A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP11851.

A State-of-the-Science Review on High-Resolution Metabolomics Application in Air Pollution Health Research: Current Progress, Analytical Challenges, and Recommendations for Future Direction

Donghai Liang,[1] Zhenjiang Li,[1] Jelle Vlaanderen,[2] Ziyin Tang,[1] Dean P. Jones,[3] Roel Vermeulen,[2] and Jeremy A. Sarnat[1]

[1]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[2]Department Population Health Sciences, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands
[3]Department of Medicine, School of Medicine, Emory University, Atlanta, Georgia, USA

Background: Understanding the mechanistic basis of air pollution toxicity is dependent on accurately characterizing both exposure and biological responses. Untargeted metabolomics, an analysis of small-molecule metabolic phenotypes, may offer improved estimation of exposures and corresponding health responses to complex environmental mixtures such as air pollution. The field remains nascent, however, with questions concerning the coherence and generalizability of findings across studies, study designs and analytical platforms.

Objectives: We aimed to review the state of air pollution research from studies using untargeted high-resolution metabolomics (HRM), highlight the areas of concordance and dissimilarity in methodological approaches and reported findings, and discuss a path forward for future use of this analytical platform in air pollution research.

Methods: We conducted a state-of-the-science review to *a*) summarize recent research of air pollution studies using untargeted metabolomics and *b*) identify gaps in the peer-reviewed literature and opportunities for addressing these gaps in future designs. We screened articles published within Pubmed and Web of Science between 1 January 2005 and 31 March 2022. Two reviewers independently screened 2,065 abstracts, with discrepancies resolved by a third reviewer.

Results: We identified 47 articles that applied untargeted metabolomics on serum, plasma, whole blood, urine, saliva, or other biospecimens to investigate the impact of air pollution exposures on the human metabolome. Eight hundred sixteen unique features confirmed with level-1 or -2 evidence were reported to be associated with at least one or more air pollutants. Hypoxanthine, histidine, serine, aspartate, and glutamate were among the 35 metabolites consistently exhibiting associations with multiple air pollutants in at least 5 independent studies. Oxidative stress and inflammation-related pathways—including glycerophospholipid metabolism, pyrimidine metabolism, methionine and cysteine metabolism, tyrosine metabolism, and tryptophan metabolism—were the most commonly perturbed pathways reported in >70% of studies. More than 80% of the reported features were not chemically annotated, limiting the interpretability and generalizability of the findings.

Conclusions: Numerous investigations have demonstrated the feasibility of using untargeted metabolomics as a platform linking exposure to internal dose and biological response. Our review of the 47 existing untargeted HRM–air pollution studies points to an underlying coherence and consistency across a range of sample analytical quantitation methods, extraction algorithms, and statistical modeling approaches. Future directions should focus on validation of these findings via hypothesis-driven protocols and technical advances in metabolic annotation and quantification. https://doi.org/10.1289/EHP11851

Introduction

Air pollution constitutes a major environmental threat to human health globally. More than 90% of the world's population lives in places where air pollution levels exceed World Health Organization guidelines.[1] Although numerous adverse health effects associated with both short- and long-term exposures to air pollution have been well documented,[2–6] detailed molecular mechanisms underlying how air pollution exposures affect various biological pathways and systems remain largely unknown. Current findings point to oxidative stress, inflammation, and genotoxicity as central mechanisms by which air pollution exposures elicit adverse health effects.[7–9] Accordingly, air pollution epidemiology has historically focused on several targeted biomarkers related to these pathways and processes, including malondialdehyde, interleukins (IL)-1, IL-6, IL-8, IL-10, tumor necrosis factor-α, vascular cell adhesion molecule 1, and exhaled nitric oxide.[7,10] However, these biomarkers are not specific to air pollution exposure and the effects of air pollution on some of these biomarkers are still underexamined. Understanding the etiological role these biomarkers play in any observed downstream biological effects, along with the development of sensitive and specific biomarkers, is a critical next step.

In the exposome era, the cumulative measure of environmental influences and associated biological responses are monitored and evaluated throughout the life span thanks to breakthroughs in several high-throughput omics technologies.[11–14] The exposome concept has ushered in a period of improved interrogation of biological responses to air pollution, affording a better understanding of the underlying molecular mechanisms driving these responses. The use of untargeted omics analytical platforms, in particular, has been a key exposomics approach for estimating exposures to complex environmental mixtures, such as urban air pollution. High-resolution metabolomics (HRM), a high-throughput method involving the identification and quantitation of thousands of metabolic features associated with exogenous exposure and endogenous processes, has emerged as a powerful tool to improve exposure estimation to complex environmental mixtures.[12,15,16] Several pioneering studies have demonstrated the applicability of using untargeted HRM as a central platform linking exposure to internal dose and biological response where specific metabolites and metabolic pathways related to air pollution exposure were identified.[17–63] However, as noted in a recent preliminary scoping review examining a limited, initial set of metabolomic analyses and exposure to air pollution,[64] more large-scaled metabolomic studies with standardized protocols are needed to investigate the underlying mechanisms accounting for air pollution toxicity on human health. Indeed, heterogeneity

Address correspondence to Donghai Liang, Gangarosa Department of Environmental Health, Emory University Rollins School of Public Health, 1518 Clifton Rd., Atlanta, GA 30322 USA. Email: donghai.liang@emory.edu

Supplemental Material is available online (https://doi.org/10.1289/EHP11851).
The authors declare they have nothing to disclose.
Received 12 July 2022; Revised 22 March 2023; Accepted 30 March 2023; Published 16 May 2023.

among study cohorts, exposure assessment methods, biospecimen types, and metabolomics platforms have hindered comparison of these initial studies. Moreover, many of the initial air pollution–related metabolomics analyses have followed largely targeted or semi-targeted protocols, focusing on a relatively limited suite of validated and known metabolite features. To our knowledge, synthesizing findings from untargeted air pollution metabolomics, specifically, has not been previously conducted. These current knowledge gaps and the rapidly growing interest among many research groups in using these platforms to examine environmental exposures and response necessitated this extensive state-of-the-science review. Here, we present review results with the goals of comprehensively summarizing and evaluating recent air pollution research from studies using untargeted HRM, identifying gaps in the peer-reviewed literature, and outlining a path forward for future use of this analytical platform. In this analysis, we highlight various methodological approaches applied at each step of the air pollution metabolomics workflow and provide an evaluation of areas of concordance and dissimilarity in the reported findings. Specifically, the key information of the air pollution metabolomics workflow in each identified study was scrutinized and summarized, including study design, sample size, type of biospecimens, air pollution assessment, untargeted metabolomics assessment, and analysis.

Methods

To summarize recent research of air pollution studies using untargeted metabolomics and to identify gaps in the peer-reviewed literature and opportunities for addressing these gaps in future designs, we used a systematic method to conduct a state-of-the-science review, and the method comprised primarily four components: *a*) the *a priori* specification of inclusion criteria for eligible studies based on the scope of the review and the research question, *b*) a reproducible process of searching and screening, *c*) an explicit and unbiased protocol of data extraction, and *d*) a straightforward synthesis and presentation of extracted data.

Eligibility Criteria

We developed the following eligibility criteria, *a priori*, in selecting papers included in the state-of-the-science review: *a*) the study was conducted in humans; *b*) the study focused on air pollution exposures; *c*) the study employed untargeted metabolomics profiling (i.e., global measurements of metabolic features) in human biospecimens; *d*) the study assessed the association between changes in metabolomic profiles and exposure to air pollution; and *e*) the study was either observational or experimental. If more than one publication was identified for the same study population, each publication would be retained if different air pollutants were investigated among the different publications. Otherwise, the publication that provided more comprehensive information on study design and analytical protocol would be included. Studies that did not meet the eligibility criteria were excluded. In addition, the following exclusion criteria were applied: *a*) The studies did not contain original data, such as reviews, editorials, or commentaries; *b*) the studies were not peer-reviewed (e.g., conference abstracts, technical reports, preprints, theses, and dissertations, working papers from research groups or committees, and white papers); *c*) the studies involved *in vitro* molecular analyses of human tissues and cells; *d*) the studies employed targeted metabolomic methods, which involved multiplexed analysis of a limited suite of known metabolites (i.e., specific compounds were hypothesized *a priori* and quantified via comparison with established reference ranges)[65,66]; and *e*) the studies were not available in English.

Search Strategy

We examined all peer-reviewed manuscripts published between 1 January 2005 and 31 March 2022 for eligible studies. This period was chosen because metabolomics analyses were rarely used in epidemiological modeling prior to 2005 and the number of articles reporting air pollution metabolomics use has grown exponentially during recent years. Articles were identified from the online databases PubMed and Web of Science. In addition, the reference lists of eligible studies were manually searched to capture articles that may have been missed in the database search. We conducted the database search using all combinations of the following two categories of terms connected with the Boolean operator AND: *a*) exposure to air pollution resulting from indoor, outdoor, or traffic-related sources: "air pollution," "air pollutant," "traffic-related pollution," "traffic-related pollutant," "traffic pollution," "traffic pollutant," "particulate matter," "coarse particle," "fine particle," "ultrafine particle," or "ozone" and *b*) HRM: "metabolomics," "metabolomic," "metabolome," "metabolic profile," or "metabolic networks and pathways." All terms were searched using both controlled vocabulary [Medical Subject Headings (MeSH) in PubMed], if applicable, and free text words in titles or abstracts. The detailed search queries were shown in Table S1.

All search results ($N = 2{,}065$) were then imported into Covidence (Veritas Health Innovation Ltd),[67] a screening and data extraction tool for Cochrane authors, which we used to complete the selection process (Figure S1). After the removal of duplicate studies, we employed a multistage screening process to select studies for inclusion. In the first stage, two researchers (Z.L. and Z.T.) conducted title and abstract screening independently and removed those irrelevant with respect to the eligibility criteria. After the screening, we combined the screening results, and the conflicts were resolved via the consensus of two senior researchers (D.L. and J.S.). Then, the two independent researchers conducted the full-text review of the screening results of the first stage. The conflicts regarding the studies to include were resolved in the same way as in the first stage. Finally, we manually examined references of included studies to see if there were additional studies to be included in the state-of-the-science review (i.e., snowballing).

