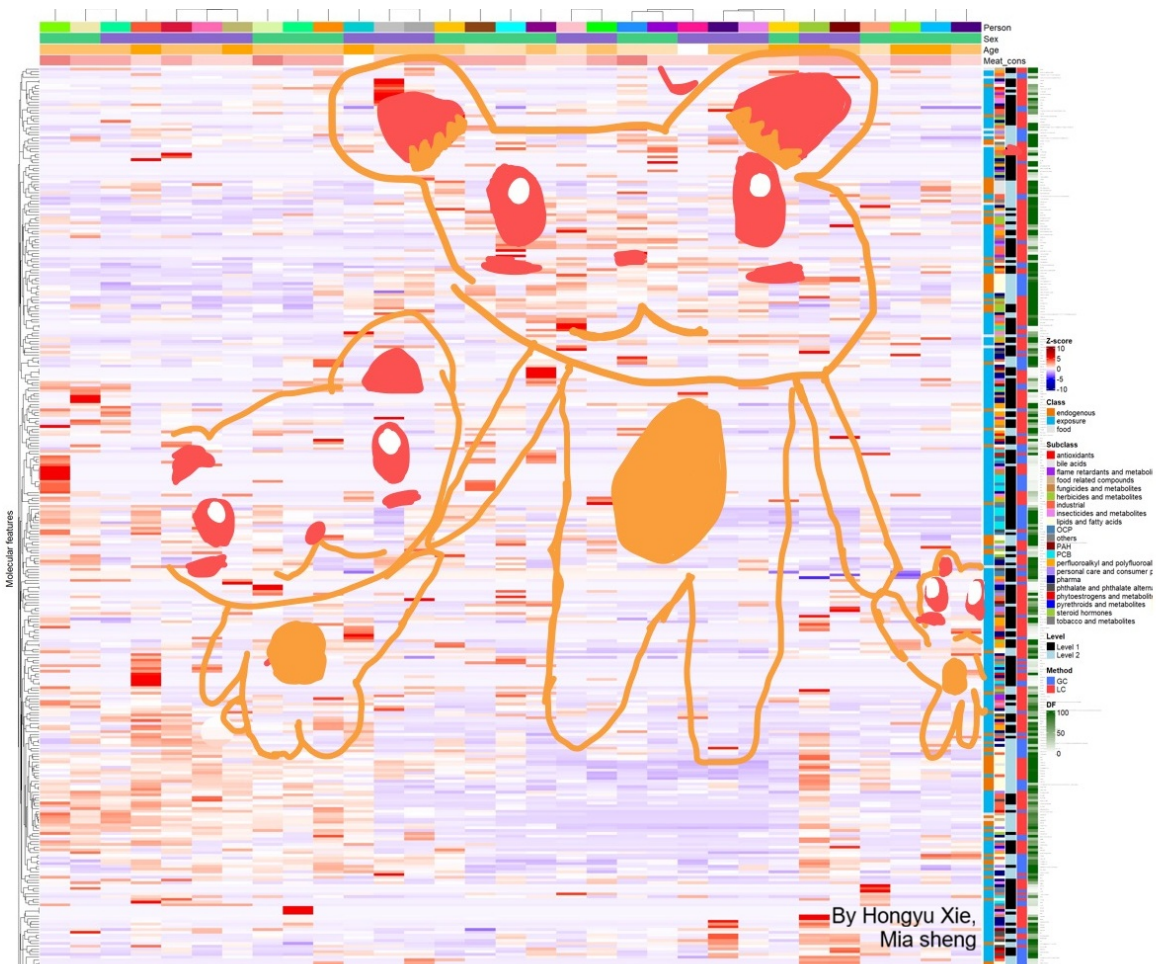


New Analytical Workflows for Comprehensive Chemical Exposomics in Human Plasma

Hongyu Xie



New Analytical Workflows for Comprehensive Chemical Exposomics in Human Plasma

Hongyu Xie

Academic dissertation for the Degree of Doctor of Philosophy in Environmental Sciences at Stockholm University to be publicly defended on Monday 25 May 2026 at 10.00 in DeGeersalen, Svante Arrhenius väg 14 and online via Zoom, public link is available at the department website.

Abstract

The ambition of chemical exposomics to measure all environmental exposures throughout the lifecourse brings analytical challenges because of the large number of environmental chemicals, their diverse physicochemical properties, their dynamic occurrence, and presence in blood at 1000-fold lower concentrations than endogenous metabolites. Gas chromatography high resolution mass spectrometry (GC-HRMS) complements liquid chromatography (LC) by expanding chemical-space coverage into neutral, hydrophobic and semi-volatile substances. However, development of GC-HRMS chemical exposomics has lagged behind LC-, and workflows are still required for sensitive and quantitative detection for multiple priority chemical classes while simultaneously enabling the discovery of novel exposures by nontarget acquisition, robust data processing and appropriate structural annotation frameworks. In Paper I, an actionable annotation scoring framework was developed to incorporate unique GC-HRMS information into annotation confidence level assignments, these GC-specific criteria were applied in Papers II, III and IV.

In Paper II, I developed and validated a chemical exposomics method using isohexane (H) to quantitatively extract prioritized analytes from protein-free acetonitrile-plasma (A-P), which significantly reduced coextracted lipid interference and enabled large-volume injections (25 μ L) to GC. The resulting HA-P method enabled highly sensitive and quantitative detection, achieving a mean method limit of quantification (MLOQ) of 0.09 ng/mL from only 200 μ L of human plasma. Application to 32 individual samples (100 μ L) allowed quantification of 51 targets and a nontarget molecular discovery of 112 additional substances (Level 2, 12.8% high annotation rate). In Paper III, the HA-P method was then applied in a longitudinal study of 46 healthy individuals in a multiomics cohort, whereby each participant donated 6 plasma samples over 2 years. Overall, the GC chemical-exposome was longitudinally unstable, with mean intraclass correlation coefficients (ICCs) of 0.24, significantly lower than for other omics profiles measured (i.e., proteomics, metabolomics, lipidomics and microbiota). Molecular networks and hierarchical clustering analysis revealed structural similarities and correlated co-exposures for numerous chemical classes that may share common exposure sources.

To enable comprehensive chemical exposomics, in Paper IV I developed and validated an integrated sample preparation method that produces two extracts from a single plasma sample, enabling high sensitivity target/nontarget analysis of separate polar and nonpolar analyte fractions. Application to 32 plasma samples allowed overall identification of 204 chemicals at Level 1 covering a wide chemical space, e.g., 11 orders of magnitude in water solubility. Hierarchical clustering of 348 total annotated features revealed a broad range of common or rare co-exposures, many of which were unique to specific individuals or correlated with endogenous metabolites, thereby revealing a high-relevance to precision public health. Overall, the combined methods and frameworks provide new tools to study the human chemical exposome, and the unprecedented datasets from their first applications will guide sampling design (based on low ICCs), comprehensive analysis and data exploration in future studies.

Keywords: *Exposome, human blood, high resolution mass spectrometry, multitargeted, nontargeted, chemical exposomics.*

Stockholm 2026
<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-254153>

ISBN 978-91-8107-600-4
ISBN 978-91-8107-601-1



Stockholm
University

Department of Environmental Science

Stockholm University, 106 91 Stockholm

NEW ANALYTICAL WORKFLOWS FOR COMPREHENSIVE
CHEMICAL EXPOSOMICS IN HUMAN PLASMA

Hongyu Xie



New Analytical Workflows for Comprehensive Chemical Exposomics in Human Plasma

Hongyu Xie

©Hongyu Xie, Stockholm University 2026

ISBN print 978-91-8107-600-4

ISBN PDF 978-91-8107-601-1

Printed in Sweden by Universitetservice US-AB, Stockholm 2026

To my beloved family

Abstract

The ambition of chemical exposomics to measure all environmental exposures throughout the lifecourse brings analytical challenges because of the large number of environmental chemicals, their diverse physicochemical properties, their dynamic occurrence, and presence in blood at 1000-fold lower concentrations than endogenous metabolites. Gas chromatography high resolution mass spectrometry (GC-HRMS) complements liquid chromatography (LC) by expanding chemical-space coverage into neutral, hydrophobic and semi-volatile substances. However, development of GC-HRMS chemical exposomics has lagged behind LC-, and workflows are still required for sensitive and quantitative detection for multiple priority chemical classes while simultaneously enabling the discovery of novel exposures by nontarget acquisition, robust data processing and appropriate structural annotation frameworks. In **Paper I**, an actionable annotation scoring framework was developed to incorporate unique GC-HRMS information into annotation confidence level assignments, these GC-specific criteria were applied in **Papers II, III and IV**.

In **Paper II**, I developed and validated a chemical exposomics method using isohexane (H) to quantitatively extract prioritized analytes from protein-free acetonitrile-plasma (A-P), which significantly reduced coextracted lipid interference and enabled large-volume injections (25 μ L) to GC. The resulting HA-P method enabled highly sensitive and quantitative detection, achieving a mean method limit of quantification (MLOQ) of 0.09 ng/mL from only 200 μ L of human plasma. Application to 32 individual samples (100 μ L) allowed quantification of 51 targets and a nontarget molecular discovery of 112 additional substances (Level 2, 12.8% high annotation rate). In **Paper III**, the HA-P method was then applied in a longitudinal study of 46 healthy individuals in a multiomics cohort, whereby each participant donated 6 plasma sam-

ples over 2 years. Overall, the GC chemical-exposome was longitudinally unstable, with mean intraclass correlation coefficients (ICCs) of 0.24, significantly lower than for other omics profiles measured (i.e., proteomics, metabolomics, lipidomics and microbiota). Molecular networks and hierarchical clustering analysis revealed structural similarities and correlated co-exposures for numerous chemical classes that may share common exposure sources.

To enable comprehensive chemical exposomics, in **Paper IV** I developed and validated an integrated sample preparation method that produces two extracts from a single plasma sample, enabling high sensitivity target/nontarget analysis of separate polar and nonpolar analyte fractions. Application to 32 plasma samples allowed overall identification of 204 chemicals at Level 1 covering a wide chemical space, e.g., 11 orders of magnitude in water solubility. Hierarchical clustering of 348 total annotated features revealed a broad range of common or rare co-exposures, many of which were unique to specific individuals or correlated with endogenous metabolites, thereby revealing a high-relevance to precision public health. Overall, the combined methods and frameworks provide new tools to study the human chemical exposome, and the unprecedented datasets from their first applications will guide sampling design (based on low ICCs), comprehensive analysis and data exploration in future studies.

Keywords: exposome, human blood, high resolution mass spectrometry, multitargeted, nontargeted, chemical exposomics

Popular Science Abstract

Before birth and throughout our lives, we are exposed to thousands of chemicals from our environment—through the air we breathe, what we eat and drink, and the products we use. It is clear from studies of certain priority contaminants that these exposures present a risk to our health, but understanding the links between disease and complex combinations of exposures, including at what dosages, for how long, and at what life stage, remains a major scientific challenge for research. This emerging field of chemical exposomics aims to comprehensively measure the full range of environmental chemicals to which we are exposed, and their potential impacts on molecular phenotype or disease, making human blood a strategic sample for research. However, tracking such a vast and diverse array of substances at low concentrations in blood—in a matrix dominated by endogenous lipids and metabolites—requires new sensitive analytical tools to be developed and validated, supported by powerful data analyses to discover and characterize novel exposures.

In this thesis, new methods were developed to greatly improve how we detect and analyze trace levels of environmental organic chemicals in human blood. By combining two advanced techniques—**gas chromatography** and **liquid chromatography**, both coupled to **high-resolution mass spectrometry**—we can now measure common priority chemicals and discover novel or rare substances with high sensitivity and accuracy, and across a wider scope of molecules than ever before. A novel scoring system was also created to help scientists more clearly communicate the identity of unknown compounds detected.

A major breakthrough here was a sample preparation method that allowed detection of trace-level chemicals in human blood with little interference from lipids during gas chromatography. This enabled the sensitive analysis of very small blood volumes and application to individuals sampled repeatedly over a

period of 2 years, providing a first look at how unique and variable a person's chemical exposures are over time. Finally, by integrating the method with a complementary liquid chromatography workflow, we effectively doubled the range of detectable molecules and provide a more comprehensive characterization of the human chemical exposome for Swedish adults.

Altogether, this work lays a strong foundation for future exposomics research. Through the demonstration of new analytical workflows, this work guides the design of future studies and opens the door to deeper molecular insights into how everyday chemical exposures influence our biology and long-term health.

Populärvetenskaplig sammanfattning

Redan innan vi föds och under hela livet utsätts vi för tusentals kemikalier via luften vi andas, det vi äter och dricker, och de produkter vi använder. Forskning har länge visat att vissa av dessa ämnen kan påverka vår hälsa negativt. Men att förstå hur sjukdomar uppstår till följd av komplexa kombinationer av kemikalier—i olika doser, under olika exponeringstider, och i olika stadier av livet—är fortfarande en stor vetenskaplig utmaning. Det framväxande forskningsfältet kemisk exposomik syftar till att kartlägga hela spektrumet av miljökemikalier som vi utsätts för att undersöka hur dessa påverkar biologiska processer och sjukdomsrisik. Blod är ett särskilt värdefullt prov i detta sammanhang. Men att spåra en så stor och varierad mängd substanser i låga koncentrationer i blod—en matris som domineras av endogena lipider och metaboliter—kräver nya känsliga analytiska verktyg som stöds av kraftfull dataanalys för att upptäcka och karakterisera det kemiska exposomet.

I denna avhandling utvecklades metoder som avsevärt förbättrar möjligheten att upptäcka och analysera mycket låga nivåer av organiska miljökemikalier i blod. Genom att kombinera två tekniker—gaskromatografi och vätskekromatografi, båda kopplade till högupplösande massespektrometri — kan både kända och tidigare oupptäckta ämnen mätas med hög känslighet och noggrannhet. Ett nytt poängsystem utvecklades för att hjälpa forskare att tydligare och mer enhetligt kommunicera graden av säkerhet vid identifiering av substanser.

Ett viktigt genombrott var en ny provberedningsmetod som gjorde det möjligt att detektera spårnivåer av kemikalier i blod, med minimal påverkan från störande lipider vid gaskromatografi. Detta möjliggjorde känslig analys av mycket små blodvolymmer och tillämpades på prover tagna från samma individer under två års tid, vilket gav en unik inblick i hur en persons kemiska ex-

poneringar förändras över tid. När metoden kombinerades med det vätskekromatografiska arbetsflödet fördubblades antalet molekyler som kunde upptäckas, vilket resulterade i den mest omfattande kartläggningen hittills av det kemiska exposomet hos svenska vuxna.

Sammantaget lägger detta arbete en stark grund för framtida forskning inom exposomik. De nya analytiska arbetsflödena som utvecklats banar väg för mer detaljerade studier av hur vardagliga kemiska exponeringar påverkar människans biologi och långsiktiga hälsa.

