20. [Human Metabolome Database: Showing metabocard for Docosapentaenoic acid (22n-3) (HMDB0006528).](http://paperpile.com/b/HmCkMR/4QhR) <https://hmdb.ca/metabolites/HMDB0006528>[.](http://paperpile.com/b/HmCkMR/4QhR)

21. [Human Metabolome Database: Showing metabocard for Arachidic acid (HMDB0002212).](http://paperpile.com/b/HmCkMR/SMNE) <https://hmdb.ca/metabolites/HMDB0002212>[.](http://paperpile.com/b/HmCkMR/SMNE)

22. [Peña-de-la-Sancha, P. *et al.* Eicosapentaenoic and Docosahexaenoic Acid Supplementation Increases HDL Content in n-3 Fatty Acids and Improves Endothelial Function in Hypertriglyceridemic Patients. *Int. J. Mol. Sci.* **24**, (2023).](http://paperpile.com/b/HmCkMR/H07L)

23. [Siri-Tarino, P. W., Sun, Q., Hu, F. B. & Krauss, R. M. Saturated fatty acids and risk of coronary heart disease: modulation by replacement nutrients. *Curr. Atheroscler. Rep.* **12**, 384–390 (2010).](http://paperpile.com/b/HmCkMR/geXW)

24. [Grundy, S. M. Influence of stearic acid on cholesterol metabolism relative to other long-chain fatty acids. *Am. J. Clin. Nutr.* **60**, 986S–990S (1994).](http://paperpile.com/b/HmCkMR/Y5Om)

25. [Tualeka, N. A. R., Martiana, N. T., Wibrata, A. & Rahmawati, P. Effect of food consumption contain glutahione anti-oxidant towards LDL cholesterol concentrations on benzene-exposed-workers at the romokalisari shoe industry, Surabaya. *Indian J. Forensic Med. Toxicol.* (2019) doi:](http://paperpile.com/b/HmCkMR/2lnU)[10.5958/0973-9130.2019.00333.5](http://dx.doi.org/10.5958/0973-9130.2019.00333.5)[.](http://paperpile.com/b/HmCkMR/2lnU)

26. [Ye, L. *et al.* Moderate body lipid accumulation in mice attenuated benzene-induced hematotoxicity via acceleration of benzene metabolism and clearance. *Environ. Int.* **178**, 108113 (2023).](http://paperpile.com/b/HmCkMR/inYq)

27. [Fu, X. *et al.* Airborne 2,5-dimethylfuran as a marker to indicate exposure to indoor tobacco and biomass burning smoke. *Atmos. Environ.* **259**, 118509 (2021).](http://paperpile.com/b/HmCkMR/TZkz)

28. [Jing, L., Chen, T., Yang, Z. & Dong, W. Association of the blood levels of specific volatile organic compounds with nonfatal cardio-cerebrovascular events in US adults. *BMC Public Health* **24**, 616 (2024).](http://paperpile.com/b/HmCkMR/ECRG)

29. [Chen, X. *et al.* Association of Smoking with Metabolic Volatile Organic Compounds in Exhaled Breath. *Int. J. Mol. Sci.* **18**, (2017).](http://paperpile.com/b/HmCkMR/5fgL)