Data Extraction

For each study, we collected the following data: citation (authors and year of publication), study design, study population, sample size, exposure assessment, biospecimen type [i.e., whole blood, plasma, serum, saliva, urine, exhaled breath condensate (EBC), dried blood spots, and bronchoalveolar lavage fluid (BALF)], experimental methods for metabolome assessment, laboratory information, statistical analysis approaches, and study findings. The information extracted for exposure assessment included type of air pollutants, exposure time window, and exposure assessment methods. The exposure assessment methods were classified into seven categories: air quality modeling, fixed-site monitoring, microenvironment monitoring, personal monitoring using portable sensors, controlled personal exposure, microenvironment modeling, and personal biomonitoring.[68] We defined microenvironmental monitoring as measuring air pollution concentrations in the small-scale built environment where people spent considerable time during the study period, such as a classroom or dormitory, and fixed-site monitoring as air pollution data collected at ambient ground-based monitoring sites. The information necessary for assessing concordance and dissimilarity in untargeted metabolomics assessment and analysis across the eligible studies was extracted and comprised untargeted metabolic profiling, feature extraction, metabolome-wide association study (MWAS) statistical method coupled with pathway enrichment analysis and

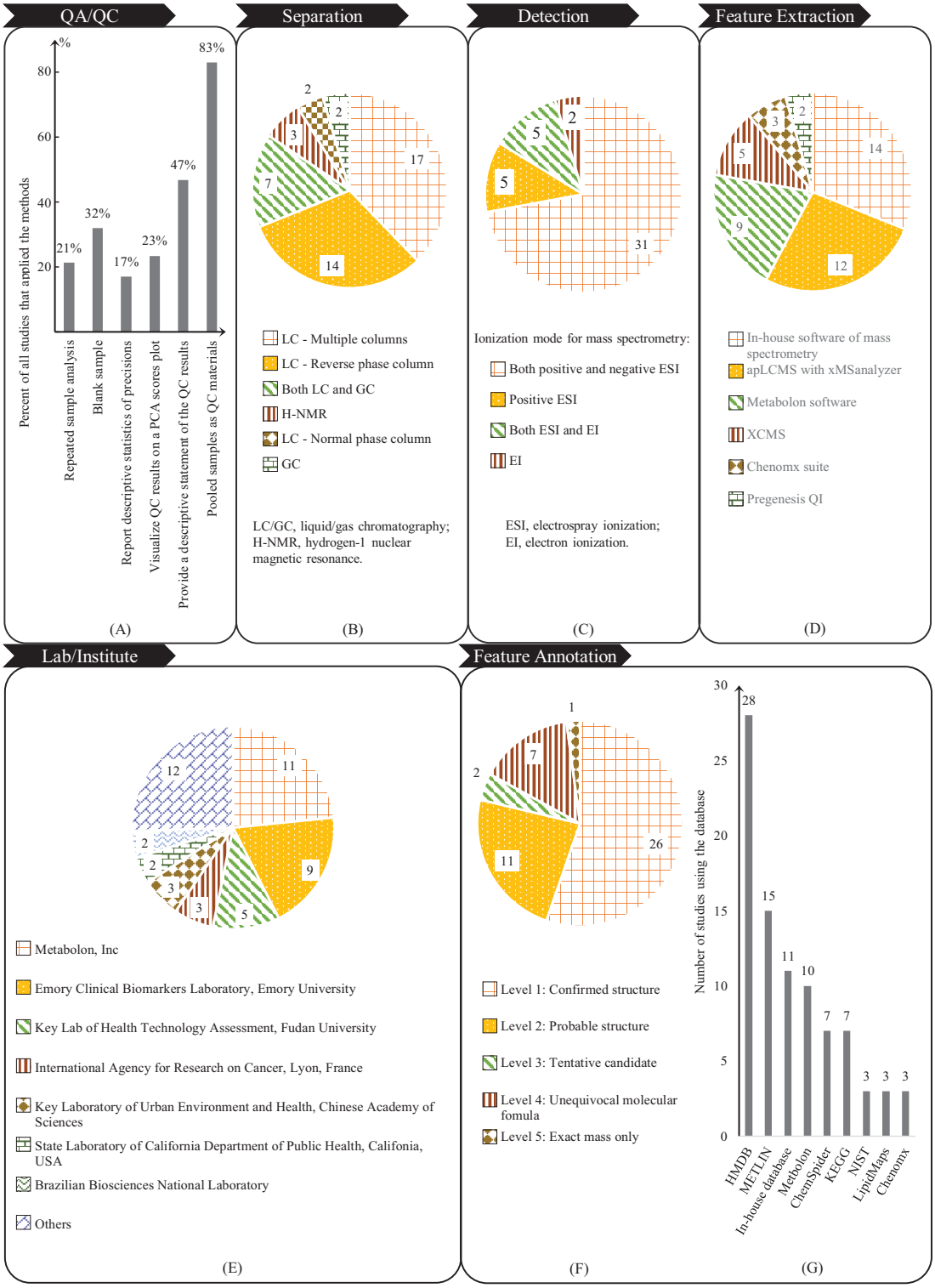

Figure 1. The experimental and statistical methods reported in the 47 eligible studies following a general workflow of the metabolome-wide association study. (A) QA process and QC materials reported. (B) Sample separation methods. (C) Mass spectrometry ionization methods. (D) Software of metabolic feature extraction. (E) Laboratory/institute conducting the experimental analysis. (F) Highest confidence level of the Metabolomics Standards Initiative metabolite identity reported. (G) Online databases used in at least three studies for putative annotation. (H) Statistical modeling approaches used to identify significant metabolic features by study design. (I) Covariates controlled in more than three studies. (J) Algorithms used for network or pathway analysis. The characteristics of experimental and statistical analysis was summarized by the frequency or percentage of each item of interest reported among all eligible studies. The *y*-axis of the bar charts denotes the number of studies, and each slice of the pie charts represents the percentage of each item reported among all 47 studies. The confidence levels refer to a communication confidence level system regarding the identification confidence of unknown molecules with level 1 as the most confident. The data can be found in Table S2 and Table 2. Note: ANOVA, analysis of variance; BMI, body mass index; HMDB, Human Metabolome Database; KEGG, Kyoto Encyclopedia of Genes and Genomes; NIST, National Institute of Standards and Technology (database); OPLS-DA, orthogonal partial least squares–discriminant analysis; PCA, principal component analysis; PLS-DA, partial least squares–discriminant analysis; QA, quality assurance; QC, quality control.

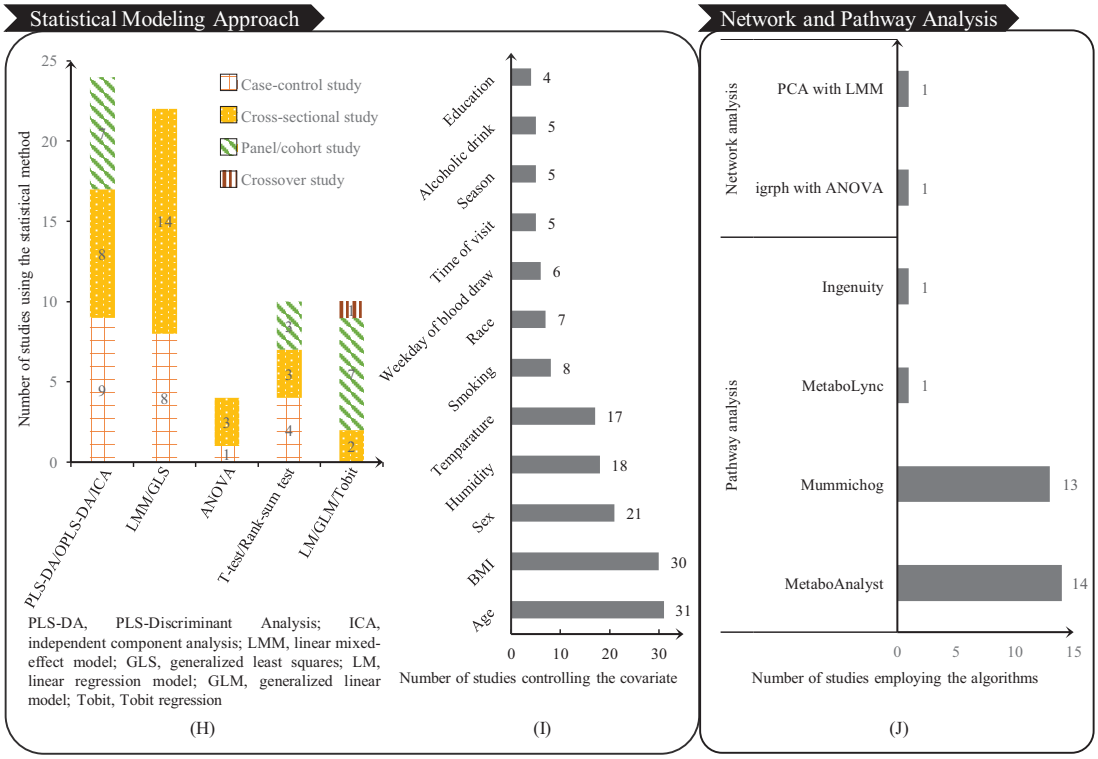

Figure 1. (*Continued.*)

chemical annotation. Untargeted metabolic profiling and feature extraction is detecting and quantifying a relatively comprehensive list of low-molecular-weight metabolites and their intermediates in biospecimens using high-throughput experimental platforms, and the information extracted for it included analytical platform, technical column, ionization mode, feature extraction algorithm, data preprocessing setting, and quality assurance/quality control (QA/QC) method. Statistical analysis approaches of MWAS consisted of three major components: *a*) approaches identifying significant metabolic features associated with the exposure, *b*) approaches identifying significant biological pathways (i.e., pathway enrichment analysis) associated with the exposure, and *c*) approaches annotating the significant metabolic features. We also extracted covariates and effect modifiers that were considered in each study if applicable. As for the study findings, we extracted the information on the significant metabolites with known identities and significant biological pathways.