科普摘要

在出生前及整个生命历程中，人类通过呼吸的空气、摄入的食物和水，以及日常接触的各类产品，暴露于成千上万种环境化学物质中。现有研究已证实，某些优先污染物对人体健康存在明确风险。然而，要理解疾病与复杂化学暴露组合之间的关系——包括暴露剂量、持续时间、生命阶段——仍面临重大挑战。新兴的“**化学暴露组学**”领域，正是为了系统表征人类所接触的全部环境化学物质及其对分子表型和疾病风险的潜在影响而发展起来的。在这一背景下，人类血液成为具有重要价值的研究样本。然而，血液是一种以内源性脂类和代谢物为主的复杂基质，要从中检测低浓度的多种污染则需要开发并验证新型高灵敏度分析工具，同时结合先进的数据分析方法，以发现并表征新的暴露来源。

本论文开发了一种新型分析方法，显著提升了人体血液中痕量环境有机污染物的检测与表征能力。该方法创新性地整合**气相色谱法**和**液相色谱法**，并分别与**高分辨率质谱**联用，使得我们不仅可以高灵敏度、高准确性地测量常见的优先污染物，还能发现新的或罕见化合物，显著拓展了检测范围。同时，本研究还开发了一种新的评分系统，可帮助科学家更清晰地报告未知化合物的鉴定结果。

本研究的一项重大突破在于开发出一种创新地样品前处理方法，实现了气相色谱分析中人体血样痕量化学物质的高灵敏检测，且几乎不受脂质基质干扰。该技术显著降低了分析所需的血液样本量，使得对两年间多次采集的个体样本进行纵向追踪成为可能，从而首次揭示了个人化学暴露的时间维度上的独特性与动态多样性。通过将该方法与互补的液相色谱流程相结合，我们成功将可检测的分子覆盖范围扩大了一倍，提供了目前针对瑞典成年人化学暴露组最全面的表征。

总体而言，这项工作暴露组学的未来发展奠定了坚实基础。通过构建创新的分析流程，本研究不仅为后续实验设计提供了方法学指导，更为深入探究日常化学暴露对人体生理机能及长期健康的影响开辟了新路径。

List of papers

1. Koelmel JP, **Xie H**, Price EJ, Lin EZ, Manz KE, Stelben P, Paige MK, Papazian S, Okeme J, Jones DP, Barupal D, Bowden JA, Rostkowski P, Pennell KD, Nikiforov V, Wang T, Hu X, Lai Y, Miller GW, Walker DI, Martin JW, Pollitt KJG. An Actionable Annotation Scoring Framework for Gas Chromatography-High-Resolution Mass Spectrometry. *Exposome 2*, osac007 (2022)
2. **Xie H**, Sdougkou K, Papazian S, Bonnefille B, Bergdahl IA, Rantakokko P, Martin JW. Chemical Exposomics in Human Plasma by Lipid Removal and Large-Volume Injection Gas Chromatography–High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* 58, 40, 17592–17605 (2024)
3. **Xie H**, Sunyer-Caldú A, Papazian S, Bonnefille B, Sdougkou K, Uhlén M, Bergström G, Martin JW. Longitudinal Chemical Exposomics by Gas Chromatography-High Resolution Mass Spectrometry in Blood Plasma Sampled Over 2-Years in a Multiomic Wellness Cohort, Manuscript
4. **Xie H**, Bonnefille B, Sunyer-Caldú A, Sdougkou K, Chen X, Rian MB, Wang Y, Papazian S, Zhu L, Martin JW. Two-for-One Chemical Exposomics in Human Plasma by an Integrated Sample Preparation Workflow for Target and Non-target Analysis by both GC- and LC-HRMS, Submitted

Contribution

Paper I

I participated in the discussion of the framework, conducted the experimental work and data analysis of spiked human serum sample extracts, and participated in the manuscript writing, reviewing and editing.

Paper II

I participated in conceptualization, played a principle role in the methodological design, experimental execution, data analysis and visualization, and took the lead role in writing and editing the manuscript.

Paper III

I participated in conceptualization, played a principle role in the methodological design, experimental execution, data analysis and visualization, and took the lead role in writing and editing the manuscript.

Paper IV

I designed the concept and led the methodological development, conducted the experimental work, data analysis and visualization, and took the lead role in writing and editing the manuscript.

List of papers not included in this thesis

1. Zhang P, Carlsten C, Chaleckis R, Hanhineva K, Huang M, Isobe T, Koistinen VM, Meister I, Papazian S, Sdougkou K, **Xie H**, Martin JW, Rappaport S, Tsugawa H, Walker DI, Woodruff TJ, Wright RO, Wheelock CE. Defining the Scope of Exposome Studies and Research Needs from a Multidisciplinary Perspective. *Environ. Sci. Technol. Letters* 8, 839–852 (2021).
2. Papazian S, D’Agostino LA, Sadiktsis I, Froment J, Bonnefille B, Sdougkou K, **Xie H**, Athanassiadis I, Krishnakant Budhavant K, Dasari S, Andersson A, Gustafsson Ö, Martin JW. Nontarget Mass-Spectrometry and In-Silico Molecular Characterization of Air Pollution from the Indian Subcontinent. *Commun. Earth Environ.* 3, 35 (2021).
3. Papazian S, Fornaroli C, Bonnefille B, Pesquet E, **Xie H**, Martin JW. Silicone Foam for Passive Sampling and Nontarget Analysis of Air. *Environ. Sci. Technol. Lett.* 10, 989–997 (2023).
4. Sdougkou K, **Xie H**, Papazian S, Bonnefille B, Bergdahl IA, Martin JW. Phospholipid Removal for Enhanced Chemical Exposomics in Human Plasma. *Environ. Sci. Technol.* 57, 10173–10184 (2023).
5. Lai Y, Koelmel JP, Walker DI, Price EJ, Papazian S, Manz KE, Castilla-Fernández D, Bowden JA, Nikiforov V, David A, Bessonneau V, Amer B, Seethapathy S, Hu X, Lin EZ, Jbebli A, McNeil BR, Barupal D, Ceresa M, **Xie H**, Kalia V, Nandakumar R, Singh R, Tian Z, Gao P, Zhao Y, Froment J, Rostkowski P, Dubey S, Coufalíková K, Seličová H, Hecht H, Liu S, Udhani HH, Restituto S, Tchou-Wong KM, Lu K, Martin JW, Warth B, Godri Pollitt KJ, Klánová J, Fiehn O, Metz TO, Pennell KD, Jones DP, Miller GW, High-Resolution Mass Spectrometry for Human Exposomics: Expanding Chemical Space Coverage. *Environ. Sci. Technol.* 58, 29, 12784–12822 (2024).
6. Sdougkou K, Papazian S, Bonnefille B, **Xie H**, Edfors F, Fagerberg L, Uhlen M, Bergstrom G, Martin LJM, Martin JW. Longitudinal Exposomics in a

- Multiomic Wellness Cohort Reveals Distinctive and Dynamic Environmental Chemical Mixtures in Blood. *Environ. Sci. Technol.* 58, 37, 16302–16315 (2024).
7. Møller MT, Birch H, Sjøholm KK, Skjolding LM, **Xie H**, Papazian S and Mayer P. Determining Marine Biodegradation Kinetics of Chemicals Discharged from Offshore Oil Platforms—Whole Mixture Testing at High Dilutions Increases Environmental Relevance. *Environ. Sci. Technol.* 58, 39, 17454–17463 (2024).
 8. Sunyer-Caldú A, **Xie H**, Bonnefille B, Lee JE, Lee B, Papazian S, Martin JW. Silicone-Foam Passive Air Samplers for Combined Target and Nontarget Chemical Profiling and Toxicity Assessment of Airborne Exposomes. *Environ. Sci. Technol.* 60, 7, 5628–5644 (2026).
 9. Chen X, Wang R, Yang Y, **Xie H**, Martin JW, Shan G, Zhu L. Characterization of Human Exposome and Risk of Reproductive Hormone Disruption by Integrating GC-HRMS Non-target Analysis, Molecular Networking and Knowledge Transfer Learning. Submitted.

Abbreviations

| | |
|-----------------|--|
| GC | Gas Chromatography |
| HRMS | High Resolution Mass Spectrometry |
| LC | Liquid Chromatography |
| HA-P | Hexane Extraction of Acetonitrile-Plasma |
| MLOQ | Method Limit of Quantification |
| ICC | Intraclass Correlation Coefficient |
| K _{ow} | Octanol-Water Partition Coefficient |
| EI | Electron Ionization |
| RI | Retention Time Index |
| PCB | Polychlorinated Biphenyl |
| XLE | Express Liquid Extraction |
| AGC | Automatic Gain Control |
| VIP | Swedish Västerbotten Intervention Programme Cohort |
| S3WP | Swedish SciLifeLab SCAPIS Wellness Profiling Program |
| EMR | Enhanced Matrix Removal |
| LLE | Liquid-Liquid Extraction |
| SPE | Solid Phase Extraction |
| IDL | Instrumental Detection Limit |
| PTV | Programmable Temperature Vaporizer |
| GNPS | Global Natural Product Social Molecular Networking |
| DIA | Data Independent Acquisition |

| | |
|------|--------------------------------------|
| HRMF | High Resolution Mass Filter |
| RSI | Reverse Search Index |
| LVI | Large Volume Injection |
| TIC | Total Ion Chromatogram |
| BDE | Polybrominated Diphenyl Ether |
| PAH | Polycyclic Aromatic Hydrocarbon |
| OCP | Organochlorine Pesticide |
| HCB | Hexachlorobenzene |
| DF | Frequency of Detection |
| HCA | Hierarchical Clustering Analysis |
| PFAS | Per- and Poly-fluoroalkyl Substances |

Contents

| | |
|--|------|
| Abstract | i |
| Popular Science Abstract..... | iii |
| Populärvetenskaplig sammanfattning | v |
| 科普摘要 | vii |
| List of papers..... | viii |
| Contribution | ix |
| List of papers not included in this thesis | x |
| Abbreviations | xii |
| Contents | xiv |
| 1. Introduction..... | 1 |
| 1.1 Chemical exposomics in blood and its challenges..... | 1 |
| 1.2 Gas chromatography - high resolution mass spectrometry (GC-HRMS) for chemical exposomics | 3 |
| 1.3 Longitudinal study of the chemical exposome in human blood plasma | 6 |
| 1.4 Analyzing the chemical exposome in blood plasma by a combination of LC- and GC-HRMS workflows..... | 7 |
| 2. Aims | 9 |
| 3. Materials and methods..... | 11 |
| 3.1 Selection of multi-class target analytes | 11 |
| 3.2 Human plasma/serum samples | 11 |
| 3.3 Sample preparation methods | 11 |
| 3.3.1 Sample preparation for nonpolar analytes (Paper I-III)..... | 11 |
| 3.3.2 Sample preparation for polar analytes (Paper IV)..... | 12 |
| 3.4 Instrumental analysis | 13 |

| | |
|--|----|
| 3.4.1 GC-HRMS analysis..... | 13 |
| 3.4.2 LC-HRMS analysis | 13 |
| 3.5 Data analysis and visualization | 14 |
| 3.5.1 Multi-class target data analysis | 14 |
| 3.5.2 Nontarget data analysis | 14 |
| 3.5.3 Visualization and statistics | 15 |
| 4. Results and discussion | 16 |
| 4.1 Annotation framework for GC-HRMS nontarget analysis in Paper I | 16 |
| 4.2 Method development and validation for GC-HRMS analysis in Paper II | 19 |
| 4.3 Multi-class target and nontarget analysis in 32 individual Swedish plasma samples in Paper II | 22 |
| 4.4 Longitudinal GC-HRMS Exposomics in a Multiomic Wellness Cohort in Paper III | 24 |
| 4.5 Sample preparation method for dual analysis by LC and GC-HRMS in Paper IV | 32 |
| 5. Conclusions and perspectives | 40 |
| Acknowledgements | 42 |
| References | 44 |

1. Introduction

1.1 Chemical exposomics in blood and its challenges

To encourage a new generation of precision health research that may eventually complement genomics, the human exposome was defined in 2005 as the totality of environmental exposures throughout an individual's lifespan, from conception onwards.¹ Extensive research has already established that specific priority chemical exposures, or environmental pollution more generally, contributes significantly to disease development,^{2,3} through inhalation, ingestion, dermal contact, and influenced by occupation and various lifestyle factors.⁴⁻⁶ Nevertheless, the vast majority of the hundreds-of-thousands of chemicals in commerce have never been monitored in people or the environment, thus the development of comprehensive methods for measuring the chemical component of the exposome—in this thesis called *chemical exposomics*—remains a vital challenge⁷ to overcome before science can fully understand the exposome and its connections to human health and disease.^{8,9}

Blood represents an ideal matrix for chemical exposomics due to its circulation to all tissues in the body, its content of small molecules (i.e., environmental chemicals and metabolites), large molecules (i.e., proteins and mRNA) and cells (i.e., DNA) for parallel analysis by other 'omic' technologies, and its routine availability in existing health studies. However, the ambition to measure the totality of environmental substances in blood quickly encounters practical analytical challenges. For starters, organic environmental contaminants in human blood have been reported to vary widely in hydrophobicity, spanning 17 orders of magnitude in the n-octanol-water partition coefficient (K_{ow}), and their concentrations range across 11 orders of magnitude (160 fM to 140 mM),^{10,11} making analysis by any single method unlikely with current technology. This is partly why historical approaches have used targeted analysis, and few studies have been able to report on multiple chemical classes in the same person. With a focus on public health, major blood biomonitoring

initiatives in the past have used a battery of targeted methods for population surveillance,¹² and have not accounted for multiple chemical classes in individuals that may interact and adversely impact on health. The challenge of the chemical exposome is partly addressed through advances in technology and commercial instruments (i.e., rapid full-scanning high resolution mass spectrometry (HRMS) instruments) but sensitive sample-preparation and instrumental methods need to be established for multi-class analysis of priority substances (Figure 1). Moreover, because there are so many chemicals in global commerce (350,000),¹³ chemical exposomics methods should also be capable of molecular discovery to identify molecules that were not previously known to be present in human blood.

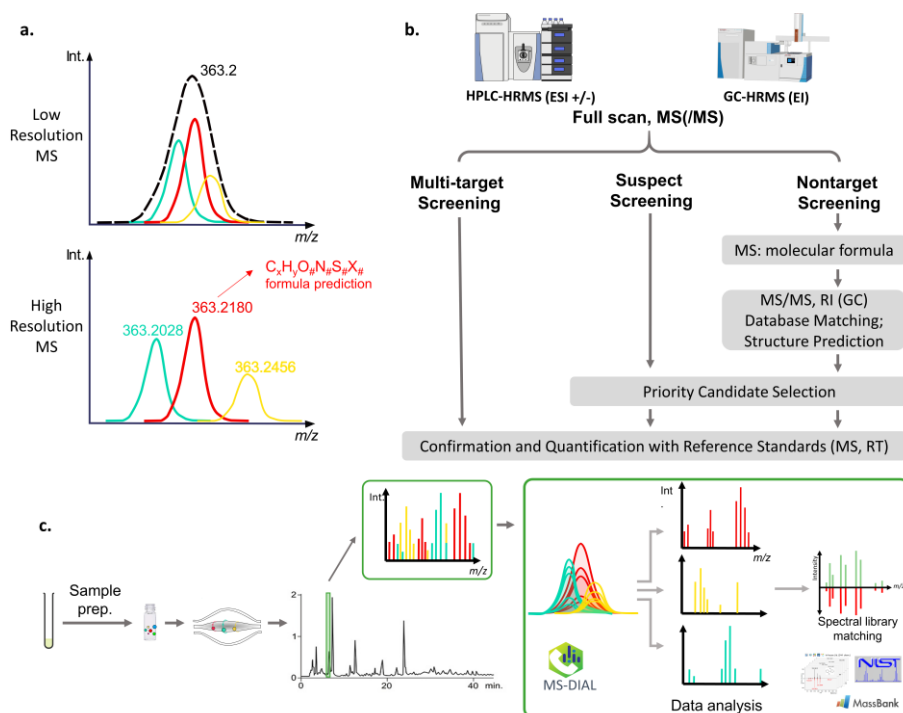


Figure 1. Advantages of high mass accuracy and mass spectral resolution (a), which supports a capability for combined target, suspect screening and nontarget analysis in one run (b), and supported by data processing workflows for spectral deconvolution and spectral matching to libraries (c). Some images in Panel C are accredited to May-Britt Rian (Stockholm University) who granted permission for their display here.