Metabolite Identification and Validation

The field of untargeted metabolomics, specifically, has led to the substantial improvement in the detection and identification of new biomarkers of environmental pollutants. Most untargeted metabolomics studies include various forms of metabolite validation as a key component in identifying and annotating HRM feature output. We categorized confidence in the identification and validation of the eligible study results into five levels, following the reporting standard proposed by Schymanski et al.[69] Briefly, level 1 represents that the proposed chemical identity has been confirmed via comparison with an authentic standard by mass spectrum and retention time; level 2 represents comparing the mass spectrum of detected chemical to library spectrum data where the spectrum–structure match is unambiguous; level 3 represents tentative candidates that are proposed without sufficient information for one exact structure only; level 4 represents a situation where only an unequivocal molecular formula is reported; and level 5 reports only exact mass. In addition, for the eligible studies using hydrogen nuclear magnetic resonance (H-NMR), we treated the reporting identities of metabolites as level 2 if the NMR spectra were compared with online databases or as level 1 if compared with a laboratory-authentic standard.

Results Synthesis

The information regarding citation, study design, and major findings of the eligible studies are summarized in Table 1. Given that the findings from the eligible studies were not amenable to meta-analysis owing to discrepancies in laboratory protocols and data analysis approaches, we collected and summarized the primary results of the studies (i.e., significant metabolites and biological pathways) in figure/table format to facilitate insight into complex metabolomics data from multiple studies. To integrate the reported metabolites and biological pathways that were associated with various air pollution exposures, we categorized the exposures of interest into eight groups: total particulate matter (PM), carbonaceous PM, PM metals/metalloids, ultrafine particles (UFPs), carbon monoxide (CO), nitrogen oxides (NO_x), ozone (O_3), and combined exposure. Furthermore, to evaluate the impact of the use of experimental and statistical approaches on the pathway enrichment results, we also summarized the biological pathways by laboratory/institute, pathway analysis algorithm, and biospecimen (Figures S3 and S4).

Results

At the first stage of the title and abstract screening, after removing duplicates, we identified 1,609 unique citations from PubMed and Web of Science for which 64 met the eligibility criteria. At the second stage of full-text reviews, 17 articles were further excluded owing to failure to meet the inclusion/exclusion criteria (Figure S1). At the final stage of reference review, no additional studies were identified. A total of 47 studies fulfilled the final

Table 1. Summary of the study design characteristics of the reviewed studies on untargeted metabolomics application in air pollution health research ($N = 47$).

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Two independent randomized crossover trials (Oxford Street II: 2 scenarios; TAPAS II: 4 scenarios)	van Veldhoven et al.[29]	Oxford Street II (London)—56 participants (Group 1: healthy volunteers ($n = 18$); Group 2: patients ($n = 19$) with COPD; Group 3: patients ($n = 19$) with clinically stable ischemic heart disease (IHD)	Oxford Street II: 336 serum samples	Oxford Street II: short-term assessment on $PM_{2.5}$, PM_{10}, UFP, BC, NO_2	International Agency for Research on Cancer, Lyon, France	There were 29 and 77 metabolic features associated with at least one pollutant in the two study populations, respectively. Different pollutants may impact different metabolic features given few features associated across pollutants. In Oxford Street II, the acyl-carnitine pathway was associated with NO_2, previously found to be involved in cardiorespiratory disease. No overlapping metabolic feature was found in both studies.
		TAPAS II (Barcelona)—28 healthy, nonsmoking, non-medication using adults, without high occupational exposures to TRAP	TAPAS II:120 serum samples	TAPAS II: short-term assessment on $PM_{2.5}$, PM_{10}, PMcoarse ($PM_{2.5}$- PM_{10}), NO_x		
Randomized, double-blind crossover trial	Li et al.[22]	55 healthy college students in Shanghai, China	110 serum samples	Short-term assessment on $PM_{2.5}$	Key Laboratory of Health Technology Assessment, Fudan University	Increases in cortisol, cortisone, epinephrine, and norepinephrine were associated with a higher level of $PM_{2.5}$ exposure. Differences in glucose, amino acids, fatty acids, and lipids were also observed across air treatments.
Randomized, semi-controlled, crossover panel study, 3 scenarios	Ladva et al.[21]	49 commuters, a subset of Atlanta Commuters Exposure (ACE-2) study	98 Plasma samples	Short-term assessment on $PM_{2.5}$, BC, pb-PAH, PNC, noise, Al, Fe, Pb, OC, WSOC	Emory Clinical Biomarkers Laboratory, Emory University	HRM identified 10-h perturbations in 110 metabolic features involved in arachidonic acid, leukotriene, and tryptophan metabolism following exposures to in-vehicle particulate metals (Al, Pb, and Fe). The perturbations in plasma metabolome 10-h after exposures were associated with 2-h changes in pro-inflammatory biomarkers, indicating an impact on amino acid, leukotriene, and anti-oxidant metabolism. The annotated compound, 20-OH-LTB4, decreased after exposures to in-vehicle particulate metals, indicating a subclinical immune response.
Randomized, double-blind, crossover trial, 2 scenarios	Chen et al.[18]	45 healthy college students in Shanghai, China	90 urine samples	Short-term assessment on $PM_{2.5}$	Key Laboratory of Health Technology Assessment of the Ministry of Health, Fudan University	Differences in 1,115 metabolites were observed between sham and real purification scenarios, and 71 were associated with $PM_{2.5}$ exposure. Among these significant metabolites were 16 lipids, 5 purines, 2 neurotransmitters, and 3 coenzymes.
Randomized, double-bind crossover study	Liu et al.[37]	44 healthy children	264 urine samples	Short-term assessment on size-fractionated PM, O_3	Key Laboratory of Urban Environment and Health, Chinese Academy of Sciences	There were 28 and 14 metabolites significantly correlated with NAI and PM, respectively, whereas 8 and 18 were associated with respiratory function and HRV, respectively. The respiratory function was improved by increased NAI and decreased PM mainly via eight pathways, promoting energy production, anti-inflammation and anti-oxidation capacity. On the other hand, HRV was improved with decreased PM via six main pathways, increasing energy production and anti-inflammation capacity. In comparison, increased NAI worsened HRV via five main pathways, lowering energy generation and anti-oxidation capacity.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Randomized crossover study, 2 scenarios	Zhang et al.[35]	39 nonsmoking, healthy college students in Beijing	78 urine samples	Short-term assessment on size-fractioned PM ($PM_{0.5}$, $PM_{1.0}$, $PM_{2.5}$, $PM_{5.0}$, and PM_{10})	Beijing Key Laboratory of Environmental Toxicology, Capital Medical University	Changes in 4 and 7 urine metabolites in men and women, respectively, were observed after 4-h exposure to PM in a subway system. $PM_{1.0}$ was found as the most influential indicator among all size-fractioned PM of cardiovascular effects and urine metabolites in both sexes. 8-OHdG and prolyl-arginine played opposing roles in male HRV and HR.
Randomized crossover trial, 2 scenarios	Du et al.[47]	35 college students	70 serum samples	Short-term assessment on UFP, $PM_{2.5}$, BC, NO_2, CO Traffic-polluted vs. traffic-free	Key Laboratory of Health Technology of Assessment, Fudan University	TRAP was significantly associated with 128 serum metabolites. Dozens of regulatory pathways—such as inflammation, oxidative stress, coagulation, endothelin-1 signaling, and renin-angiotensin signaling—were altered in response to TRAP shown in the multi-omics analysis.
Crossover study, 2 scenarios	Liu et al.[54]	31 healthy adults (11 males, 20 females)	62 serum samples	Short-term assessment on Co, Ni, Cd, Cu, Ag, Ba in $PM_{2.5}$ and PM_{10}	Department of Global Health, School of Medicine, Wuhan University	Brief exposures to heavy metals were associated with changes in serum cardiovascular-related metabolites among young adults, including increased SM(d18:1/17:0) and Sphingomyelin, and decreased GlcCer(d16:1/18:0) and galabiosylceramide. The changes were also accompanied by activation of the sphingolipid metabolism pathway. Inhaled heavy metals posed noncarcinogenic and carcinogenic risks, among which Ni and Cd were found to be main contributors. Increased exposure to heavy metals may increase health risks by inducing cardiovascular-related metabolites disturbance via activating the sphingolipid metabolism pathway.
Randomized crossover study	Vlaanderen et al.[30]	31 healthy nonsmoking volunteers (21 females and 10 males)	493 serum samples	Short-term assessment on $PM_{2.5}$, PM_{10}, PMcoarse ($PM_{2.5-}$ PM_{10}), NO_2, NO_x, O_3, PNC	International Agency for Research on Cancer, Lyon, France	Air pollutants were associated with changes in metabolic features after 2 h following exposure. Several significant metabolic features were also associated with acute health response, including FEV1. Tyrosine, guanosine, and hypoxanthine were annotated among significant features. Eight pathways, including tyrosine metabolism, were significantly enriched by mummichog.
Randomized, semi-controlled, crossover panel study, 3 scenarios	Liang et al.[23]	24 commuters with asthma+21 without asthma, a subset of Atlanta Commuters Exposure (ACE-2) study	140 plasma samples	Short-term assessment on 27 air pollutants (PNC, noise, pb-PAH, $PM_{2.5}$, BC, EC, OC, WSOC, As, Cr, Ni, V, Al, Ca, Ce, Mg, Ba, Cd, Co, Fe, K, Mn, P, Pb, Sb, Ti, Zn)	Emory Clinical Biomarkers Laboratory, Emory University	Alterations of several metabolic pathways related to inflammation and oxidative stress were detected by pathway analysis, including leukotriene, vitamin E, cytochrome P450, and tryptophan metabolism. In these pathways, 45 unique metabolites were identified, including arginine, histidine, and methionine. Exposure to TRAP might perturb interrelated molecular networks centering on arginine metabolism.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Crossover clinical study	Miller et al.[26]	24 healthy young adults (20 males and 4 females)	48 serum samples	Short-term assessment on O_3	Metabolon, Inc.	Increases in monoacylglycerol, glycerol, and medium- and long-chain free fatty acids involved in lipid mobilization and catabolism were associated with O_3 exposure. O_3 exposure was also associated with increased serum cortisol, corticosterone, and circulating mitochondrial beta-oxidation-derived metabolites, such as acyl-carnitines. Pathway analysis indicated alterations of seven pathways belonging to lipid metabolism, including sphingolipid metabolism, endocannabinoid synthesis, fatty acid metabolism, beta-oxidation, dicarboxylic acid metabolism, steroid hormone biosynthesis, and phospholipid metabolism.
Two-arm crossover study, 9 of 23 returned for a 2nd randomized crossover study	Cheng et al.[19]	23 nonsmoking, healthy adults	BALF	Short-term assessment on O_3	Metabolon, Inc.	There were 28 metabolites differentially expressed following 1-h post-O_3 exposure compared with filtered air exposure, and 41 were differentially expressed at 24-h postexposure. The changes at 1 h suggested acute phase increased the cellular response to oxidative stress, whereas the changes at 24 h were consistent with the ongoing repair of airway tissue.
Randomized controlled, double-blind crossover trial	Surowiec et al.[28]	15 nonsmoking healthy volunteers (8 males and 7 females)	30 BALF and 30 BW samples	Short-term assessment on pure biodiesel exhaust	Swedish Metabolomics Center (SMC), Sweden	BALF and BW samples had different metabolite profiles with 46 metabolites showing different levels. Exposure to biodiesel exhaust was associated with the levels of 1-monostearylglycerol, sucrose, inosine, nonanoic acid, and ethanolamine, niacinamide (in BAL) and pentadecanoic acid and lactic acid (in BW).
Randomized single-blind crossover trail, two scenarios	Cruz et al.[45]	15 healthy males	45 serum samples	Short-term assessment on $PM_{2.5}$, PM_{10}, NO_xTRAP vs. filtered air	Brazilian Biosciences National Laboratory, Brazil	Incomplete fatty acid metabolism was detected under the TRAP condition 10 min after exercise by enrichment analysis, and an overactivity of ketone body metabolism at 10 min and at 1 h after exercise was detected with TRAP.
Randomized crossover trial, 2 scenarios	Cruz et al.[46]	10 healthy males	40 serum samples	Short-term assessment on $PM_{2.5}$, total PM, NO, NO_2. TRAP vs. filtered air	Brazilian Biosciences National Laboratory, Brazil	Comparing the samples taken at 30, 60, and 90 min after experiment to the baseline, 12, 16, and 18 metabolites were identified as discriminants between TRAP and filtered conditions, respectively.
Cohort study	Nassan et al.[40]	456 White men	648 fasting blood plasma samples	Long-term assessment on $PM_{2.5}$, NO_2, O_3	Metabolon, Inc.	Statistically significant associations with several metabolites were observed (58 with $PM_{2.5}$ and 15 with NO_2, whereas there were none with O_3). One of five ICA factors (factor 2) was associated with $PM_{2.5}$. Eight altered metabolic pathways were associated with long-term exposure to $PM_{2.5}$ and which were related to inflammation, oxidative stress, immunity, and nucleic acid damage and repair.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Cohort study	Nassan et al.[63]	456 White men	648 fasting blood plasma samples	Short- and long-term assessment on $PM_{2.5}$ and 14 of its species (UFP, BC, Na, Fe, Al, Si, K, Ni, V, S, Se, Pb, Zn, Ti)	Metabolon, Inc.	In short-term $PM_{2.5}$ exposure, UFP was associated with the greatest number of metabolic features, followed by Ni, V, K, Si, and Al. Long-term exposures to BC, V, Zn, Ni, Fe, Cu, and Se were associated with at least one feature. Metabolic pathways, including glycerophospholipid, sphingolipid, and glutathione metabolism, were perturbed by $PM_{2.5}$ species, which were involved in inflammation, oxidative stress, immunity, and nucleic acid damage and repair.
Cohort study	Nassan et al.[39]	456 White men	648 fasting blood plasma samples	Short-term assessment on $PM_{2.5}$, NO_2, O_3	Metabolon, Inc.	In the adjusted models, exposure to NO_2 was associated with 19 metabolites. Six metabolic pathways were perturbed by short-term exposure to air pollution and temperature and which involved inflammation, oxidative stress, immunity, and nucleic acid damage and repair.
Cohort study	Loo et al.[36]	53 Chinese women	152 dried blood spot samples	Long-term assessment on HAP Short-term assessment on $PM_{2.5}$, BC	Key Laboratory of Health Technology Assessment, Fudan University	Metabolites, such as amino acids, acyl-carnitines, lysophosphorylcholines, sphinganine, and choline, were detected in the DBS specimens. The approach was able to detect the differences in personal exposure to HAP via DBS metabolome profiles.
Panel study	Chu et al.[44]	78 college students	624 plasma samples	Short-term assessment on $PM_{2.5}$, PM_{10} Filtered vs. non-filtered	State Key Laboratory of Natural Medicines, China Pharmaceutical University	$PM_{2.5}$ exposure was associated with 120 metabolic features, and 25 of them were identified, most of which were phospholipids. PM_{10} exposure was associated with 42 features, and no distinctive metabolites were found. The levels of LysoPC (P-20:0) and LysoPC [P-18:1(9z)] changed significantly before and after air purifier intervention.