The above inherent challenges are further intensified by the complex blood matrix, which contains abundant lipid classes and major endogenous metabolites that may obscure comparatively minor signals from environmental chemicals. While multi-class target, and combined target/nontarget analytical methods are emerging for chemical exposomics in blood, current sample preparation techniques are often adapted from metabolomics protocols^{14,15} and not tailored to overcome the unique challenges of the chemical exposome. This presents a critical limitation, as chemical exposomics in blood will require detection of trace-level exogenous compounds that are typically present at concentrations that are three orders of magnitude lower than endogenous substances. This challenge is exacerbated considering the low sample volumes (i.e., 50–200 μ L) of blood that are collected and archived in typical cohort studies, and where availability of more than one sample is often constrained.

1.2 Gas chromatography - high resolution mass spectrometry (GC-HRMS) for chemical exposomics

With the aim of developing methods for the comprehensive characterization of environmental contaminants in human blood, we previously summarized the chemical-space for 299 representative target analytes that are routinely biomonitoring in major biomonitoring programmes,¹⁶ including HBM4EU and US NHANES (Figure 2). Due to their broad range of water solubilities (18 orders of magnitude) and K_{ow} (15 orders of magnitude), their comprehensive analysis by the combination of liquid chromatography (LC) and GC mass spectrometry techniques would be required.^{16–18}

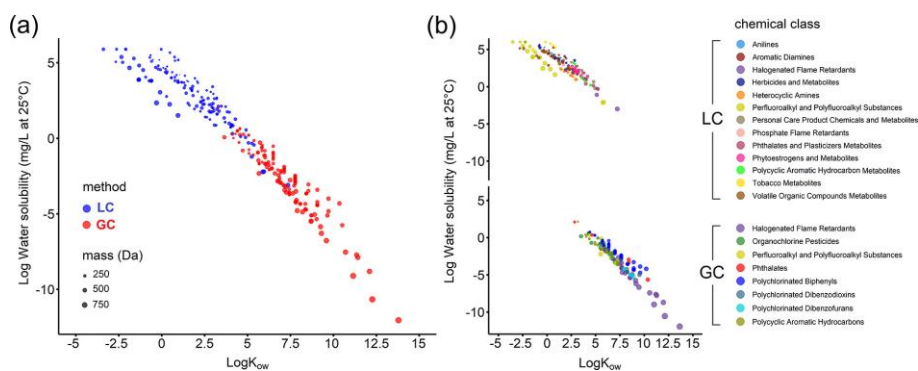


Figure 2. Full coverage of the chemical exposome will require multiple instrumental strategies, as shown by the chemical space of 299 exogenous analytes routinely monitored in human biofluids in biomonitoring programmes. a) Measurement of the analytes will require a combination of LC- and GC-based approaches that are b) dependent on the chemical class. From Zhang et al.¹⁶

GC-HRMS analysis is a critical complement to LC-HRMS, as GC- with an electron ionization (EI) interface to mass spectrometry enables the detection of semi-volatile, neutral and lower polarity analytes that are not adequately ionizable by LC-based atmospheric pressure ionization methods, such as by electrospray. However, commercially available instruments combining GC with rapid full-scanning HRMS detectors is relatively recent,¹⁹ and GC-based nontarget analysis remains an emerging field, and still lacks common standard criterion for communicating confidence of molecular discoveries. The establishment of a standardized annotation framework is crucial for communicating uncertainty in compound identification or annotation, particularly in HRMS-based nontargeted and suspect screening studies where tentative structural assignments and the potential for false positives must be considered. While Schymanski et al. introduced one of the first confidence scoring frameworks for LC-HRMS/MS,²⁰ its direct applicability to GC-HRMS was strained due to fundamental differences in chromatography, ionization, and data acquisition or processing workflows. A preceding framework for GC-HRMS (with EI) categorized feature annotation confidence into four levels, ranging from confirmed identifications (Level 1), spectra matched to a database (Level 2a) or

experimental data (Level 2b), molecular formula assigned (Level 3), or a molecular feature defined only by its exact mass (Level 4).²¹ Although this represented progress, the proposed confidence parameters may still lead to high false-positive and false-negative rates because EI is a hard ionization technique, resulting in some produced spectra missing a molecular ion, or other important fragment ions, and thus spectral evidence alone is not a strong enough criteria. GC's long-established retention time index (RI), primarily by the Kovats method which standardizes the retention time of analytes by the relative elution order of a linear alkane standard mixture, partly compensates for this limitation and provides additional confidence. For example, the RI of chemical polychlorinated biphenyl (PCB)-169 was calculated as 2522 using Kovats index equation,^{22,23} as it eluted between n-C₂₅H₅₂ and n-C₂₆H₅₄. This RI value remains stable in a narrow range, regardless of instruments or methods used.

In **Paper I**, in collaboration with a global community of GC-HRMS users, and considering the distinctive chemical evidence obtained from these powerful new instruments, we developed a new annotation scoring framework for nontarget GC-HRMS to aid and standardize communication of confidence in annotating the chemical structure or formula of unknown substances.

The application of GC-HRMS to chemical exposomics in human blood has gained increasing attention because of the complementary chemical information provided, relative to LC-HRMS. During the early phases of this thesis research, two sample preparation methods for chemical exposomics of blood plasma by GC-HRMS were reported which were simple, rapid and potentially scalable for a wide range of multiclass analytes.^{24,25} However, these were not optimized to enhance or maximize method sensitivity, nor evaluated for matrix effects caused by major endogenous interferences, in particular from major lipid coextractives. For example, I suspected major interferences in the express liquid extraction (XLE) method, in part because no nontarget analytes

were discovered in its application to authentic plasma samples. Abundant lipids are typically removed in targeted GC-HRMS workflows, as these are known to compromise method robustness and reduce sensitivity.^{26–30}

In this thesis it was an overarching hypothesis that abundant lipid co-extractives might also hinder molecular discovery by nontarget GC-HRMS, not only due to suppression of analyte signals, but due to the complex spectral interferences (i.e., lipid fragments) throughout the chromatogram that might make it difficult to deconvolute the spectra (Figure 1c) of minor environmental chemicals. Additionally, because I used Orbitrap™ HRMS analyzers in the current thesis work, it is important to understand that a high signal of background ions from the matrix (i.e., lipid fragments) can reduce ion injection times to the Orbitrap due to the automatic gain control (AGC) setting on these instruments; designed to prevent space-charging and loss of mass accuracy.³¹ This may compromise method sensitivity for trace analytes. The benefits of cleanup for nontarget screening in silicone personal passive sampler extracts by GC-Orbitrap-HRMS has been previously documented.²¹

In **Paper II**, a simple and potentially scalable protocol, using isohexane (H) to liquid–liquid extract protein-free acetonitrile-plasma (A-P), was developed and validated. The HA-P method was optimized for small-volume samples of human plasma (100–200 μL), with emphasis on minimizing lipid co-extractives and maximizing method sensitivity for 103 priority target analytes by large volume injection. The validated methodology was subsequently applied to a subset of adult plasma samples ($n = 32$, 100 μL each) in a combined test of multi-class target and nontarget chemical exposomics.

1.3 Longitudinal study of the chemical exposome in human blood plasma

As an individual's location and environmental conditions are dynamic, and lifestyle factors vary over time and across individuals, understanding both the

temporal variability and between-person differences in plasma chemical exposomes is critical, but remains an important data-gap in the design and interpretation of human health studies. While exposome studies are inherently encouraged to record exposures ‘over the life-course’, longitudinal study designs, i.e., where participants exposures are recorded more than once over time, remain rare in the exposome field. Moreover, recent examples of longitudinal studies have mainly focused on LC-HRMS metabolomics, and only over a few months in a small group of individuals.^{32–35} Applying GC-HRMS in a longitudinal study of the plasma exposome will enable fundamental understanding of the temporal stability of each substance, while also allowing a more comprehensive analysis of an individual's chemical exposure profile. Nonpolar chemical exposures are most often determined for specific priority chemical classes only,^{6,36–39} ignoring emerging, or previously overlooked chemicals in blood. In **Paper III**, we analyzed the chemical exposome of 46 healthy Swedish health adults, with each participant providing six plasma samples over a two-year period for multiomics analysis. Comprehensive multi-class target and nontarget analysis by the HA-P method and GC-HRMS revealed comprehensive chemical exposomes that could be visualized through hierarchical clustering, examined for co-exposure patterns, temporal trends and sex-differences, and most importantly to quantify the inter- and intra- individual variability over time through measures of intraclass correlation coefficients (ICCs). The chemical exposome ICCs were then be compared to parallel measures of plasma proteomes, metabolomes and gut microflora.

1.4 Analyzing the chemical exposome in blood plasma by a combination of LC- and GC-HRMS workflows

Comprehensive chemical exposomes in human blood plasma require the dual application of both LC- and GC-HRMS analysis to cover the numerous chemicals known to be present with a wide range of physicochemical properties.^{16,18} However, multiple samples or sufficiently large sample volumes are

not commonly available in many cohort biobanks to enable sensitive and quantitative analyses by dual-mode chemical exposomics, particularly in multiomic investigations where various other molecular profiling technologies also require a separate sample aliquot. Thus, for future studies prioritizing chemical exposomics, it would be ideal if one typical sample aliquot (e.g., 100-200 μ L) could be used to create two extracts, one for LC- and one for GC-HRMS. Very few studies have reported combined LC and GC analysis of the same blood sample aliquot, and the supporting methods suffer from low recoveries for many target analytes, which affects accuracy and method sensitivity, and nontarget analysis was not considered.^{40,41} In **Paper IV**, starting from the validated and sensitive HA-P method for GC-HRMS chemical exposomics⁴² we preserve the extracted A-P subnatant (rather than disposing it) and treat this by an amended method of Sdoukhou et al. for LC-HRMS chemical exposomics,⁴³ and validate this integrated workflow for ‘2-for-1 chemical exposomics’, whereby one plasma aliquot enables both GC- and LC-HRMS analysis of the chemical exposome. The validated method was contrasted with the direct method of Sdoukhou et al. and successfully applied to the same 32 samples (100 μ L each) examined in **Paper II**, thereby revealing many new exposures and co-exposures of high relevance to precision health research.

2. Aims

The major aim of this thesis was to develop new, comprehensive and sensitive chemical exposomics sample preparation workflows for blood plasma in support of target and nontarget chemical exposomics by GC- and LC-HRMS. Considering the unique challenges, and few previous applications of GC-HRMS in this field, a secondary aim was to apply the method(s) to human samples from existing biobanks while developing and testing data processing, data visualization, and molecular annotation workflows in support of quantifying or characterizing diverse (co-)exposures in unprecedented detail. To the extent possible, this work would include confirmation of novel environmental substances in human blood. To inform future large-scale investigations, the third and final aim was to apply the GC-HRMS method in the first longitudinal study of the nonpolar chemical exposome, with a major objective to quantify the inter- and intra-individual variation for hundreds of target and nontarget analytes over a 2-year period. The specific aims are integrated and are spread across 4 papers composing this thesis:

In **Paper I** I aimed to establish a standardized annotation framework for GC-HRMS nontarget features by integrating unique GC-HRMS data to the existing LC-HRMS framework (Schymanski scale) to ensure clear and transparent communication of confidence in chemical identification or annotation.

In **Paper II** I aimed to develop, optimize, and validate a sample preparation method that minimizes coextracted lipid interference, thereby enabling large-volume injection for sensitive and quantitative analysis of multi-class target analytes while simultaneously screening for unanticipated chemicals in individual human plasma samples by nontarget GC-HRMS.

In **Paper III** I aimed to apply the validated HA-P extraction and GC-HRMS workflow with open-science data processing to plasma samples from a unique

longitudinal and multiomic wellness cohort. A primary objective was to quantify ICCs for all targets and confidently annotated analytes, to compare these measures of ‘stability’ to other biomolecular profiles measured in the same participants, and to characterize and visualize patterns in comprehensive chemical exposomes across hundreds of samples to maximize knowledge of relevance to precision health.

Paper IV describes my work to find sample processing efficiencies that can bridge GC- and LC-HRMS chemical exposomics workflows in an integrated manner. Here I aimed to validate an integrated method for combined target and nontarget analysis of chemical exposomes in single small-volume sample aliquots by both GC- and LC-HRMS analysis; while maintaining sensitivity and chemical-space coverage. In addition to many novel chemical discoveries, in the unprecedented combined dataset it was my objective to visualize and describe unique and common patterns of (co-)exposure, and exposome-metabolome interactions among all samples from individual participants.

3. Materials and methods

3.1 Selection of multi-class target analytes

Chemicals reported in human biofluids (i.e., blood, urine) were collected from comprehensive biomonitoring projects (e.g. HBM4EU, US NHANES). For **Papers II-III**, 103 representative targets from 6 chemical classes were selected, covering a wide range of physicochemical properties. These compounds, typically exhibiting high detection frequencies, were used for method development and validation, while also serving as a priority target analytes in chemical exposomics by GC-HRMS.

3.2 Human plasma/serum samples

Commercial human serum (Merck, Germany, human male AB plasma-derived, origin USA) was used in **Paper I** (software and annotation framework evaluation) and **Paper II** (method development and validation). In **Paper II** and **Paper IV**, the optimized and validated chemical exposomics workflows were applied to 32 individual human plasma samples (100 μ L) from the Swedish Västerbotten Intervention Programme (VIP) cohort.⁴⁴ For quality control purposes in **Paper II-IV**, a pooled reference plasma was also prepared from previously-thawed residual plasma samples from the VIP cohort. For the longitudinal study in **Paper III**, the 276 plasma samples (50–200 μ L each) were from a subset of 46 individuals who were each sampled 6 times over 2 years in the Swedish SciLifeLab SCAPIS Wellness Profiling (S3WP) program.⁴⁵

3.3 Sample preparation methods

3.3.1 Sample preparation for nonpolar analytes (**Paper I-III**)

In **Paper II**, method development included testing various lipid removal approaches, including sulfuric silica gel columns, enhanced matrix removal (EMR, Agilent, USA) dispersive solid phase extraction, and liquid-liquid extraction (LLE) with different organic solvent mixtures (isohexane mixed with toluene, ethyl acetate, chloroform, and pure isohexane at different volumes).

The final and optimized sample preparation method used for individual plasma sample analysis in **Paper II and III** started with loading plasma/serum samples into a glass tube (muffle furnace), followed by adding acetonitrile ($4 \times$ plasma volume, i.e., 400 μ L for 100 μ L in the VIP cohort, and 800 μ L for 200 μ L plasma in the S3WP cohort) spiked with 8 μ L of 100 ng/mL surrogate internal standards before protein precipitation, vortexing and centrifugation. The protein-free supernatant was transferred to a new tube for liquid-liquid extraction with 1.2 mL, and subsequently 0.6 mL, of isohexane. The two isohexane extracts were combined and evaporated by a gentle N_2 flow to 100 μ L, and then transferred to the injection vial spiked with 6 μ L of methoxychlor- d_{14} (100 ng/mL) as a volumetric internal standard. The serum (500 μ L, $n = 3$) in **Paper I** was prepared using the same method as above with proportional solvent volume adjustments, except that all 112 tested standards (90 native and 22 labeled chemicals, 6.7–33.6 ng/mL in serum) were spiked into the extract before the injection (details in **Paper I**).