Panel study, 5 visits (including baseline)	Zhao et al.[62]	76 healthy seniors (37 males, 39 females)	350 serum samples	Short-term assessment on $PM_{2.5}$	Dian Diagnostics	$PM_{2.5}$ exposure was associated with 253 metabolites, 11 pathways, a higher insulin resistance index, and increased functional biomarkers, suggesting that $PM_{2.5}$ exposure may contribute to systemic inflammation and altered sphingolipid metabolism.
Panel study	Gao et al.[49]	74 apprentices and instructors attending a welding skill training center	509 plasma samples	Short-term assessment on $PM_{2.5}$Welding/classroom area and morning/afternoon	Metabolon, Inc.	Sphingosine 1-phosphate (S1P) and sphingasine 1-phosphate (SA1P) exhibited significant interaction effects between welding day and time. S1P, SA1P and sphingosine shared similar trends over time: high relative levels in the morning of a non-welding day declining by afternoon, but with lower starting levels on a welding day and no decline.
Panel study	Walker et al.[60]	73 nonsmoking males	292 plasma samples	Short-term assessment on EC, OC, $PM_{2.5}$	Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai	Metabolic changes associated with EC exposure suggest increased lipid peroxidation products, biomarkers of oxidative stress, thrombotic signaling lipids, and metabolites associated with endothelial dysfunction from altered NO metabolism, whereas OC exposures were associated with anti-oxidants, oxidative stress biomarkers, and critical intermediates in NO production. Correlation with whole blood RNA gene expression provided additional evidence of changes in processes related to endothelial function, immune response, inflammation, and oxidative stress.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Panel study, 2 scenarios	Huan et al.[52]	63 healthy college students	126 serum samples	Short-term assessment on $PM_{2.5}$ High-exposure vs. low-exposure season	Tongji Medical College	Metabolic profiling was separated clearly between high- and low-exposure seasons by OPLS-DA in males. The changes in 14 serum metabolites were significantly associated with $PM_{2.5}$ exposure in males. The metabolites with decreased levels were involved in heme metabolism, energy metabolism and oxidative stress, phospholipid metabolism, and tryptophan metabolism. Thyrotropin-releasing hormone, glutathione, and phosphatidylethanolamine were increased, which were related to energy metabolism and oxidative stress, and phospholipid metabolism.
Panel study, 4 visits	Tang et al.[59]	54 college students	175 plasma samples and 204 saliva samples	Short-term assessment on oxidative potential of $PM_{2.5}$	Emory Clinical Biomarkers Laboratory, Emory University	Leukotriene metabolism was associated with most FPMOP indicators in both plasma and saliva. Changes in metabolic features associated with water-soluble and -insoluble FPMOPs suggested different patterns of perturbed pathways. Five identified metabolites (hypoxanthine, histidine, pyruvate, lactate/glyceraldehyde, and azelaic acid) were associated with FPMOP, indicating perturbations in acute inflammation, nucleic acid damage and repair, and energy perturbation. Metabolic signals specific to FPMOP, not to PM mass, were observed, suggesting that FPMOP might be a more sensitive, health-relevant measure for elucidating PM toxicity.
Panel study, 4 visits	Liang et al.[24]	54 college students (near dorm: 24 vs. far dorm: 30), the Dorm Room Inhalation to Vehicle Emission (DRIVE) study	175 plasma samples and 204 saliva samples	Short-term assessment on BC, CO, NO, NO_2, NO_x, $PM_{2.5}$	Emory Clinical Biomarkers Laboratory, Emory University	In total, 20,766 and 29,013 metabolic features extracted from plasma and saliva samples, respectively. More than 10,000 metabolic features were shared by plasma and saliva samples. There were 1,291 metabolic features associated with at least one traffic indicator, such as BC, CO, NO_x, or $PM_{2.5}$. Elicitation of inflammatory and oxidative stress pathways, such as leukotriene and vitamin E metabolism, were associated with traffic exposures. Ten of the significant features were identified, including arginine, histidine, gamma-linolenic acid, and hypoxanthine.
Panel study	Liu et al.[55]	46 healthy children (24 boys, 22 girls)	138 urine samples	Short-term assessment on O_3, $PM_{2.5}$, PM_{10}, BC	Laboratory of Institute of Urban Environment, Chinese Academy of Sciences	Nine metabolites mainly involved in amino acid and bile acid metabolism were associated with O_3, four of which were associated with cardiorespiratory function indices. The metabolites related to bile acid and RNA showed greater decreases in boys. Furthermore, cholestane-3,7,12,25-tetrol-3-glucuronide mediated 26.67% of the positive association between O_3 and heart rate.
Panel study, 1 visit	Baker et al.[17]	37 workers (exposed: 20 vs. unexposed: 17)	37 urine samplesStudy 2012: 16 plasma samples	Short-term assessment on inhalable Mn	Department of Medicinal Chemistry, University of Washington	Fifteen metabolic features showed a significant difference between the exposed and unexposed training sets, and 9 metabolic features showed a difference in the validation set as well. Some of the 9 features showed an exposure–response relationship for exposure levels divided as lower vs. higher exposure.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Panel study, 4 visits	Xia et al.[33]	43 nonsmoking college students in Shanghai	145 serum samples	Short-term assessment on O_3	Key Laboratory of Health Technology Assessment, Fudan University	Personal O_3 exposure was significantly associated with 36 metabolites. Among these, 25 showed a positive association, primarily belonging to fatty acids, phosphatidylcholines, lysophosphatidylcholines, phosphatidyl ethanolamines, and lysophosphatidyl ethanolamines.
Panel study, 3 scenarios	Mu et al.[27]	26 nonsmokers 30–65 years of age, a subset of the Beijing Olympics Air Pollution (BoaP) study	78 serum samples	Short-term assessment on PM_1, $PM_{2.5}$, PM_7, PM_{10}, and total suspended particulates (TSPs)	Metabolon, Inc.	Network partitioning identified four modules with 69 known metabolites significantly changed across three time points. All known molecules in the first module ($n = 33$) were lipids; the second module, primarily dipeptides ($n = 24$); and the third module, mostly unknown. Enriched pathways included long- and medium-chain fatty acids, polyunsaturated fatty acids (n3 and n6), eicosanoids, lysolipid, dipeptides, fatty acid metabolism, and purine metabolism [(hypo) xanthine/inosine–containing pathways].
Panel study, 2–5 visits	Selley et al.[58]	21 healthy, nonsmoking university students	86 urine samples	Short-term assessment on UFP (use of PNCs as an indicator), NO_X, CO, BC, and $PM_{2.5}$	Medical Research Council Toxicology Unit, University of Cambridge	Total PNC at the exposure site was associated with decreases in urinary taurine and dimethylamine, and so was UFP produced during both aircraft landing and take-off. A significant decrease in pyroglutamate was also associated with UFP, specifically contributed by both landing and take-off exposure. Non-aviation UFPs had minimal impact on the urinary metabolome, which did not significantly change the overall response to airport UFP exposure.
Panel study	Feng et al.[48]	20 nonsmoking healthy adults	40 plasma samples	Short-term assessment on $PM_{2.5}$, BC, NO_2, CO, SO_2	Center of Biomedical Analysis, Tsinghua University	Metabolites altered by short-term air pollution exposure were mainly involved in amino acid metabolism.
Two-stage, self-controlled panel study	Wei et al.[32]	11 nonsmoking boilermakers without diabetes (combined analysis: $n = 14$; study 2011: $n = 11$; study 2012: $n = 8$)	Study 2011: 22 plasma samples	Short-term assessment on $PM_{2.5}$	Metabolon, Inc.	Unsaturated fatty acids were negatively associated with respirable welding fume exposure, showing an exposure–response relationship. A decrease in eicosapentaenoic acid, docosapentaenoic acid n3, and docosapentaenoic acid n6 were associated with the exposure. Enriched pathways involved the unsaturated fatty acid pathway.
Panel study, 4 visits	Ladva et al.[20]	4 commuters, a subset of Atlanta Commuters Exposure (ACE-2) study	16 samples for each of these biometrics: plasma, EBC, saliva	Short-term assessment on traffic-related pollutants, not specific ($PM_{2.5}$, PNC, organic components, transition metals)	Emory Clinical Biomarkers Laboratory, Emory University	Correlations across all pairwise comparisons of metabolic features in plasma, EBC, and saliva were moderate-to-strong, with the strongest between EBC and saliva. After controlling for participant and sampling time, associations of mean feature intensities between matrix pairs were positive and significant. Six features in all three biosamples were matched to a known mobile-source air toxics list.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Two independent case–control studies	Jeong et al.[43]	*a*) 139 patients with asthma (cases) + 196 participants without asthma (controls), nested in the Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults (SAPALDIA) *b*) 166 cases cardio-cerebrovascular+155 controls, nested in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort	SAPALDIA: 335 serum samples; EPIC: 321 serum samples	SAPALDIA: long-term assessment on $PM_{2.5}$, UFP (PNC, LDSA), NO_2 EPIC: long-term assessment on $PM_{2.5}$, UFP(PNC), NO_2	International Agency for Research on Cancer, Lyon, France	A strong association of adult-onset asthma (AOA) with UFP was observed. Perturbation of linoleate metabolism was simultaneously associated with air pollution exposure, AOA, and cardio-cerebrovascular diseases.
Cross-sectional study	Menni et al.[25]	523 participants, a subset of Twins UK cohort	Fasting blood	Long-term assessment on $PM_{2.5}$, PM_{10}	Metabolon, Inc.	Eight of 21 metabolites associated with altered lung function were also associated with both $PM_{2.5}$ and PM_{10}. The strongest association both with $PM_{2.5}$ and FEV1 was seen with alpha-tocopherol levels.
Cross-sectional study	Chen et al.[41]	252 participants [111 high exposure (40 children and 71 elderly) vs. 141 low exposure (70 children and 71 elderly)]	252 urine samples	Short-term assessment on heavy metals (V, Cr, Ni, As, Cu, Sr, Cd, Hg, Tl and Pb), and PAHs	Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University	There were 45 metabolites identified in children related to the exposure and 42 in older subjects; 11 were present in both cohorts. Pathway analysis identified age-dependent biological pathways, including tryptophan and phenylalanine metabolism in children, and glycine, serine, and threonine metabolism in older participants.
Cross-sectional study	Ritz et al.[57]	241 children	241 serum samples	Short-term assessment on $PM_{2.5}$	State Laboratory of California Department of Public Health, California, USA	After controlling for confounding factors, partial least square regression detected 402 and 182 features in HILIC positive and C18 negative profiles, respectively, as significantly associated with increasing $PM_{2.5}$ exposure in the third trimester. Pathway enrichment analysis identified oxidative stress and inflammation pathways as being altered, primarily involving lipid metabolism.
Cross-sectional study	Hood et al.[51]	200 women receiving ART	200 serum samples	Short-term assessment on $PM_{2.5}$	Emory Clinical Biomarkers Laboratory, Emory University	We identified 17 pathways solely associated with acute $PM_{2.5}$ exposure and involved in amino acid, lipid, energy, and nutrient metabolism. Fifteen pathways were solely associated with long-term $PM_{2.5}$ exposure and mainly involved in pro-inflammatory, anti-inflammatory, amino acid, and energy metabolism. Seven pathways were associated with the majority of exposure windows and were primarily involved in anti-inflammatory and lipid metabolism.
Cross-sectional study	Gaskins et al.[50]	200 women receiving ART	200 serum samples	Short-term assessment on $PM_{2.5}$, NO_2, O_3, BC	Emory Clinical Biomarkers Laboratory, Emory University	There were 190 metabolic features and 18 pathways that were associated with both air pollution and birth outcome across chromatography columns. Eight identified metabolites suggested amino acid and nutrient metabolism changes with downstream effects on oxidative stress and inflammation. Six identified metabolites fell into two intuitive clusters—"anti-oxidants" and "oxidants"—which might partly mediate the association between air pollution and live birth. Tryptophan and vitamin B3 metabolism were enriched pathways linking air pollution exposure to decreased odds of live birth.