3.3.2 Sample preparation for polar analytes (**Paper IV**)

The bottom acetonitrile-plasma layer after LLE described above was gently evaporated at room temperature under N_2 for 5 minutes to remove residual isohexane. Citric acid (10 μ L, 50%) and an additional 400 μ L of acetonitrile were added separately to reach the same volume (800 μ L) as the direct LC method of Sdougkou et al.⁴³ The mixture was vortexed and loaded onto Hybrid Solid phase extraction (SPE)-Phospholipid cartridges, sequentially preconditioned with 10 mL of acetonitrile (0.5% citric acid), 12 mL methanol, and finally 2 mL of acetonitrile (0.5% citric acid). Elution was performed with 1 mL acetonitrile (0.5% citric acid) and 2 mL of 1% NH_4COOH in methanol. The eluent was mixed with 40 μ L of 25% aqueous ammonia solution, vortexed, and centrifuged (5000 rpm, room temperature, 12 min). The supernatant was transferred to an Eppendorf tube, evaporated to under 1 mL by N_2 , vortexed, then further evaporated to 100 μ L. After sonication, the extract was

transferred to a filtration tube. The evaporation tube was rinsed with 100 μL methanol, vortexed, and added to the same filtration tube. After centrifugation (10000 rpm, room temperature, 10 min), the final extract was transferred to an injection vial spiked with 8 μL diuron- d_6 (105 ng/mL).

3.4 Instrumental analysis

3.4.1 GC-HRMS analysis

In **Paper II**, instrumental method development encompassed comparing 15 m and 30 m DB-5MS columns by evaluating the instrumental detection limits (IDLs) for 69 targets from 4 chemical classes at 1 μL injection, along with evaluation of different injection volumes (1, 5, 10, 20, 25, 30 μL) on the 30 m column. Method reproducibility was assessed by 60 continuous large-volume injections of plasma extract. The final nonpolar plasma extracts, in **Paper II-III**, were analyzed using a GC-HRMS system (TRACE 1300 series, Q-Exactive, ThermoFisher Scientific, US) operating in EI full scan mode (34–750 m/z). Ion transfer line and source temperatures were 300 $^{\circ}\text{C}$. Chromatographic separation employed a DB-5MS column (30 m \times 0.25 mm \times 0.25 μm , Agilent) with temperature gradient program starting at 30 $^{\circ}\text{C}$ for 1 min, increasing to 50 $^{\circ}\text{C}$ at 20 $^{\circ}\text{C}/\text{min}$, then ramping to 170 $^{\circ}\text{C}$ at 25 $^{\circ}\text{C}/\text{min}$, to 250 $^{\circ}\text{C}$ at 6 $^{\circ}\text{C}/\text{min}$, and then increasing to 315 $^{\circ}\text{C}/\text{min}$ at 25 $^{\circ}\text{C}/\text{min}$ with a 12 min hold. The carrier gas was He at 1.3 mL/min. Injections of 25 μL extracts were made to a programmable temperature vaporizer (PTV) injector with baffle liner in large-volume mode. The serum extract in **Paper I** was analyzed in the same system as above, but with 1 μL injections prior to the implementation of the optimized large volume injection method.

3.4.2 LC-HRMS analysis

In **Paper IV**, polar plasma extracts were analyzed using an ultrahigh pressure LC-HRMS (Ultimate 3000, Orbitrap HF-X, Thermo Scientific, US) with the method as described in our previous study.⁴³ Briefly, the data was acquired

with parallel full scan (MS1, 90–1000 m/z) and data-independent MS/MS acquisition (DIA) with four m/z precursor windows. The extracts (20 μL) were injected on an Acquity BEH C18 column (130 Å, 1.7 μm , 3 \times 100 mm, Waters) with vanguard pre-column, employing a binary gradient elution (0.4 mL/min) with mobile phases of (A) water containing 1 mM ammonium fluoride and (B) methanol.

3.5 Data analysis and visualization

3.5.1 Multi-class target data analysis

In **Papers II** and **III**, target analysis to quantify the 103 target analytes was performed using TraceFinder (v.5.0, Thermo Scientific) with solvent-based external calibration curves (7 points, 0.005–5 ng/mL, $n = 3$). Isotopically labeled surrogate internal standards ($n = 26$) were used to correct for recovery and other variations in sample preparation. Quantification of polar target analytes in **Paper IV** was based on peak areas exported from MS-DIAL, and solvent-based external calibration curves with internal standards (9 points, 0.005–50 ng/mL, $n = 3$) were used. Isotopically labeled surrogate internal standards ($n = 34$) were used to correct the recoveries and variations during sample preparation.

3.5.2 Nontarget data analysis

In **Paper I**, raw HRMS files were processed (by co-author collaborators) in Compound Discoverer (v. 3.2, Thermo Scientific) for peak picking, spectral deconvolution, alignment and integration of peak areas. The exported data were further processed for spectral library matching, scoring and filtering by SIF-GC software. In **Papers II**, **III** and **IV**, nontarget analysis, including peak picking, spectral deconvolution, spectral matching, alignment and peak integration, was conducted in MS-DIAL.⁴⁶ Libraries for spectral matching included MassBankEU,⁴⁷ Global natural product social molecular networking (GNPS),⁴⁸ NIST, and an in-house HRMS library.

3.5.3 Visualization and statistics

Further data processing, visualization and statistical analyses employed Excel (Microsoft office 2019), Python (v.3.7.3)⁴⁹ and Jupyter Notebook (v.5.7.8),⁵⁰ and R (v.4.3.2)⁵¹ and RStudio (v.2023.12.1+402)⁵² with various packages, including plotly (v.5.24.1), pingouin (v. 0.5.5), ggplot2 (v.3.4.4), and scipy.stats (1.13.1). PCA analysis was performed in SIMCA (v. 17.0, Umetrics).

4. Results and discussion

4.1 Annotation framework for GC-HRMS nontarget analysis in **Paper I**

GC-HRMS instrumental methods provide data-rich information with some unique advantages and limitations for compound identification and feature annotation. These include retention time and RI, a rather universal but hard ionization process, EI, which often results in fragmentation of the molecular ion, as well as high spectral resolution, with benefits to accurate mass and accurate isotopic patterns.

Despite availability of softer and more selective ionization modes for GC-MS (e.g., chemical ionization), the most universal and commonly used ionization technique has long been EI (normally at 70 eV). EI spectra are dominant in spectral libraries, and for nontarget analysis serve as the most important structural evidence for molecular annotation. However, as a hard ionization process (the molecular ion is an unstable cationic radical), EI results in fragmentation of the molecular ion, which can reduce confidence for annotation due to low abundance or absence of the molecular ion in the acquired EI spectrum.

Nevertheless, GC also has unique advantages that can partly compensate for the limitations of molecular ion fragmentation, specifically the higher separation efficiency (relative to LC) and relatively consistent RI, primarily measured using the Kovats method, which can provide additional annotation confidence, including for closely related isomers. In fact, RI is commonly stored in GC-MS spectral libraries, thus providing a secondary layer of diagnostic information.^{22,23,53} With the combination high spectral resolution and accurate mass (low- to sub-ppm mass accuracy on Orbitrap instruments),⁵⁴ the molecular formula (of fragments or molecular ion) can be predicted for a third

layer of evidence supporting annotation. Building on the Schymanski confidence criteria for LC-HRMS, and incorporating GC-specific analytical evidence, an actionable scoring framework for GC-HRMS annotation communication was proposed (also Figure 3):

Level 1: Confirmed identification (retention time, EI spectrum, and reference ions) using in-house library. Retention time matches ($RSD < 1\%$) standard database generated in-house by the same method. Observation of a minimum 2 EI spectral reference peaks of the standard at the correct ratios ($< 20\%$), or EI spectral match with in-house library > 600 . If present, an accurate mass agreement (< 5 ppm error) for the molecular ion.

Level 2: Probable structure or close isomer using external libraries (RI match, molecular ion and EI match to exact mass library or including metrics incorporating accurate mass information). RI match (< 50 and $< 1.5\%$), or predicted RI (< 100 difference) using a validated method. Reverse dot-product EI spectral match score (> 600) and dot-product score (> 500). Match to an exact mass library (reverse dot-product score > 600), or using metrics incorporating exact mass (e.g., high reverse high resolution mass filter score (HRMF)⁵⁵ > 75).

Level 3: Tentative candidate (EI accurate mass spectral match, or EI match with metrics incorporating accurate mass) using an external library; alternatively, RI match with accurate mass fragment matches. Reverse dot-product EI spectral match (> 600) and dot-product (> 500). Match to an exact mass library (reverse dot-product > 600) or using metrics incorporating exact mass (e.g. HRMF > 75). Alternatively, RI match within 100 (predicted or experimental) and EI spectral match based on in silico accurate mass spectra, rule-based ion presence or absence (at least three ions, accurate mass), or detection of a molecular ion (accurate mass).

Level 4: Chemical group or exact chemical formula. Level 4A: Identification of unequivocal chemical formula. Only one formula is possible given by

the exact mass of the precursor ion and isotopic distribution directly or after only including atoms. Level 4B: Possible chemical series (e.g. homologous series with repeating chemical constituents). Follows mass defect series (e.g. using Kendrick mass defect). Or shows defined subnetwork clustering. Level 4C: Possible chemical class (chemicals grouped based on structural motifs or similarity) with one or more fragments.

Level 5: Unknown feature (retention time and reference mass or deconvoluted spectrum).

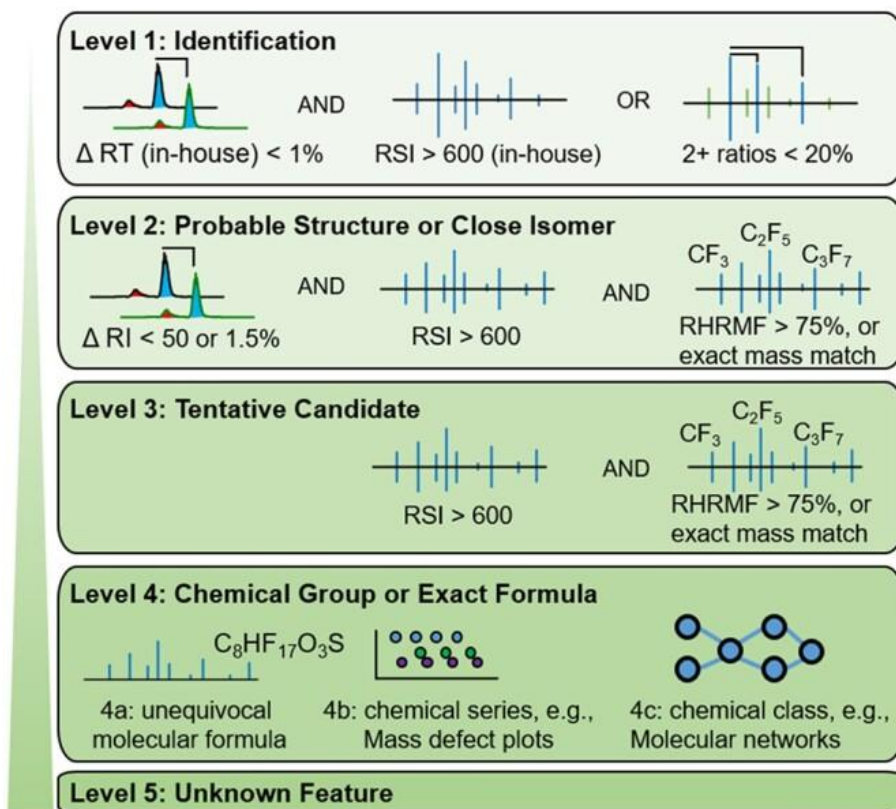


Figure 3. Proposed schema for assigning five levels of confidence in compound annotation using common evidence in GC-HRMS analysis. Adapted from Paper I. Acronyms: retention time (RT), reverse search index (RSI), RI match (ΔRI), and RHRMF. From Paper I.

GC-HRMS data from spiked human serum and air samples were used to validate the annotation score framework, and a supporting open science software (SIF-GC). For the serum extract, 80 of the 90 spiked native standards were detected and spectrally deconvoluted by the nontarget data processing (Compound Discoverer, Thermo Fisher Scientific), of which 61 were assigned as Level-2 annotation, but the correct exact structure was only identified for 28 compounds. These data emphasize the remaining challenges in nontarget analysis, not only of false-negatives (e.g., due to detection limits, poor chromatography or interferences), but also the overall false positive rate (54% for Level-2 assignments, and 60% for Level-3 assignments). However, when excluding cases of subtle isomeric differences (e.g. position of methyl or chlorine substitution), the false-positive rate was only 13% for the Level-2 assignments, and 25% for Level-3 assignments. This framework establishes a systematic approach to assign and communicate confidence of feature annotations by integrating quantitative scoring metrics alongside qualitative methods for ranking candidates to yield a confidence assignment.

4.2 Method development and validation for GC-HRMS analysis in **Paper II**

GC column selection. Among the 69 target analytes tested on 15 m and 30 m columns, two analytes (pentachlorobenzene and PCB-3) were excluded from the comparison because of very low IDLs (i.e., $\ll 2.5$ fg) that were below the tested concentration range. Sensitivity comparisons revealed superior performance for the 30 m column, on which 51.5% of analytes (35/68) exhibited 2–5 times lower IDLs, 22 analytes had equivalent IDLs on the two columns, and only 10 analytes (primarily high-boiling-point PBDEs) had better sensitivity on the shorter 15 m column. Therefore, the 30 m column was selected for its enhanced sensitivity for most priority analytes, and thus for its presumed broader utility in nontarget molecular discovery.

Performance of large volume injection (LVI). LVI in the range of 1–30 μL was tested to maximize method sensitivity, and 25 μL demonstrated an optimal balance between signal response and reproducibility (**Paper II** Figure S6). The robustness of the LVI-GC-HRMS instrumental method was later confirmed via 60 continuous injections of spiked human serum extracts over 4 days, which resulted in a median RSD of only 5.1% among the target analytes (range 1.6–24.5%, **Paper II** Figure S7).

Sample preparation for nonpolar analytes in human plasma. For development of the sample preparation method, 103 target analytes from 6 contaminant classes were evaluated. Initial testing of a traditional lipid removal technique, using acidic silica gel columns, was abandoned because PAHs and phthalates were degraded and had very low absolute recoveries (<5%, details in **Paper II** Figure S2). Consequently, acidic conditions were also avoided in subsequent method development, which ruled out certain plasma protein precipitation methods (e.g. using trichloroacetic acid, or phosphoric acid).⁵⁶ It was therefore an ideal starting point to use acetonitrile for the protein precipitation step, a common technique in metabolomics and consistent with our LC-HRMS chemical exposomics workflow.⁴³

Following protein precipitation with acetonitrile, different LLE systems were tested to maximize analyte recoveries into nonpolar solvent while minimizing co-extracted lipids. Pure isohexane achieved optimal recoveries, low chromatographic interference from any co-extracted lipids, and was practical due to the immiscibility of the top nonpolar layer (**H**exane) and bottom polar layer (**A**cetonitrile-**P**lasma). The procedure was generally termed the HA-P method, and despite further testing with additions of polar solvents (i.e., ethyl acetate, chloroform, and toluene) could not be further improved, as any polar solvent addition increased miscibility with the A-P subnanant, which lowered recoveries of polar analytes, and increased co-extraction of lipids. The added

benefits of EMR dispersive lipid removal material was also tested, but provided no additional lipid removal because the HA-P extract was already clean of most lipids, and lowered recoveries for hydrophobic target analytes.

Eventually, the HA-P method was optimized only by using a sequential isohexane extraction (i.e., 1.2 mL followed by 0.6 mL) after acetonitrile precipitation, achieving quantitative recoveries for most multi-class target analytes. Compared to a literature chemical exposomics method (XLE),²⁴ the HA-P produced visibly clearer extracts (transparent vs. lipid-rich yellow) and cleaner total ion chromatograms (TICs), with fewer lipid/fatty acid interferences when applied to pooled Swedish plasma (200 μ L, in triplicate, **Paper II** Figure 2). The clean extract by HA-P enabled effective LVI of 25 μ L of the extract, and consistent detection of targets, whereas the literature method, with only 2 μ L injections (and major matrix interference) resulted in fewer target analyte detections (**Paper II** Figure 3).

Validation. Method validation for all target analytes using the optimized HA-P method with LVI (25 μ L) included assessment of matrix effects, internal standard corrected recoveries, calibration linearity, and method limits of quantification (MLOQs). The median MLOQ was 0.088 ng/ mL (range 0.005–4.8 ng/mL; Figure 4a), with 99 of 103 target analytes demonstrating excellent calibration curve linearity ($R^2 > 0.99$). Among the 91 detectable analytes (out of 103) at 1 ng/mL spiking level, the mean matrix effect was null (mean 100% response; range 76–140%) and the mean recovery was 104% (range 22–132%; Figure 4b). Notably, three phthalates, specifically dimethyl phthalate, diethyl phthalate, and diethoxy ethyl phthalate, showed suboptimal recoveries (22–41%), owing to their lower hydrophobicity ($\log K_{ow} = 1.6, 2.5,$ and 2.1) and their presumed preferential partitioning into the plasma-acetonitrile polar layer during extraction. All other analytes, which had $\log K_{ow} > 3$, were quantitatively recovered, demonstrating the broad suitability of the HA-

P method for hydrophobic compounds. While this specificity limits the universal applicability of HA-P, it underscores the necessity of complementary LC-HRMS workflows for comprehensive chemical exposomics,¹⁶ particularly for ionic and polar contaminants with log K_{ow} values below 3.

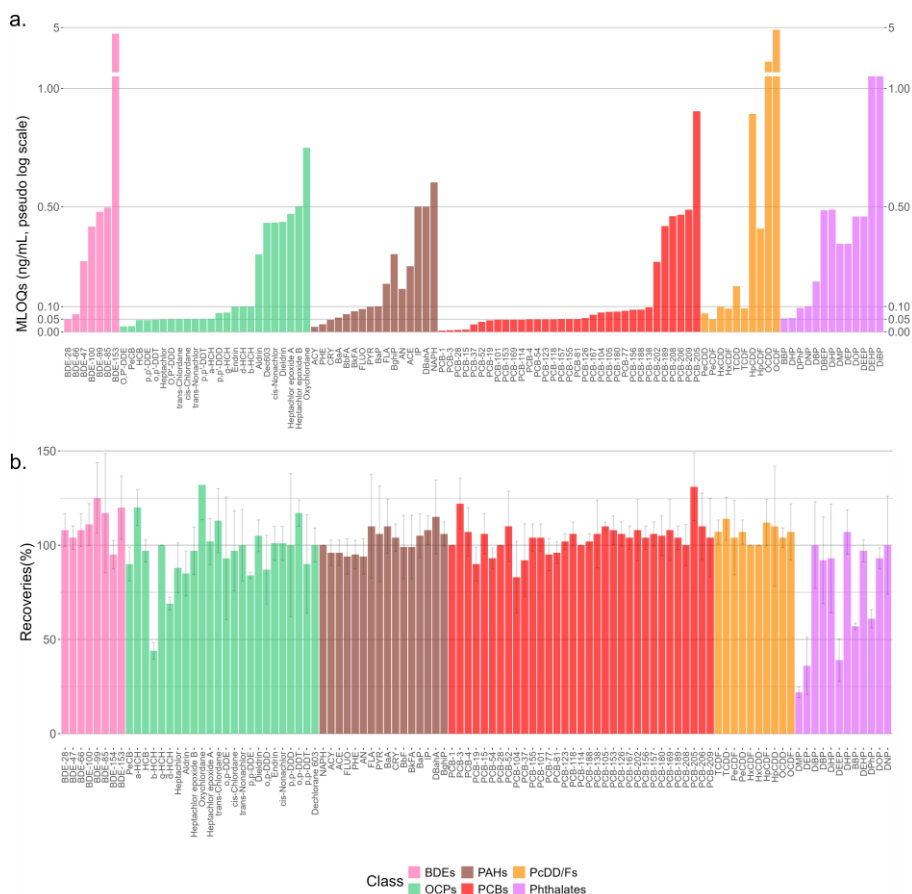


Figure 4. Method validation results for multi-class target analytes, arranged by classes, spiked to 200 μ L of human serum and analyzed by the HA-P method with LVI, showing (a) MLOQ ($n = 4$) and (b) internal standard corrected analyte recoveries ($n = 3$, 1 ng/mL spiking levels). Adapted from Paper II.

4.3 Multi-class target and nontarget analysis in 32 individual Swedish plasma samples in Paper II

The GC-HRMS target list containing 103 nonpolar analytes was applied for multi-class quantitative analysis in 32 individual plasma samples (100 μ L

each) from the VIP cohort. Out of 103 analytes, 51 were detected in at least one individual (details in **Paper II** Table S8), comprising seven dioxin-like PCBs (#105, 114, 118, 123, 156, 157 and 167), 14 non-dioxin-like PCBs (#1, 3, 4, 19, 15, 28, 52, 37, 101, 138, 153, 202, 180 and 205), 9 polycyclic aromatic hydrocarbons (PAHs), 12 organochlorine pesticide (OCPs), 1 polybrominated diphenyl ether (BDE), and 8 phthalates. Among these, 28 analytes had detection frequencies exceeding 20%, with gender-specific distribution shown in Figure 5.

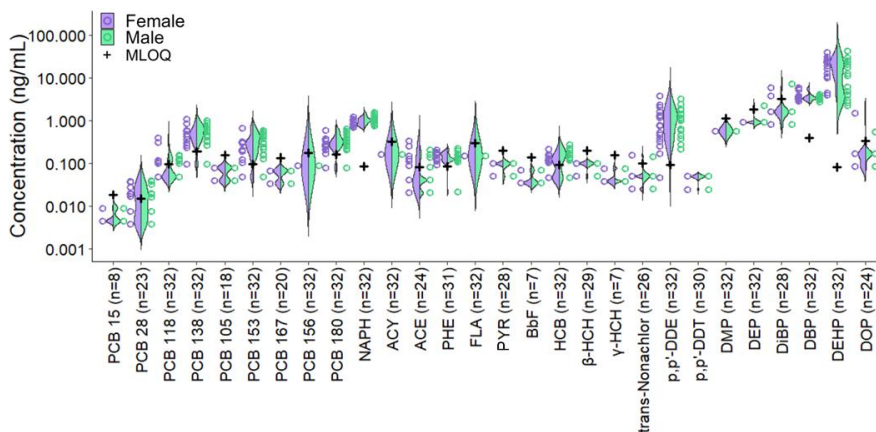


Figure 5. Violin plots showing concentration distributions for 28 target analytes (detection frequencies > 20%) in 32 individual Swedish adult plasma samples by sex (female in purple, male in green). Black '+' symbols indicate the target analyte MLOQ. For data visualization, detectable signals below the MLOQ are plotted as MLOQ/2, and non-detects are plotted as MLOQ/4. Adapted from Paper II, details in Paper II Table S7.

After processing the nontarget data with MS-DIAL and blank filtration, 875 molecular features were detected among the plasma samples (details in **Paper II** Table S9). Of these, 112 were annotated as Level 2 (Koelmel criteria, in **Paper I**), achieving a relatively high annotation rate of 12.8%. Among these, 7 were further confirmed at Level 1 using authentic standards (**Paper II**, Figures S12–18), including two related analytes 2,4-di-tert-butylphenol and tris(2,4-di-tert-butylphenyl) phosphite (used as antioxidants and UV stabilizers in rubber and plastics). The co-occurrence of the two antioxidants was

reported in indoor dust in 2018,⁵⁷ and 2,4-di-tert-butylphenol was detected from human urine and blood samples.^{58,59} But as far as we know, this was the first time that tris(2,4-di-tert-butylphenyl) phosphite detected in human bio-monitoring, likely due to the compound factor (high boiling point), higher sensitivity from LVI and lower matrix interference in the HA-P extracts.

4.4 Longitudinal GC-HRMS Exposomics in a Multiomic Wellness Cohort in **Paper III**

In this study, GC-HRMS-based quantitative analysis of 103 target analytes and nontarget analysis was performed in 276 plasma samples (50–200 μ L each) from 46 participants in the S3WP cohort, with six samples per participant over a 2-year period. A total of 52 target analytes were detected, including 1 BDE, 1 Dioxin, 12 OCPs, 14 PAHs, 14 PCBs and 10 phthalates; two additional PCB congeners were detected and confirmed at Level 1 (PCB-99 and PCB-146). Violin plots show the gender-specific distributions, with three PCBs (#156, 138, 180) having significantly higher concentrations in males. Age-dependent upward trends were observed for 4 PCBs (#146, 138, 153, 180) and hexachlorobenzene (HCB) (details in **Paper III** Figure 1-2).

For the nontarget analysis we compared MS-DIAL v4.9 (stable version but which was only able to process 256 of the sample files) and v5.4 (newer version that processed all 276 files). Processing 256 files in v4.9 yielded 1,510 features after blank subtraction, and 467 spectral library matches. In contrast, v5.4 yielded only 1,204 features after blank subtraction, and 242 spectral matches. The differences in annotations were likely because of the new added features: “Dot product score cut off” and “Reverse dot product score cut off”.

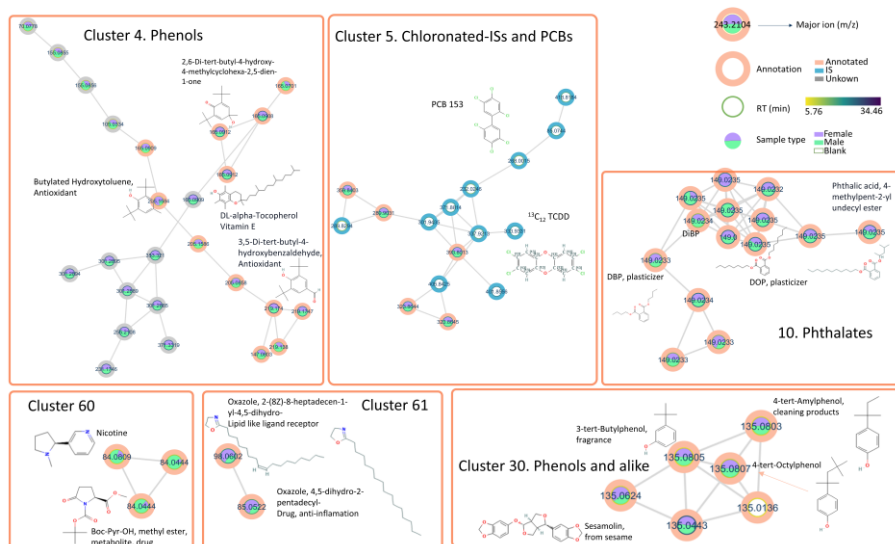


Figure 6 Example Clusters in Molecular Network of nontarget features detected among 276 plasma samples from 46 individuals, analyzed by the HA-P method and GC-HRMS. Each node presented one feature, labeled with the base-peak ion (m/z), the outer circle colors shows the annotation status (orange, blue and grey are annotated, internal standard, and unknown, respectively), the interior thinner circle color indicates retention time, and the inner pie chart shows the response ratio between the males (green), females (purple), and blank (white); empty pie charts indicate features only detected in blanks. All features were blank subtracted prior to networking. Adapted from **Paper III** Figure 3.

GNPS molecular networking was performed using the spectra of features exported from MS-DIAL v4.9 due to file format compatibility. The features (i.e., substances) were clustered by spectral similarity (examples in Figure 6, details in **Paper III** Figure 3). Among the 276 plasma samples, the molecular network identified 131 molecular family clusters derived from 726 features (out of 1,510). The two largest clusters—Cluster 1 with 89 features and Cluster 2 with 75—consisted primarily of endogenous compounds. Besides endogenous chemicals (e.g., sterols, hormones in Cluster 2, 11-13), various xenobiotics appeared, including aromatic compounds (Clusters 3, 6), phenolics (Cluster 4, 20, 30), PAHs (Cluster 16), and phosphates (Cluster 33). Each node (feature) in the network was also encoded with feature-specific metadata, including precursor ion mass-to-charge ratio (m/z), annotation confidence level, retention time, and gender-specific response ratio based on an average response among

the detectable samples. Notably, dimethyl aminoethyl esters (personal care products, Cluster 15) and oxazole antibiotics (Cluster 61) were more abundant in females, while in Cluster 60 nicotine and boc-L-pyroglutamic acid methyl ester (medical use) were higher in males.

Longitudinal Stability of the Chemical Exposome. Further data analysis focused on the 532 consensus molecular features detected in nontarget data-processing by both versions of MS-DIAL. This set of substances included 159 xenobiotics identified at Level 1 or structurally annotated at Level 2, 45 endogenous substances, and 328 unknown non-annotated features.

As introduced in our previous study on the polar chemical exposome, analyzed by LC-HRMS,⁴³ ICC is a nondimensional ratio comparing the interindividual variance to the total variance (i.e., the sum of inter- and intra-individual variance). As such, the ICC provides a relative measure of the longitudinal stability of a substance, with values ranging from 0 to 1. In the current study context, higher ICC values indicate greater consistency in the rank-ordering of individuals based on the substance's concentration across repeated sampling events. In this 2-year study, with six sampling points, ICC was calculated for each identified, annotated and unknown feature. For the chemical exposome subset of annotated substances, the mean and median ICC values were 0.24 and 0.17, respectively. These were similar to those observed for endogenous substances (mean and median ICCs of 0.27 and 0.24), any difference was not statistically significant ($p = 0.25$, Mann-Whitney test).

A key goal of chemical exposomics is to identify associations between levels of environmental substances measured in blood and biomolecular disruption, which may, for example, be measured in the same samples by metabolomics, proteomics, or transcriptomics.⁸ To that end, we compared the chemical exposome ICCs to previously reported ICCs—derived from the same individuals (i.e., same samples, different aliquots)—for the plasma proteome, lipidome, metabolome, and gut microbiome.³⁴ For improved accuracy, the

previously reported ICCs (originally based on 99 individuals) were recalculated using only the 46 individuals included in this chemical exposomics sub-study. The recalculated mean ICCs were 0.62 for the proteome, 0.41 for the lipidome, 0.51 for the metabolome, and 0.38 for the gut microbiota. These endogenous biomolecular mean ICCs were all significantly higher than for the nonpolar chemical exposome (mean 0.24) (Figure 7a), which indicated generally unique and unstable individual chemical exposome and requires more frequent repeat tests.

Longitudinal Exposure Types. Also as introduced in our previous study of the polar chemical exposome, analyzed by LC-HRMS,⁴³ this most detailed longitudinal exposomics study presents new opportunities for classifying exposures based on the combination of each substance's longitudinal stability (ICC) and frequency of detection (DF). A brief discussion of the four stability-frequency exposure types and their implications for exposome-studies are included below and illustrated in Figure 7b.