Table 1. (*Continued.*)

		Study design				
Type of study	Reference	Study population	Biometric	Air pollutant indicator	Laboratory information	Primary results
Cross-sectional study	Li et al.[38]	180 American adults	180 plasma samples	Short-term assessment on CO, NO_2, O_3, $PM_{2.5}$, EC, OC	Emory Clinical Biomarkers Laboratory, Emory University	The identified pathways associated with one or more air pollutants are involved in various biochemical processes, including nucleic acids damage and repair (pyrimidine metabolism and purine metabolism), nutrient metabolism (e.g., fatty acid beta-oxidation, tryptophan metabolism, vitamin A metabolism), and acute inflammation (e.g., histidine metabolism, tyrosine metabolism, alanine and aspartate metabolism).
Cross-sectional study	Yan et al.[34]	160 pregnant women (high exposure: 98 vs. low exposure: 62)	160 midpregnancy serum samples	Short-term assessment on CO, NO_x, $PM_{2.5}$	State Laboratory of California Department of Public Health, California, USA	PLS-DA detected 181 and 251 metabolic features (HILIC and C18 columns, respectively) discriminated between the high- and low-exposure group. Pathway enrichment analysis indicated altered oxidative stress and inflammation-related pathways, including linoleate, leukotriene, and prostaglandin metabolisms.
Cross-sectional study	Misra et al.[56]	134 women	134 D-quame samples (stratum corneum)	Long-term assessment on PAH, PAH metabolites	Metabolon, Inc.	At a higher level, increased enrichment was observed in samples from Baoding (more polluted) that included amino acid and fatty acid metabolism. Elevated levels of *N*-acetyl amino acids, gamma-glutamyl amino acids, and urea cycle intermediates were observed in more polluted areas.
Cross-sectional study	Liao et al.[53]	98 young adults with childhood asthma history	98 serum samples	Short- and long-term assessment on $PM_{2.5}$, PM_{10}, NO_2, O_3, and NRAP	Metabolon, Inc.	Network analysis found that exposures to air pollutant mixture were connected to 357 gene markers and 92 metabolites. One-year and 1-month averaged $PM_{2.5}$ and NO_2 levels were associated with several amino acids involved in serine, glycine, and beta-alanine metabolism. Lower serum levels of carnosine, aspartate, and choline were associated with worse asthma control.
Cross-sectional study, 1 visit	Walker et al.[31]	59 nonsmoking, non-Hispanic White residents (≥40 y old) of Somerville and Dorchester/South Boston, a subset of Community Assessment of Freeway Exposure and Health (CAFEH) study	59 plasma samples	Long-term assessment on UFP	Emory Clinical Biomarkers Laboratory, Emory University	Five metabolites, including arginine, aspartic acid, glutamine, cystine, and methionine sulfoxide, were differentially expressed in low- vs. high-exposure groups. There were 316 metabolic features associated with UFP, consistent with increased lipid peroxidation, endogenous inhibitors of NO, and vehicle exhaust exposure biomarkers. Network and pathway analysis identified 38 altered pathways, including variations related to inflammation, endothelial function, and mitochondrial bioenergetics.
Cross-sectional study	Huang et al.[42]	41 Chinese elderly participants (patients with COPD: 23; healthy spouses: 18), nested in an existing cohort	41 urine samples	Short-term assessment on indoor $PM_{2.5}$	Key Laboratory of Urban Environment and Health, Chinese Academy of Sciences	Seven metabolites, primarily involved in purine, amino acid, and glycolysis metabolism, were significantly correlated to $PM_{2.5}$ exposure. Ten COPD-related metabolic features—indicating perturbation of amino acid, lipid and fatty acid, and glucose metabolism—were detected. $PM_{2.5}$ and its main species were associated with COPD-related features.

Table 1. (*Continued.*)

Type of study	Reference	Study design				Primary results
		Study population	Biometric	Air pollutant indicator	Laboratory information	
Cross-sectional study	Wu et al.[61]	40 workers (29 UPG workers vs. 20 non-UPG workers)	40 urine samples	Short-term assessment on PM_1, $PM_{2.5}$, PM_{10}, PAHs in $PM_{2.5}$, 16 unmetabolized PAHs in urine samples	Key Laboratory of the Three Gorges Reservoir Eco-environment, Chongqing University	The concentrations of unmetabolized 5–6 rings PAHs in the urine samples from UPG workers were significantly higher than in non-UPG workers. In total, 16 metabolites were different in UPG workers and which were involved in amino acid metabolism, nucleotide metabolism, lipid metabolism, carbohydrate metabolism, and metabolism of cofactors and vitamins. Particle-bound PAHs might induce glucose metabolism disorders through the hypoxia-inducible factor 1 signaling pathway.

Note: The table is ordered by study design and sample size. 8-OHdG, 8-oxo-2′-deoxyguanosine; 20-OH-LTB4, 20-Hydroxy-leukotriene B4; Ag, silver; Al, aluminum; ART, assisted reproductive technology; As, arsenic; Ba, barium; BAL, bronchioalveolar lavage; BALF, BAL fluid; BC, black carbon; BW, bronchial wash; Ca, calcium; Cd, cadmium; Ce, cerium; Co, cobalt; CO, carbon monoxide; COPD, chronic obstructive pulmonary disease; Cr, chromium; Cu, copper; EBC, exhaled breath condensate; EC, elemental carbon; Fe, iron; FEV1, forced expiratory volume in 1 second; FPMOP, fine particulate matter oxidative potential; GlcCer(d16:1/18:0), Glucosylceramide(d16:1/18:0); HAP, household air pollution; HILIC, hydrophilic interaction liquid chromatography; HR, heart rate; HRV, heart rate variability; ICA, independent component analysis; K, potassium; LDSA, lung deposited surface area; Mg, magnesium; Mn, manganese; Na, sodium; NAI, negative air ion; NARP, near-roadway air pollutants, represented by NO_x from freeway and non-freeway sources; Ni, nickel; NO, nitric oxide; NO_2, nitrogen dioxide; NO_x, nitrogen oxides; NRAP, near-road air pollution; O_3, ozone; OC, organic carbon; PAH, polycyclic aromatic hydrocarbon; Pb, lead; pb-PAH, particle-bound polycyclic aromatic hydrocarbon; $PM_{2.5}$, fine particulate matter; PM_{10}, coarse particulate matter; PNC, particle number concentration; S, sulfur; Sb, tin; Si, silicon; SM(d18:1/17:0), Sphingomyelin(d18:1/17:0); Ti, titanium; TRAP, traffic-related air pollution; UFP, ultrafine particles; UPG, underground parking garage; V, vanadium; WSOC, water-soluble organic carbon; Zn, zinc.

eligibility criteria and were included in this state-of-the-science review (Figure S1). Most articles (38/47) were recent publications written during the past 5 y (2018–2022).

Study Design and Sample Type

Multiple research study designs were used to apply untargeted metabolomics in air pollution research (Table 1). The largest number were crossover studies (15/47) and panel studies (15/47), followed by cross-sectional studies (12/47), cohort studies (4/47), and a case–control study (1/47). The crossover studies randomly allocated participants to study arms, where each arm consisted of a sequence of at least two exposure conditions or exposed participants to controlled exposure. The participants in 12 eligible studies were recruited from a general population, whereas another 12 studies focused on patients with a particular precondition of clinical concern or the susceptible populations included children, older adults, or pregnant people (Figure S2A). Most of the studies (35/47) had a relatively small sample size (i.e., <100 participants). Blood-based sample was the most commonly used biometric for air pollution metabolomics analysis, with 35 studies measuring blood-based metabolites (18 serum, 15 plasma, 2 whole blood), followed by 9 studies using urine, 3 using saliva, 2 using BALF, 1 using bronchial wash (BW), 1 using EBC, and 1 using stratum corneum (i.e., the outermost skin layer). Four studies used more than one biospecimen for metabolic profiling: 1 study collected BAL and BW samples[28]; 1 collected plasma, saliva, and EBC[20]; and 2 collected plasma and saliva samples.[24,59] Most of these studies focused on short-term (e.g., exposure windows from 2 h to 3 months) exposure to air pollution (42/47), 3 of which considered short- and long-term exposure concurrently. In these articles, the demographics of the study population varied widely by different life stages [(i.e., children, pregnant people, adults, and elderly) and different health status (i.e., healthy participants, participants with underlying conditions such as asthma, chronic obstructive pulmonary disease (COPD), cardio-cerebrovascular disease, or ischemic heart diseases].

Air Pollution Assessment

More than half of the eligible studies measured multiple air pollutants with PM, including its species, as the most common pollutant measured {coarse PM ($PM_{10-2.5}$) 11/47, fine PM [PM ≤2.5 μm in aerodynamic diameter ($PM_{2.5}$)] 38/47, ultrafine PM (UFP) 7/47, and carbonaceous species (e.g., organic carbon and elemental carbon) 17/47}. NO_x (16/47) and O_3 (11/47) were the two most commonly measured gaseous air pollutants among the eligible studies. As for the assessment methods, fixed-site monitoring (13/47) and personal monitoring using portable sensors (10/47) were the most widely used methods for collecting ambient, indoor, or traffic-related air pollution, followed by air quality modeling (9/47) and microenvironment monitoring (9/47) (Figure S2B).

Metabolic Profiling and Feature Extraction

Figure 1 shows the details of the experimental and statistical methods used in the eligible studies following a general workflow of the MWAS. Specifically, metabolic profiling and feature extraction is the first analytical step of the untargeted metabolomics application. Liquid chromatography (LC) was the predominant separation method in the eligible publications we reviewed (Figure 1B) and was used in 89% (42/47) studies. Among those studies, 24 used multiple chromatography columns, with 7 using both LC and gas chromatography (GC), and 17 using multiple LC columns, where C18 hydrophobic chromatography columns and hydrophilic interaction liquid chromatography (HILIC) columns were the predominant combination. Most of the studies

(31/47) used both positive and negative electrospray ionization (ESI) for detection in mass spectrometry (MS). In addition, 3 studies used H-NMR as the platform to conduct untargeted metabolomics. The majority of eligible studies conducted some forms of QA/QC on the metabolic profiling, with pooled samples added into each analytical batch as QC materials (Figure 1A). Laboratories conducting the metabolic profiling included Metabolon, Inc. (11/47); Emory Clinical Biomarkers Laboratory (9/47); Key Laboratory of Health Technology Assessment, Fudan University (5/47); International Agency for Research on Cancer (3/47); Key Laboratory of Urban Environmental and Health, Chinese Academy of Sciences (3/47); and others (16/47) (Figure 1E). The laboratory facilities extracted metabolic features differently from each other. A range of R packages, including apLCMS, xMSanalyzer, and XCMS, were employed to extract features from the metabolomic profiles, with different algorithms applied to conduct feature detection, quantification, feature alignment, batch effect adjustment, and noise removal (Figure 1D).

MWAS Statistical Approach

After conducting feature detection and extraction, the next step in untargeted metabolomics workflow is to conduct MWAS statistical analyses to identify metabolic features that are associated with the exposure of interest (e.g., air pollution exposures). Table 2 summarizes detailed information on the MWAS statistical modeling approaches applied in each untargeted metabolomics application in air pollution research. Consistent with the diversity of study designs available, a range of approaches using inferential statistics were employed. In general, most studies employed both univariable and multivariable analyses to identify significant metabolic features, with adjustment for potential confounders. Some researchers applied a multistage approach by combining both statistical modeling approaches and dimension reduction techniques rather than using one single approach for untargeted MWAS (23/47). For example, some studies first employed dimension reduction methods, such as partial least squares–discriminant analysis (PLS-DA), before conducting MWAS statistical modeling to identify the significant features associated with air pollution exposures. In this way, only metabolic features that had a variable importance in projection higher than a prescribed cutoff point could pass to the next stage of analysis, which was usually a regression model. Although the choice of inference statistics depended on study design, PLS-DA/orthogonal PLS-DA (OPLS-DA) was the most common dimension reduction approach used in eligible studies (18/47), whereas linear mixed-effect models and generalized linear models were the most used statistical modeling approaches (22/47) (Figure 1H).