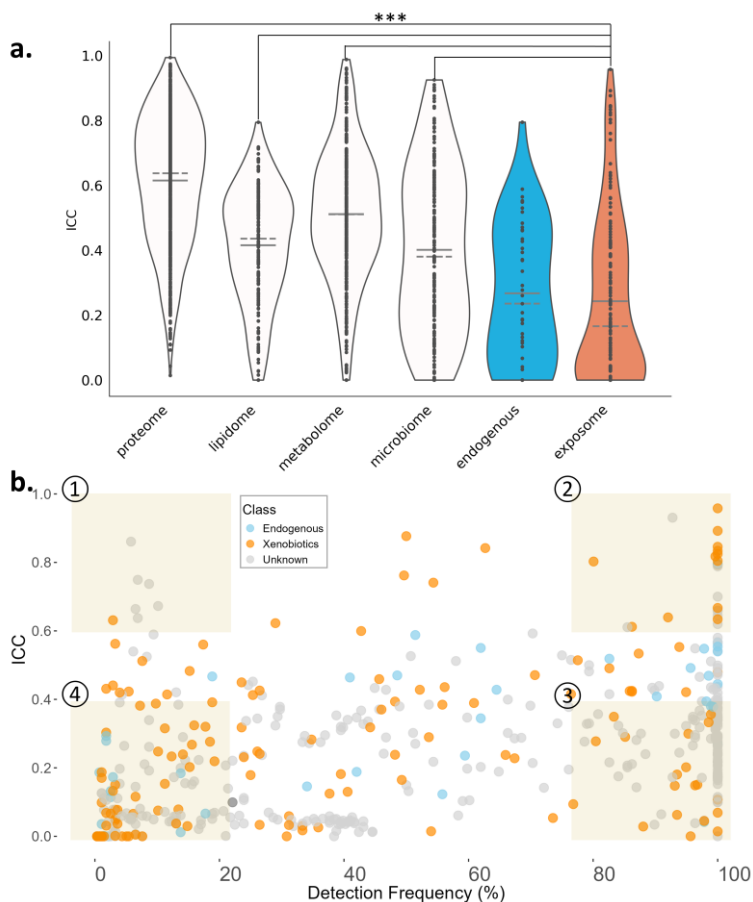
Type 1 (Rare-Stable): High stability (ICC > 0.6) but rare detection (DF < 20%). These exposures are rarely detected in the population, but are consistently present in some individuals, likely due to unique behaviours, occupations or environmental exposure sources. Although these exposures may not have great impacts on public health, for an individual's health the exposure may be consequential. Without a longitudinal sampling design, the importance of such exposures may go unnoticed. The only annotated substance in this category was o, p'-DDT (ICC = 0.63, DF = 2.9%), only detected in 8 samples from 5 individuals, a relatively minor component of the technical pesticide that was banned in Sweden in the 1970s. These 5 individuals also had relatively higher concentration in p, p'-DDT, but not its transforming product, p, p'-DDE. This probably indicated unique source of exposure or metabolism.

Type 2 (Common-Stable): High stability (ICC > 0.6) and frequent detection (DF > 80%). Substances in this category include persistent legacy pollutants

with long biological half-lives, and here included PCBs (#99, 118, 146, 138, 153, 156, 180), HCB, and p, p'-DDE. A more contemporary chemical in this category was 2,4-di-tert-butylphenol, linked to production of antioxidants and UV-stabilizers for plastics and rubber, and was also detected in the VIP cohort (**Paper II**). Many substances in this category have been studied by traditional environmental epidemiology and do not necessarily require longitudinal sampling for accurate exposure assessment. But highly detected phenolic antioxidants should be prioritized in future exposome and toxicity studies.

Type 3 (Common-Unstable): Low stability ($ICC < 0.4$) but frequent detection ($DF > 80\%$). Substances in this category may have broad impacts on public health, due to wide exposure, but will need to be monitored longitudinally for accurate exposure classification because of low ICCs. Annotated substances included 18 chemicals (**Paper III**, Table S5), notably including the organochlorine pesticide trans-nonachlor, and the ubiquitous antioxidant butylated hydroxytoluene.

Type 4 (Rare-Unstable): Low stability ($ICC < 0.4$) and infrequent detection ($DF < 20\%$). To understand the health impacts of substances in this category will require very large sample sizes (for statistical power) and a longitudinal design with multiple sampling events to minimize exposure misclassification. In small under-powered investigations, we have previously argued⁶⁰ to omit such substances from statistical analysis to minimize the penalty (i.e., Bonferroni correction)⁶¹ for multiple comparisons. Here, 40 annotated substances and 25 targets fell into this category, including the sporadically detected PCB-126, PCB-169, and a phthalate derivative, likely reflecting transient exposures to compounds with short biological half-lives. Since the quantified concentrations of these chemicals were close to their MLOQs, developing more sensitive analytical methods may help reduce the number of substances in this category and enhance exposome research.



*Figure 7. Longitudinal stability of the chemical exposome compared to other biomolecular profiles in the same study participants. (a) ICC data presented in violin plots separately for the chemical exposome and the endogenous metabolites as well as for the proteome, lipidome, metabolome, and microbiome. The solid and dashed lines show median and mean ICC values, respectively, and asterisks indicate significantly different ICC means between the chemical exposome, endogenous, all other molecular profiles ($p < 0.001$; Kruskal-Wallis test followed by Dunn' pairwise comparison test). (b) Intraclass correlation coefficients (ICC) of the 532 consensus substances (Level 1, Level 2 and unknowns) versus their detection frequencies (DF), for the chemical exposome (orange, $n = 159$), endogenous metabolites (blue, $n = 45$) and unknown features (gray, $n = 328$). Four regions of interest are highlighted on the plot: top left (Type 1; rare-stable, low DF, high ICC), top right (Type 2; common-stable, high DF, high ICC), bottom right (Type 3; common-unstable, high DF, low ICC), bottom left (Type 4; rare-unstable, low DF, low ICC). Adapted from **Paper III** Figure 4.*

To understand the connection between hundreds or thousands of environmental exposures and health, the high-dimensional nature of the datasets may

first need to be reduced. One approach may be through Hierarchical Clustering Analysis (HCA), which groups correlated exposures, and the resulting heat maps are also an effective means to visualize individual exposomes, including to identify unique or common co-exposures in the study population.⁴³ For the HCA here in the longitudinal S3WP study, the response of each substance or feature ($n = 532$) were averaged for the 6 visits of each individual over 2 years (Figure 8a). Various chemical co-exposures are highlighted in the HCA heatmap, and were generally classified here as ‘common’ or ‘rare’ based on the number of individuals with the co-exposure (Figure 8a). These are discussed in details in **Paper III**, but here only illustrative examples of environmental co-exposures are mentioned. Group I-L (Figure 8a) are notable correlated endogenous substances (i.e., metabolome) but will not be discussed further. Zoomed-in sections of the HCA show correlations for nine stable POP targeted analytes (Level 1, Figure 8b), three unstable phenolic antioxidants (Level 2, Figure 8c) and other substances.

Co-exposure Groups A-C (Figure 8a) included common exposures present in most or all people, and included substances used in daily-use products such as in plastics, cosmetics, food or cleaning products. As an example, in Group A, diphenyl ether (ICC = 0.07, DF = 100%) and 1,6-dimethyl naphthalene (ICC = 0.53, DF = 87.3%) grouped together. These have previously been reported to co-occur in recycled plastic materials, breast milk, and treated sewage sludge,^{62–64} but this is the first report of their correlated co-exposure in human plasma. Another interesting example was the correlated co-exposure of phenolic antioxidants (12 were detected overall), with 4 clustering with the two above substances in Group A. It is uncertain why these Group A substances correlated, but one consideration is that these come from liquid soaps or personal care products (including bug repellents) contained in plastic bottles that are then applied to the skin and absorbed into circulation.

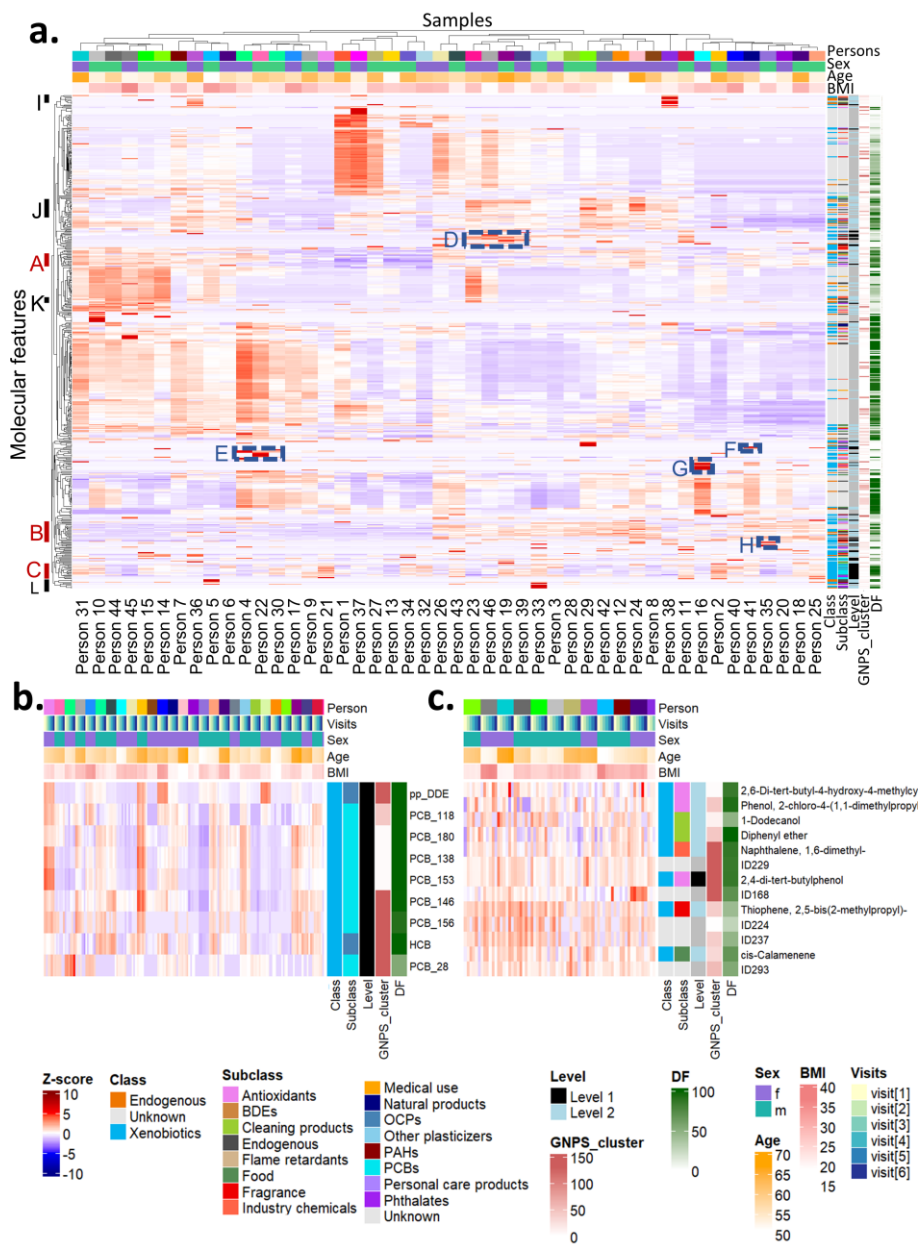


Figure 8. Clustered exposomes and correlated exposures in S3WP participants. A hierarchical cluster analysis heatmap with dendrograms displays exposome profiles of 46 individuals, (a) averaged across six clinical visits for 532 features (159 exposures, 45 endogenous, 328 unknowns). Relative exposure is shown as a Z-score, with more intense exposures in red, and lower exposures in blue according to the scale shown. The metadata of molecular features (or substances) are color-coded by class, subclass, identification confidence, GNPS cluster number, and detection frequency (DF), while individuals are color-coded by sex, age, and BMI. Groups of interest are high-

*lighted: common co-exposures (red, A–C), rare co-exposures (blue, D–H), and endogenous metabolites (black, I–L). Panel (b) and (c) show zoomed-in subsets of the HCA heatmap, but expanded to show 6 visits per individual and separate stable and unstable exposure groups across visits. Adapted from **Paper III** Figure 5 and S6.*

Highlighted co-exposure Groups D-H (Figure 8a) were relatively rare exposure clusters, potentially reflecting unique behaviours or lifestyle factors. As an example, 4 individuals are highlighted in Group D which was a correlated cluster of the heavy octachlorinated PCB congener (# 205) with 6 relatively large PAHs (1 three- ring: phenanthrene, 3 four-ring: benz[a]anthracene, chrysene, fluorene, and 2 five-ring: benzo[b]fluoranthene, benzo[k]fluoranthene), and the fatty aldehyde (4-nonenal, (4E)-) which could be endogenous or from its additive use in flavoring. The co-occurrence of all these, many of which have high K_{ow} , may relate to a dietary intake of fatty fish, meats or dairy items.

The HCA heatmap revealed that features with similar structural functional groups exhibited correlated exposure patterns, aligning with information from GNPS molecular networking (**Paper IV** Figure S9). For example, Group A and C in Figure 8 included many nodes that were clustered in Cluster 4 and 5 respectively in molecular networking. This methodological synergy can be a tool to simplify interpretation of the chemical exposome and to provided insights on the source and structure of unknown features.

4.5 Sample preparation method for dual analysis by LC and GC-HRMS in **Paper IV**

The unique phase separation observed in the HA-P method, yielding a non-polar isohexane layer (upper) and a polar acetonitrile-plasma layer (lower), enabled the development of a combined workflow for dual LC- and GC-HRMS analysis of the same plasma sample. This integrated method merges the HA-P method⁴² with our previously reported LC method,⁴³ requiring only an adjustment to the citric acid addition step (Figure 9). The integrated workflow preserves the HA-P workflow, with the top nonpolar layer (isohexane)

prepared for GC analysis, but the bottom polar layer (acetonitrile-plasma) is not disposed, rather it is further treated for phospholipid removal and an ensuing LC analysis. Here, the combined method is termed as HA-P GC/LC.

For GC analysis, the workflow remains exactly as in the HA-P method (**Paper II**). However, the further processing of the polar layer (subnatant) was slightly different than in Sdougkou et al.,⁴³ as here the protein precipitation was achieved with pure acetonitrile (without citric acid), followed by LLE, after which the addition of citric acid is made prior to SPE for phospholipid removal. Thus, to validate this integrated method, we only needed to evaluate the recoveries and matrix effects for the polar analytes by LC-HRMS analysis, and these were contrasted with values already published for the direct LC method of Sdougkou et al.⁴³

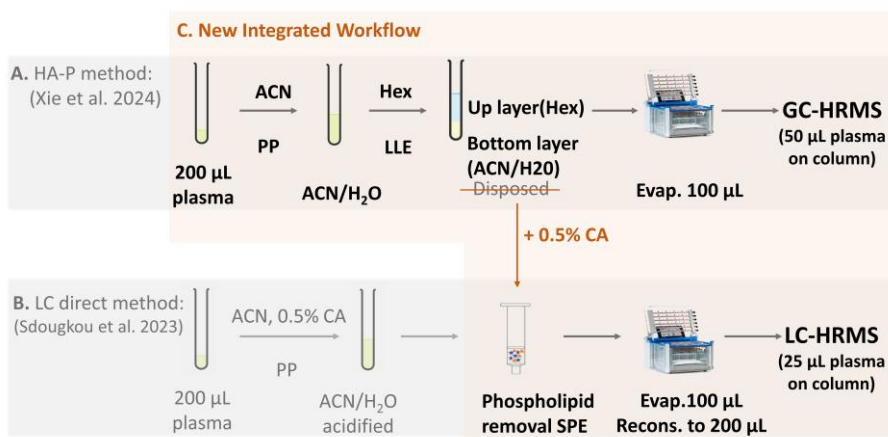


Figure 9 Schematic of sample preparation methods for polar and nonpolar chemical exposomics by existing LC- and GC-HRMS protocols (A, B), and showing the integrated workflow that enables both LC and GC chemical exposomics from a single plasma sample (C). From **Paper IV** Figure 1.

Recoveries. Among 76 target analytes, mean absolute recoveries showed no significant difference ($p = 0.26$) between the combined method and direct LC method (82% and 86%, respectively) in a two-tailed paired t-Test. While absolute recoveries (without internal standard correction, Figure 9) of most (56

out of 76) LC-analytes were actually slightly higher by the combined workflow, perhaps due to less matrix interference in the acetonitrile-plasma, 10 analytes had lower recoveries. These included four relatively hydrophobic chemicals ($\log K_{ow} > 3.79$) that likely partitioned into the isohexane phase, while the others (two polar analytes, four phosphates or phosphorothioates) were likely retained on the SPE cartridge, thus their recovery might be optimized in future work. Nevertheless, these relatively low recoveries can be largely corrected by the corresponding internal standards; median corrected recovery for these was 90%, range 68–136% (**Paper IV** Table S2-3).

Matrix effects. The mean absolute matrix effects differed significantly between the methods (107% for direct LC vs. 88% for HA-P GC/LC), though both methods achieved near-neutral suppression/enhancement (i.e., 100% indicates no matrix effect). For certain chemicals in the integrated workflow (details in **Paper IV** Figure 2), the matrix effect was lower (closer to 100%), and likely attributable to the isohexane-mediated removal of interferences prior to LC-HRMS.

Application to the VIP cohort. The optimized HA-P GC/LC method was applied to analyse the leftover plasma/acetonitrile supernatant layer from 32 individual (100 μ L each) plasma samples from the VIP cohort, that were previously analyzed by the HA-P workflow and GC-HRMS analysis in **Paper II**. Among the 76 priority target analytes by LC-HRMS, 46 were detected in at least one sample and quantified (**Paper IV** Table S4). These 46 target analytes belonged to 16 chemical classes and included 16 perfluoroalkyl and polyfluoroalkyl substances (PFAS), 7 herbicides, 3 flame retardants, 3 personal care products, 3 phytoestrogens, 4 steroid hormones, and other miscellaneous substances. For the associated nontarget analysis, 57669 features were detected after blank filtration, with 226 annotated (Level 1 and 2) including the 46 original targets, 98 additional Level 1 confirmations (**Paper IV** Table S5), and 82 Level 2 annotations according to Schymanski et al.²⁰

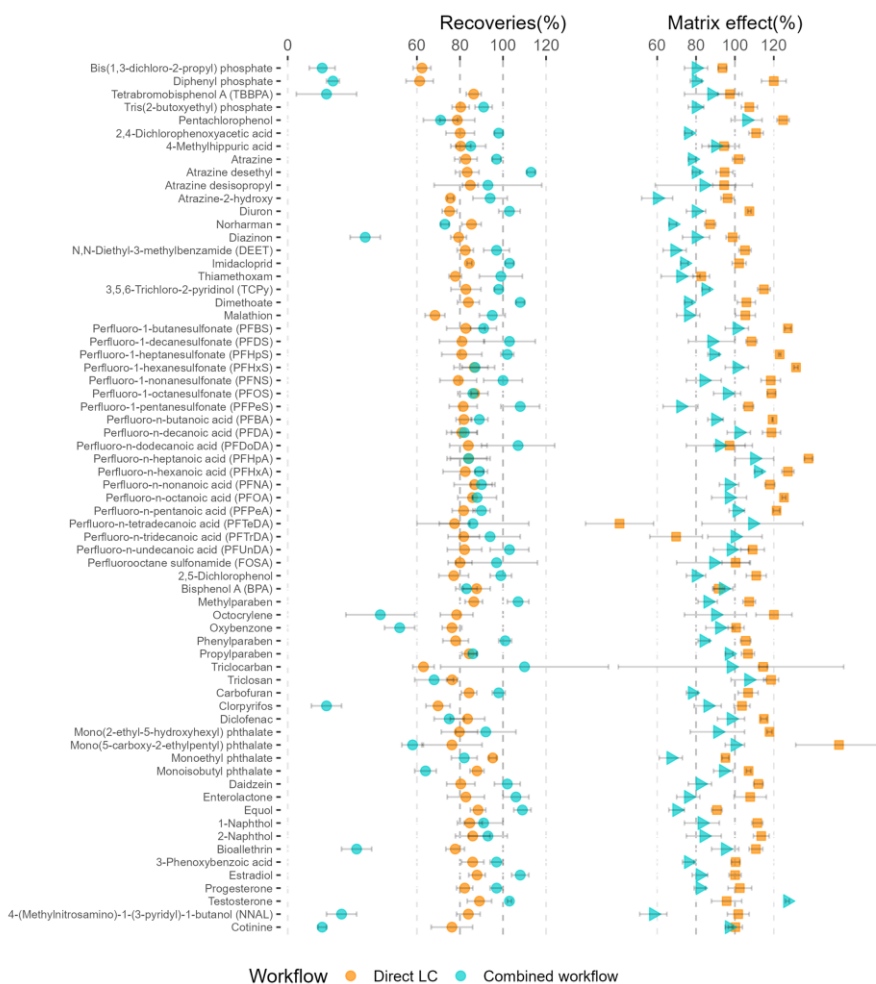


Figure 10 Absolute recoveries and matrix effects of polar analytes by the direct LC method (orange) and by the current integrated method (light blue). Error bars represent standard deviation, and compound order by RT from top to bottom. From **Paper IV** Figure 2.

In total, 204 detected chemicals were identified at Level 1 confidence through the integrated LC and GC multi-class target and nontarget analysis of 32 individual plasma samples. Of major significance, these substances span around 30 chemical classes, a broad range of molecular size, and over 10-orders of magnitude in physicochemical properties: molecular mass 128.1 to 714.0 Da, $\log K_{ow}$ -0.303 to 9.52, water solubility 1.73×10^{-5} to 1×10^6 mg/L. To cover this range of analytes by traditional targeted methods would have

been much more costly and required much larger sample volumes from each individual.

Here in **Paper IV**, HCA was again applied to 348 annotated features detected by the integrated LC- and GC- HRMS methods, and comprising both endogenous metabolites and environmental chemicals. The resulting heatmap (Figure 10) revealed that, while demographic factors such as age, gender, and meat consumption did not cluster strongly, distinct groups of co-exposures and metabolites could be visualized. Five clusters (Figure 9, Groups A–E, details in **Paper IV** Figure S2) represented common exposures with high detection frequencies in the population, reflecting widespread chemical exposures. Importantly, in many instances the LC- and GC-analytes clustered together, thereby demonstrating high-value in measuring such a wide chemical space by the integrated workflow.

Group A (7 LC-analytes) included steroid hormones and PFAS, which not only clustered together but also showed expected sex-related differences in plasma concentrations, supporting known biological and exposure pathways. Notably, PFHxS and PFHpS clustered with testosterone, as the same trend observed by others.^{60,65}

Other clusters, such as Group B (10 LC- and 3 GC-analytes), highlighted the co-occurrence of PFAS with endogenous and industrial compounds (e.g. chlorinated phenol (2-chloro-4-(1,1-dimethylpropyl) phenol, Level 2), insecticides (malaoxon), and two drugs or drug metabolites). Group C (1 LC- and 12 GC-analytes) was dominated by persistent organic pollutants (POPs; i.e., HCH, multiple PCBs (#167, 138, 153, 180, 105, 118), DDE, trans-nonachlor and HCB)), and lipophilic compounds (e.g. 1-heptadecene (Level 2) and 1-dodecanol (Level 2)), potentially due to a dietary exposure such as meat consumption. Groups D (6 LC- and 1 GC-analytes) and E (4 LC- and 1 GC-ana-

lytes) further illustrated mixtures of airborne semi-volatile pollutants, industrial chemicals, and persistent PFAS, underscoring the diversity and complexity of chemical exposures to which people are routinely coexposed.

In addition to these common co-exposure clusters in the population, the analysis identified several other Groups (F–J, Figure 9, details in **Paper IV** Figure S3) characterized by rare or individual-specific co-exposures, often present in only one or two participants. These rare clusters revealed highly personalized chemical profiles, such as unique combinations of pharmaceuticals, pesticides, industrial additives, and legacy pollutants, which may reflect specific occupational, environmental, or lifestyle factors. For instance, Group F (10 LC-analytes) was composed primarily of exogenous chemicals detected in individual 10, reflecting a distinctive exposure to an anthropogenic chemical mixture. It included a drug (LN80), the insecticide malathion, the insect repellent icaridin, the antifungal agent climbazole, and nootkatone; a natural compound used both as an insect repellent and flavoring agent. The co-exposure also features several industrial chemicals, such as 1,3-benzothiazole; 4,6-di-tert-butyl-cresol (BHT, an antioxidant additive, Level 2); tris(1-chloro-2-propyl) phosphate (a flame retardant, Level 2); and nitrophenols (2,4-dinitrophenol and 4-nitrophenol). Together, these compounds suggest overlapping environmental or occupational sources, potentially coupled with individual-specific behaviors, product use, or medical treatments. This distinctive mixture underscores a complex exposure scenario unique to individual 10, and highlights the capacity of comprehensive nontarget chemical exposomics to uncover insights of relevance to precision health.

Group G (4 LC- and 15 GC-analytes) highlighted rare hydrocarbon and flame retardant exposures in a few participants. Group H (4 LC- and 9 GC-analytes) consisted entirely of xenobiotic compounds, including banned or regulated persistent organic pollutants and current-use pesticides, again pointing to unique or rare exposure events.

The HCA also identified four subcluster Groups (K–N) of endogenous metabolites, some of which correlated with nearby exposure markers, possibly indicating a metabolic responses to environmental chemicals that could be explored in larger studies.

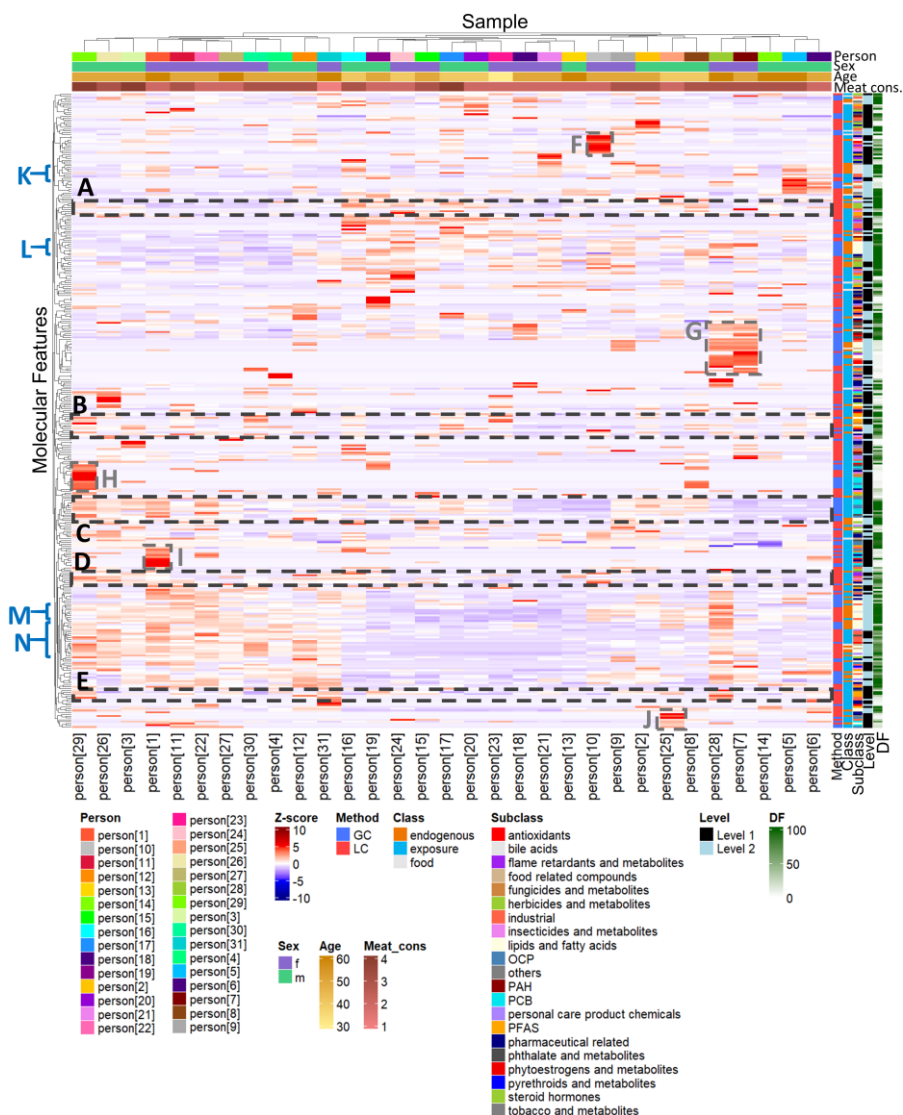


Figure 11. Hierarchical clustering of exposome profiles from 32 VIP study participants based on 348 annotated features (Level 1 and 2, including 254 environmental exposures, 81 endogenous metabolites and 13 food-related compounds, measured by GC- and LC-HRMS). Relative exposure is shown as a Z-score, with more intense exposures in red, and lower exposures in blue according to the scale shown. Chemical

*classes are color coded by compound class, subclass, identification confidence level, analytical method, and detection frequency (DF), while individuals are color-coded by sex, age, and meat consumption. Groups of interest are highlighted on the heatmap: A–E (black) denote common co-exposures; F–J (grey) highlight rare or individual-specific exposure patterns; and K–N (blue) represent clusters of correlated endogenous metabolites, see zoomed views in Figures S1–S5). Common co-exposures occurred in most participants, rare exposures appeared in isolated individuals or small subsets (e.g. in up to 4 or 5 participants). Adapted from **Paper IV** Figure 5.*

These findings demonstrate the power of the integrated GC- and LC-HRMS workflows to comprehensively profile the human exposome from single low-volume plasma samples, enabling both broad chemical coverage and the detection of individualized exposure patterns that may be relevant to future health and disease. This integrated approach advances exposomics research and lays a groundwork for future large-scale and multiomic studies aimed at understanding the interplay between environmental exposures, metabolism, biomolecular phenotypes and health outcomes.

5. Conclusions and perspectives

Building upon the widely applied Schymanski criteria for expressing confidence in LC-HRMS nontarget annotation, here we proposed an enhanced annotation framework for GC-HRMS that incorporates retention time index and other GC-MS specific criteria into the annotation framework to more clearly reflect confidence in molecular annotations by GC-HRMS-based workflows. This GC-optimized schema (the Koelmel criteria), offers a standardized reporting framework that will be familiar to users of the Schymanski criteria, but which will improve the quality and impact of GC-based studies in the environment and of the human exposome.

The HA-P method developed here for GC-HRMS-based chemical exposomics demonstrated high sensitivity which was achieved through an optimized extraction that minimizes lipids, thereby effectively enabling large-volume injection (25 μ L) for maximal method sensitivity. Its application to 32 individual plasma samples (100 μ L) confirmed its real-world capabilities for multi-class target quantification and simultaneous nontarget discovery and confirmation of novel substances. As important context, previously published GC-HRMS-based chemical exposomics methods failed to uncover or confirm any specific nontarget analytes, which we attribute to major lipid interferences that limit sensitivity on Orbitrap instrument. The quantitative strength of the HA-P method was furthermore evident for 4 PCBs (#118, 138, 153, 180) and 2 organochlorine pesticides (HCB and p,p'-DDE) which had statistically significant decreasing temporal trends in the small dataset.

Application of the HA-P method and GC-HRMS analysis in a longitudinal study of 46 individual healthy adults revealed greater instability (i.e., lower ICCs) for chemical exposures compared to profiles of biomolecules measured in the same people (proteomic, metabolomic, lipidomic, and microbiomic), indicating that multiple sampling events will be essential in future powerful studies of the human chemical exposome. Nevertheless, through application

of powerful tools, such as GNPS molecular networking and HCA, we demonstrated at the cohort scale how complex high-dimensional exposome datasets can be effectively explored and simplified for future exposome-wide association studies.