Possibly owing to the nature of high dimensionality within metabolomics data, multiple correction tests were used in 31 of 47 studies and included such tests as the Bonferroni,[70] Benjamini-Hochberg,[71] and Storey-Tibshirani[72] approaches. However, likely owing to the limited statistical power resulting from a lack of exposure contrast or the small sample size, the significance cutoff for selecting individual features for downstream pathway enrichment analysis was usually selected based on raw *p*-values without adjusting for multiple testing in many studies (15/32).

Pathway Enrichment Analysis and Chemical Annotation and Confirmation

In untargeted metabolomics applications, pathway enrichment analysis is a key step in predicting the potential biological function and pathways of significant metabolic signals, which helps reduce the complexity of metabolomics data, improve interpretation of biological function, and generate hypotheses.[73] Many pathway enrichment methods have been developed for untargeted metabolomics applications. All of these methods, except for mummichog, use putatively annotated compound names or chemical identities as the input for the pathway enrichment analysis. In our opinion, extra caution should be taken when interpreting the enrichment results generated from the tools that only use *m/z* values or putative annotation as input, given that they do not require *a priori* knowledge of true signal identity and may result in an increased risk of false-positive discoveries. Of these publications, 34 studies included pathway enrichment analyses (32/34) or network analysis (2/34). MetaboAnalyst was the most used bioinformatic tool for pathway enrichment or network analysis (14/34), followed by mummichog (13/34). Of note, among these methods, mummichog is the only approach using *m/z* values as the input for the pathway analysis (Figure 1J).[74]

The last step of the MWAS typically involves chemical annotation and confirmation, where researchers aim to confirm the chemical identities of the significant features previously identified in the MWAS model and pathway enrichment analysis. Although the levels of confidence for reported metabolites vary between studies, 79% of the eligible studies confirmed metabolic features by tandem MS (MS/MS) and co-elution with authentic standards (i.e., level-1 evidence; $N = 27$) or by MS/MS and matches with online databases or *in silico*–predicted spectra (i.e., level-2 evidence; $N = 12$) based on the Metabolomics Standards Initiative (MSI) standards.[69] The Human Metabolome Database (HMDB) and METLIN were the two most commonly used publicly available online databases (Figure 1G).

Overview of the Annotated and Confirmed Metabolites Associated with Air Pollution

Eight hundred sixteen unique metabolites were identified as associated with air pollution exposures across the 47 included studies (Excel Table S2). Figure 2 summarizes the 35 unique metabolites reported repeatedly in >5 independent studies and which were confirmed with level-1 or -2 evidence. Of these metabolites, 20 were organic acids and derivatives, 5 were organoheterocyclic compounds (hypoxanthine, tryptophan, urate, allantoin, and uracil), 5 were lipids and lipid-like molecules (glycerol 3-phosphate, azelate, sebacate, linoleate, and docosahexaenoate), 2 were organic nitrogen compounds (carnitine and sphingosine), 1 was an organic oxygen compound (glycerate), 1 was a nucleotide (adenosine 5′-monophosphate), and 1 was a benzenoid (benzoate). Hypoxanthine was reported by 13 independent studies, followed by histidine, serine, and aspartate, which were reported by 9 independent studies. Ten metabolites—hypoxanthine, serine, aspartate, phenylalanine, arginine, azelate, threonine, glycerate, adenosine 5′-monophosphate, and lysine—were associated with six combined air pollutant groups.

We observed inconsistencies in the directions of these air pollutant–metabolite associations across the included studies that may be, in part, due to metabolomic differences by biospecimen type (Figure 2; Excel Table S3). Specifically, for the same associated air pollutant, the directions of the association of several metabolites were consistent in the same biospecimen but were inconsistent across different biospecimens. For example, lactate was found to be positively associated with total PM in serum, but in negative relation in plasma.[24,40,59,63] Creatine was inversely associated with O_3 in serum[26] but positively associated with O_3 in BALF.[19] High reproducibility was present between plasma with serum, BALF, and stratum corneum, but not with urine.

Table 2. Summary of the statistical modeling approaches, pathway analysis and feature annotation approaches of the reviewed studies on untargeted metabolomics application in air pollution health research.

	Statistical modeling approach					Pathway analysis		Feature annotation	
Reference	Model	Dimension reduction	Covariates	Interaction	MCT	Algorithm	Sig. feature criteria	Conf. level	Match database
Jeong et al.[43]	GLM	NA	SAPALDIA: age, sex, study area, bench time, fasting time, sine and cosine functions of venipuncture time with periods of 24 and 12 h, and their multiplicative interaction terms with fasting time	SAPALDIA: fasting time and an indicator for geocoding quality	BH	Mummichog	$p > 10$th percentile	1	Authentic chemical standard
			EPIC: age, center of recruitment, sex, BMI, smoking status, education level					4	Mummichog
Menni et al.[25]	GLM	NA	Age, sex, BMI, height, metabolite batch, family relatedness	NA	Bonferroni	NA	NA	2	Metabolon's library
Chen et al.[41]	Student's *t*-test	NA	NA	NA	Bonferroni	MetaboAnalyst	$q < 0.05$	3	HMDB, KEGG, ChEBI
Huang et al.[42]	Mann-Whitney test, Partial correlation analysis	PCA, PLS-DA	Age, sex, BMI, COPD, past smoking, alcohol drinking	NA	NA	NA	NA	2	HMDB
Li et al.[38]	Tobit, MLR	NA	Year of visit, season, weekday, apparent temp	NA	BH	Mummichog	$p < 0.05$	1	Authentic chemical standard
Walker et al.[31]	Correlation analysis	PLS-DA	NA	NA	BH	Mummichog	$q \leq 5$th percentile and Pearson $\lvert r \rvert \geq 0.6$	1	Authentic chemical standard
								3	HMDB
								4	
Yan et al.[34]	NA	PLS-DA	Maternal age, maternal race/ethnicity, maternal education	NA	NA	Mummichog	$p < 0.05$	1	Authentic chemical standard
								4	HMDB, KEGG, LipidMaps
Loo et al.[36]	MLR	OPLS-DA	Age, BMI, sodium intake, physical activity	NA	NA	NA	$p < 0.05$ and VIP > 1	1	Authentic chemical standard
								4	HMDB, METLIN
Nassan et al.[40]	LMM	ICA	PM mass, NO_2, O_3, temp, RH, age, BMI, cigarette pack-years, alcohol intake, SES, season	NA	ENT	MetaboAnalyst	$p < 0.01$	1	Metabolon database
Nassan et al.[63]	LMM	ICA	Temp, RH, age, BMI, cigarette pack-years, alcohol intake, SES, season	Type II diabetes, obesity	ENT	MetaboAnalyst	$p < 0.01$	1	Metabolon database
Miller et al.[26]	ANOVA	NA	NA	NA	NA	MetaboLync pathway analysis	$p < 0.05$	1	Metabolon database
Surowiec et al.[28]	Student's *t*-test, Wilcoxon's *t*-test	PCA, OPLS-DA	NA	NA	NA	MetaboAnalyst	$p < 0.05$	1	In-house spectra library, Chenomx
Li et al.[22]	LMM	PCA, OPLS-DA	Age, sex, BMI, random intercept for each participant, temp, RH	NA	NA	NA	NA	3	NIST 11 standard mass spectral database, Fiehn database, HMDB, METLIN
Vlaanderen et al.[30]	LMM	NA	Age, BMI, sex, temp, RH, season, peak areas measured at baseline, random intercept and slope	NA	Storey-Tibshirani	Mummichog	Combined criteria	1	HMDB, METLIN, mzCloud
Cheng et al.[19]	Paired *t*-test	PLS-DA	NA	NA	NA	MetaboAnalyst Network analysis	$p < 0.05$	1	Metabolon database

Table 2. (*Continued.*)