The integrated HA-P GC/LC method was quantitatively validated for dual LC/GC-HRMS chemical exposomics using only 1 low-volume plasma aliquot. This expands the chemical space coverage to include both polar and nonpolar substances from around 30 chemical classes, thereby allowing hundreds of substances to be detected that span over 10-orders of magnitude in physical properties. To cover this range of analytes by traditional targeted methods would have been much more costly and required much larger sample volumes from each individual. Even in the small sample set here, the unprecedented chemical coverage revealed correlations among GC- and LC-amenable compounds that revealed new insights into exposure sources, and which raise new questions for the associated health impacts of complex environmental exposures in the population, and indeed in individuals with unique or rare co-exposures. The HA-P GC/LC method establishes a foundational capability for future large-scale cohort studies, in particular for exposome studies that must preserve plasma aliquots for multiple multiomic analyses. Future research should consider scaling-up the HA-P GC/LC method for higher-throughput in much larger human exposome studies that are now being envisioned at the international scale.⁶⁶

Acknowledgements

First and importantly, I would like to thank my supervisor, Jon, for the opportunity of this position. I would not have achieved so much without your professional and rigorous scientific guidance and training. All this had a big impact on my research life, which has not been easy but very fruitful and meaningful. Meanwhile, I would like also thank my co-supervisors Lillemor and Oskar, who have been there to talk and offer support.

Thanks to all my collaborators from all around the world, it has been an honour for me to participate in all these amazing and interesting projects. And I want to show my gratitude to the group members for the fun time with critical discussion and supportive activities together, I would not have made it without you. Thanks to my academic twin sister, Kalli. It was a really nice time working and supporting each other all the way, and working out the projects. Thanks to my officemates: Bénilde for all the interesting serious scientific discussions :), your help with figure coding and time after work; May Britt for the fun, hard work and spare time together; the skiing trip was always impressive to me. I would like to thank Stefano for the nice collaborations and the help with data analysis. I still remember the first day when we started together, time flies and marks. Thanks Adrià, you are like a warm light coming through the window, and also Solveig for the discussions, and interesting talks and-conference trips. Thank you, Ioannis A. for the instrumental training, discussions and help. Thanks to my previous and new colleagues, Jean, Mai, Amaia, Yan and Daeho for being kind and fun. Thanks to Oskar's group at ScilifeLab, and to colleagues at C Corridor in Arrhenius for the interesting and supportive Fika and lunch times. I would like to thank Jon Benskin, Anna Sobek, Margareta, Karin Almerén, Karin Bergman, Zhe Li, Xiaodi Shi and other colleagues at ACES for being kind and supportive. I would like to thank my friends Juan, Adam, and Dora. I appreciate the time together, the things I learnt from you and your warm support.

感谢我在瑞典的中国朋友们，你们的存在让我感觉到家的温暖：英英、小云、婷茹、菲姐、婷姐、敏瑜、瑀焯、雪佩、丽涛、兔子、婕妤及她们的家人们，还有那些羽毛球小伙伴和其他未能提及的朋友们。在所有的历历过往中一起走过的欢乐，在我困难时给予的支持。我时常感慨自己是多么幸运能在瑞典遇到你们，有你们在身边。

谢谢我一生的战友翔，多远多难的路我们一起走过都成为风景。感谢我爸妈、亲爱的小妹谢红妍和其他家人们对我无限的爱和支持，一路磕磕绊绊，有你们做我强大的后盾我就有前行的勇气。感谢我两个亲爱的女儿，米娅和米粒，我这博士的一路有你们的陪伴，感谢这些年我们共同的成长，你们的到来给我们家庭增加了更多期待和幸福维度。

References

1. Wild, C. P. Complementing the genome with an ‘exposome’: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1847–1850 (2005).
2. Landrigan, P. J. *et al.* The Lancet Commission on pollution and health. *Lancet* **391**, 462–512 (2018).
3. Rappaport, S. M. & Smith, M. T. Environment and Disease Risks. *Science* **330**, 460–461 (2010).
4. Heinrich, U. Exhaust-Specific Carcinogenic Effects of Polycyclic Aromatic Hydrocarbons and Their Significance for the Estimation of the Exhaust Exposure-Related Lung Cancer Risk. in *Assessment of Inhalation Hazards* 301–313 (Springer, Berlin, Heidelberg, 1989). doi:10.1007/978-3-642-74606-2_28.
5. Grove, R. & Sanchez, O. Human Exposure to Environmental Pollutants and Associations with Non-Hodgkin Lymphoma: A Review. *Expo. Heal.* **14**, 99–109 (2022).
6. Donat-Vargas, C. *et al.* Persistent organochlorine pollutants in plasma, blood pressure, and hypertension in a longitudinal study. *Hypertension* **71**, 1258–1268 (2018).
7. Jones, D. P. Sequencing the exposome: A call to action. *Toxicol. Rep.* **3**, 29–45 (2016).
8. Stokstad, E. Why were salmon dying? The answer washed off the road: Common tire chemical implicated in coho salmon kills. *Science* **370**, 1145 (2020).
9. What Killed the Eagles? EPA Researchers Help Solve 25+ Year Mystery | US EPA. <https://www.epa.gov/sciencematters/what-killed-eagles-epa-researchers-help-solve-25-year-mystery>.
10. Rappaport, S. M., Barupal, D. K., Wishart, D., Vineis, P. & Scalbert, A. The blood exposome and its role in discovering causes of disease. *Environ. Health Perspect.* **122**, 769–774 (2014).
11. David, A. *et al.* Towards a comprehensive characterisation of the human internal chemical exposome: Challenges and perspectives. *Environ. Int.* **156**, 106630 (2021).
12. *National Health and Nutrition Examination Survey (NHANES 2013-2014)*. https://wwwn.cdc.gov/Nchs/Data/Nhanes/Public/2013/DataFiles/PCBPOL_H.htm.

13. Wang, Z., Walker, G. W., Muir, D. C. G. & Nagatani-Yoshida, K. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ. Sci. Technol.* **54**, 2575–2584 (2020).
14. González-Domínguez, R., Jáuregui, O., Queipo-Ortuño, M. I. & Andrés-Lacueva, C. Characterization of the Human Exposome by a Comprehensive and Quantitative Large-Scale Multianalyte Metabolomics Platform. *Anal. Chem.* **92**, 13767–13775 (2020).
15. Jamnik, T. *et al.* Next-generation biomonitoring of the early-life chemical exposome in neonatal and infant development. *Nat. Commun.* **13**, 2653 (2022).
16. Zhang, P. *et al.* Defining the Scope of Exposome Studies and Research Needs from a Multidisciplinary Perspective. *Environ. Sci. Technol. Lett.* **8**, 839–852 (2021).
17. Rostkowski, P. *et al.* The strength in numbers: comprehensive characterization of house dust using complementary mass spectrometric techniques. *Canada Anal. Bioanal. Chem.* **411**, 1957–1977 (2019).
18. Manz, K. E. *et al.* Non-targeted analysis (NTA) and suspect screening analysis (SSA): a review of examining the chemical exposome. *J. Expo. Sci. Environ. Epidemiol.* **33**, 524–536 (2023).
19. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).
20. Schymanski, E. L. *et al.* Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
21. Travis, S. C., Kordas, K., Aga, D. S. & Aga, D. S. Optimized workflow for unknown screening using gas chromatography high-resolution mass spectrometry expands identification of contaminants in silicone personal passive samplers. *Rapid Commun. Mass Spectrom.* **35**, e9048 (2021).
22. Kováts, E. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helv. Chim. Acta* **41**, 1915–1932 (1958).
23. The IUPAC Compendium of Chemical Terminology. *The IUPAC Compendium of Chemical Terminology* (2025) doi:10.1351/GOLDBOOK.
24. Hu, X. *et al.* A scalable workflow to characterize the human exposome. *Nat. Commun.* **12**, 5575 (2021).

25. Manz, K. E. *et al.* Targeted and Nontargeted Detection and Characterization of Trace Organic Chemicals in Human Serum and Plasma Using QuEChERS Extraction. *Toxicol. Sci.* **185**, 77–88 (2021).
26. Rylander, L., Dyremark, E., Strömberg, U., Östman, C. & Hagmar, L. The impact of age, lactation and dietary habits on PCB in plasma in Swedish women. *Sci. Total Environ.* **207**, 55–61 (1997).
27. Jönsson, B. A. G. *et al.* Inter-population variations in concentrations, determinants of and correlations between 2,2',4,4',5,5'-hexachlorobiphenyl (CB-153) and 1,1-dichloro-2,2-bis (p-chlorophenyl)-ethylene (p,p'-DDE): A cross-sectional study of 3161 men and women from Inuit and European populations. *Environ. Heal. A Glob. Access Sci. Source* **4**, 27 (2005).
28. Schaeffer, D. J., Dellinger, J. A., Needham, L. L. & Hansen, L. G. Serum PCB profiles in Native Americans from Wisconsin based on region, diet, age, and gender: Implications for epidemiology studies. *Sci. Total Environ.* **357**, 74–87 (2006).
29. Kraft, M. *et al.* Quantification of all 209 PCB congeners in blood—Can indicators be used to calculate the total PCB blood load? *Int. J. Hyg. Environ. Health* **220**, 201–208 (2017).
30. Metcalfe, C. D. *et al.* Methods for the analysis of endocrine disrupting chemicals in selected environmental matrixes. *Environ. Res.* **206**, 112616 (2022).
31. Gorshkov, M. V., Good, D. M., Lyutvinskiy, Y., Yang, H. & Zubarev, R. A. Calibration Function for the Orbitrap FTMS Accounting for the Space Charge Effect. *J. Am. Soc. Mass Spectrom.* **21**, 1846–1851 (2010).
32. Gao, P. *et al.* Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res.* **32**, 1199–1214 (2022).
33. Oosterwegel, M. J. *et al.* Variability of the Human Serum Metabolome over 3 Months in the EXPOsOMICS Personal Exposure Monitoring Study. *Environ. Sci. Technol.* **57**, 12752–12759 (2023).
34. O'gorman, A. *et al.* Identification of a plasma signature of psychotic disorder in children and adolescents from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. *Transl. Psychiatry* **7**, e1240 (2017).
35. Tamayo-Uria, I. *et al.* The early-life exposome: Description and patterns in six European countries. *Environ. Int.* **123**, 189–200 (2019).
36. Stableski, J., Lind, L., Salihovic, S., Lind, P. M. & Kärman, A. Longitudinal changes in persistent organic pollutants (POPs) from

- 2001 to 2009 in a sample of elderly Swedish men and women. *Environ. Res.* **165**, 193–200 (2018).
37. Cao, X. *et al.* Cross-sectional and longitudinal associations of dichlorodiphenyltrichloroethane (DDT) metabolites exposure with lung function alternation in the Chinese general adults. *Sci. Total Environ.* **905**, 167729 (2023).
 38. Svarcova, A. *et al.* Integration of five groups of POPs into one multi-analyte method for human blood serum analysis: An innovative approach within biomonitoring studies. *Sci. Total Environ.* **667**, 701–709 (2019).
 39. Foster, W. G. *et al.* Human maternal and umbilical cord blood concentrations of polybrominated diphenyl ethers. *Chemosphere* **84**, 1301–1309 (2011).
 40. Braun, G., Krauss, M. & Escher, B. I. Recovery of 400 Chemicals with Three Extraction Methods for Low Volumes of Human Plasma Quantified by Instrumental Analysis and In Vitro Bioassays. *Environ. Sci. Technol* **57**, 19363–19373 (2023).
 41. Braun, G. *et al.* Neurotoxic mixture effects of chemicals extracted from blood of pregnant women. *Science* **386**, 301–309 (2024).
 42. Xie, H. *et al.* Chemical Exposomics in Human Plasma by Lipid Removal and Large Volume Injection Gas Chromatography High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* **58**, 17592–17605 (2024).
 43. Sdougkou, K. *et al.* Phospholipid Removal for Enhanced Chemical Exposomics in Human Plasma. *Environ. Sci. Technol.* **57**, 10173–10184 (2023).
 44. Norberg, M., Wall, S., Boman, K. & Weinehall, L. The Västerbotten Intervention Programme: background, design and implications. *Glob. Health Action* **3**, 4643 (2010).
 45. Tebani, A. *et al.* Integration of molecular profiles in a longitudinal wellness profiling cohort. *Nat. Commun.* **11**, 4487 (2020).
 46. Takeda, H. *et al.* MS-DIAL 5 multimodal mass spectrometry data mining unveils lipidome complexities. *Nat. Commun.* **15**, 9903 (2024).
 47. MassBank | MassBank Europe Mass Spectral DataBase. <https://massbank.eu/MassBank/>.
 48. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
 49. Van Rossum, G. & Drake, F. L. Python 3 Reference Manual. *Nature*

- 585**, 357–362 (2009).
50. Kluyver, T. *et al.* *Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows*. (IOS Press, 2016). doi:10.3233/978-1-61499-649-1-87.
 51. R: The R Project for Statistical Computing. <https://www.r-project.org/>.
 52. Posit | The Open-Source Data Science Company. <https://posit.co/>.
 53. Qu, C., Schneider, B. I., Kearsley, A. J., Keyrouz, W. & Allison, T. C. Predicting Kováts Retention Indices Using Graph Neural Networks. *J. Chromatogr. A* **1646**, 462100 (2021).
 54. Domínguez, I., Arrebola, F. J., Martínez Vidal, J. L. & Garrido Frenich, A. Assessment of wastewater pollution by gas chromatography and high resolution Orbitrap mass spectrometry. *J. Chromatogr. A* **1619**, 460964 (2020).
 55. Kwiecien, N. W. *et al.* High-Resolution Filtering for Improved Small Molecule Identification via GC/MS. *Anal Chem* **87**, 8328–8335 (2015).
 56. Polson, C., Sarkar, P., Incledon, B., Raguvaran, V. & Grant, R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **785**, 263–275 (2003).
 57. Liu, R. & Mabury, S. A. Unexpectedly High Concentrations of a Newly Identified Organophosphate Ester, Tris(2,4-di-tert-butylphenyl) Phosphate, in Indoor Dust from Canada. *Environ. Sci. Technol.* **52**, 9677–9683 (2018).
 58. Arvstrand, L. Target analysis of synthetic phenolic antioxidants in human serum, prioritized by using an exposure index applied to the Swedish Products Register. (Stockholm University, 2021).
 59. Tang, S. *et al.* Prenatal Exposure to Emerging Plasticizers and Synthetic Antioxidants and Their Potency to Cross Human Placenta. *Environ. Sci. Technol* **56**, 8507–8517 (2022).
 60. Sdougkou, K. *et al.* Longitudinal Exposomics in a Multiomic Wellness Cohort Reveals Distinctive and Dynamic Environmental Chemical Mixtures in Blood. *Environ. Sci. Technol* **58**, 16302–16315 (2024).
 61. Van der Weele, T. J. & Mathur, M. B. SOME DESIRABLE PROPERTIES OF THE BONFERRONI CORRECTION: IS THE BONFERRONI CORRECTION REALLY SO BAD? *Am. J. Epidemiol.* **188**, 617 (2018).
 62. Sullivan, C., Mitchelmore, C. L., Hale, R. C. & Van Veld, P. A.

- Induction of CYP1A and DNA damage in the fathead minnow (*Pimephales promelas*) following exposure to biosolids. *Sci. Total Environ.* **384**, 221–228 (2007).
63. Zhang, Y. *et al.* Screening for compounds with bioaccumulation potential in breast milk using their retention behavior in two-dimensional gas chromatography. *Environ. Int.* **190**, 108911 (2024).
 64. Zhang, M. Y. *et al.* Identification and quantification of volatile organic compounds in Chinese recycled HDPE: Consequences for its applicability. *Food Packag. Shelf Life* **47**, 101424 (2025).
 65. Wang, Y. *et al.* Perfluoroalkyl substances and sex hormones in postmenopausal women: NHANES 2013–2016. *Environ. Int.* **149**, 106408 (2021).
 66. Vermeulen, R. *Human Exposome Research: Potential, Limitations and Public Policy Implications*. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2025\)765791](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2025)765791).