	Statistical modeling approach					Pathway analysis		Feature annotation	
Reference	Model	Dimension reduction	Covariates	Interaction	MCT	Algorithm	Sig. feature criteria	Conf. level	Match database
Ladva et al.[21]	LMM	NA	Asthma status, age, sex, BMI, race	NA	BH	Mummichog	$q < 0.05$	4	Mummichog
C. Chen et al.[18]	LMM	PCA, PLS-DA	Age, sex, BMI, temperature, relative humidity, random intercept	Sex, temperature	BH	NA	NA	4	HMDB, METLIN
Liang et al.[23]	LMM	NA	Asthma status, weekday, age, sex, race, BMI, baseline feature intensity	Asthma status	BH	Mummichog	$q < 0.05$	1 4	Authentic chemical standard METLIN, ChemSpider, HMDB, KEGG
van Veldhoven et al.[29]	Oxford Street II: GLSMM	NA	Age, sex, BMI, health status, caffeine intake	NA	Bonferroni	NA	NA	1	Authentic chemical standard
	TAPAS II: GLSMM		Age, sex, BMI, physical activity	Physical activity				4	HMDB, METLIN
Zhang et al.[35]	Student's *t*-test, LMM	PCA	Age, BMI, noise, temp, RH, number of measured days, random intercept	Sex	NA	NA	NA	2	HMDB, METLIN, Chemspider, *in silico*
Liu et al.[37]	LMM	PLS-DA	Age, sex, BMI, class, day of measure, temp, RH, noise	NA	NA	NA	NA	2 4	HMDB
Wei et al.[32]	LM, LMM	NA	Age, medication use, (a random intercept of exposure for each study for LMM)	NA	BH	PCA	NA	1	Metabolon database
Baker et al.[17]	*t*-test	NA	NA	NA	BH	NA	NA	5	NA
Ladva et al.[20]	Correlation analysis, MLR	NA	Subject effects, sampling time effects	NA	NA	NA	NA	4	HMDB, ChemSpider
Liang et al.[24]	LMM	NA	Dorm location, age, sex, BMI, race, move-in days, time point	NA	BH	Mummichog	Adjust $p < 0.05$	1 4	Authentic chemical standard METLIN, ChemSpider, HMDB, KEGG, U.S. EPA
Xia et al.[33]	Paired *t*-test, LMM	PCA, OPLS-DA	Sampling day, temp, RH, age, sex, BMI, allergic status, random-effect intercept	NA	NA	NA	NA	4	HMDB, METLIN
Mu et al.[27]	ANOVA	NA	Time point, sex, BMI, vegetable and fruit intake, transportation mode	NA	BH	Network analysis, module analysis	$q < 0.05$	1	Metabolon database
Chu et al.[44]	LMM, ANOVA	PCA, OPLS-DA	Temp, RH, age, sex, BMI, random intercept	Cardiopulmonary function	BH	NA	$q < 0.05$ and VIP > 1	1 2	Authentic chemical standard HMDB, METLIN, LipidMaps
Cruz et al.[46]	NA	PCA, PLS-DA	NA	NA	NA	MetaboAnalyst	VIP > 1	2	Chenomx
Cruz et al.[45]	NA	PCA, PLS-DA	NA	NA	NA	MetaboAnalyst	VIP > 1	2	Chenomx
Du et al.[47]	LMM	OPLS-DA	Age, sex, BMI, temp, RH, random intercept	Sex	BH	Ingenuity pathway analysis	$q < 0.05$	4	NIST 11 standard mass spectral database, Fiehn database, HMDB
Feng et al.[48]	Student's *t*-test	OPLS-DA	NA	NA	FDR	MetaboAnalyst	$q < 0.05$ and VIP > 1	1 2	In-house spectra library HMDB, METLIN
Gao et al.[49]	ANOVA	PCA	Circadian rhythm	Smoking status, BMI	BH	NA	NA	1	Metabolon database

Table 2. (*Continued.*)

	Statistical modeling approach					Pathway analysis		Feature annotation	
Reference	Model	Dimension reduction	Covariates	Interaction	MCT	Algorithm	Sig. feature criteria	Conf. level	Match database
Gaskins et al.[50]	MLR	NA	Age, BMI, smoking status, education, temp	NA	BH	Mummichog	$p < 0.05$ and $q < 0.05$	1 4	In-house spectra library METLIN, ChemSpider, HMDB, KEGG
Hood et al.[51]	MLR	NA	Temp, age, BMI, education, smoking status	NA	BH	Mummichog	$p < 0.005$	1	In-house spectra library
Huan et al.[52]	LMM, GLM	PCA, OPLS-DA	Age, BMI, temp, air pressure, random intercept	Sex	BH	NA	NA	2	HMDB, METLIN, KEGG
Liao et al.[53]	MLR	Network analysis, sPLS	Sex, age, BMI, race/ethnicity, smoking history, secondhand smoking, asthma medication use	Asthma control status, sex, ethnicity, BMI	BH	MetaboAnalyst	$p < 0.05$ and correlations ≥ 0.65	1	Metabolon database
Zhao et al.[62]	LMM	NA	Age, sex, BMI, annual income, education, cooking, and drinking habits, blood cotinine level, weekday of visit, temp, RH	Sex	BH	MetaboAnalyst	$q < 0.05$	2	HMDB, Discovery HD4 Metabolomics platform
Walker et al.[60]	LMM	NA	Age, day of visit, BMI, sampling time	NA	BH	Mummichog	$q < 0.2$	1 4	In-house spectra library HMDB
Selley et al.[58]	LMM	NA	Vectors of pharmaceutical signals, temp, RH, secondary pollutants, urinary concentrations of acetaminophen and ibuprofen	NA	NA	NA	NA	2	Chenomx profiling, HMDB
Ritz et al.[57]	MLR	PLS	Maternal age, maternal race, birth year, preterm birth, parity, and neighborhood SES	NA	NA	Mummichog	$p < 0.05$ and VIP ≥ 2	1 4	In-house spectra library HMDB
Nassan et al.[39]	LMM	ICA	Coexposures, age, BMI, SES, cigarette pack-years, alcohol intake, season at blood draw, RH	Type II diabetes, obesity	ENT	MetaboAnalyst	$p < 0.01$	1	Metabolon database
Liu et al.[54]	LMM	PCA, OPLS-DA	Age, sex, BMI, random intercept	NA	BH	MetaboAnalyst	VIP > 1 and $q < 0.05$	4	HMDB, LipidMaps
Liu et al.[55]	LMM, Mann-Whitney test	PLS	Age, sex, BMI, class, long-term time trend, coexposure, temp, RH, noise, random intercept	Sex, BMI	FDR	MetaboAnalyst	$p < 0.05$, $q < 0.05$, and VIP > 2	2	HMDB
Misra et al.[56]	Student's *t*-test, sCCA, v-test	NA	NA	NA	BH	NA	$q < 0.05$	1	Metabolon database
Tang et al.[59]	LMM	NA	Dorm location, age, sex, BMI, race, move-in days, time point	NA	BH	Mummichog	$q < 0.2$	1 4	In-house spectra library HMDB, KEGG, ChemSpider, METLIN
Wu et al.[61]	Student's *t*-test	PCA, OPLS-DA	NA	NA	NA	NA	NA	NA	NA

Note: The confidence levels refer to the communication of identification confidence based on the Metabolomics Standards Initiative (MSI) standards.[69] Briefly, these included identified metabolites (level 1), putatively annotated compounds (level 2), putatively characterized compound classes (level 3), and unknown compounds (level 4). Adjust, adjusted; ANOVA, analysis of variance; BH, Benjamini-Hochberg procedure; BMI, body mass index; ChEBI, Chemical Entities of Biological Interest; conf, confidence; COPD, chronic obstructive pulmonary disease; DBS, dried blood spot; Dorm, dormitory; ENT, number of effective/independent tests; EPA, Environmental Protection Agency; EPIC, European Prospective Investigation into Cancer and Nutrition; FDR, false discovery rate; GLM, generalized linear model; HMDB, Human Metabolome Database; ICA, independent component analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LMM, linear mixed-effect model; MCT, multiple comparison test; MLR, multiple linear regression; NA, not available; NIST, National Institute of Standards and Technology (database); OPLS-DA, orthogonal partial least squares–discriminant analysis; PCA, principal component analysis; PLS-DA, partial least squares–discriminant analysis; PM, particulate matter; RH, relative humidity; SAPALDIA, Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults; sCCA, sparse Canonical Correlation Analysis; SES, socioeconomic status; sig, significant; temp, temperature.

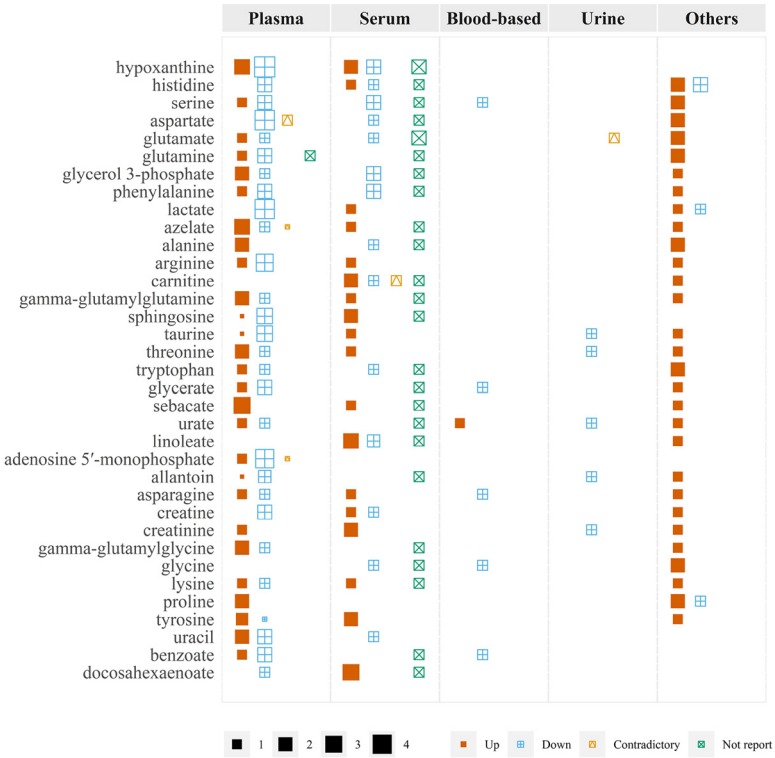

Figure 2. Annotated metabolites (level 1 or 2) associated with the exposure to air pollution reported in >5 independent studies of 37 studies. The *x*-axis represents metabolites ordered by the number of studies that reported the metabolites (from top to bottom). The data can be found in Excel Table S3. Thirty-seven studies reported metabolites with level-1 or -2 evidence (47 studies were included in this review). The confidence levels refer to the communication of identification confidence based on the Metabolomics Standards Initiative (MSI) standards.[70] Briefly, these included identified metabolites (level 1), putatively annotated compounds (level 2), putatively characterized compound classes (level 3), and unknown compounds (level 4). The "blood-based" category includes fasting blood and dried blood spots. The "others" category includes saliva, bronchial wash, bronchoalveolar lavage fluid, and stratum corneum. The size of the point corresponds to the number of studies that reported the metabolite. "Up" means that the metabolite's relative concentration is positively associated with air pollution exposure level. "Down" means that the metabolite's relative concentration is negatively associated with air pollution exposure level. "Contradictory" means that the study reported contradictory changes in the metabolite with air pollution exposure level. "Not report" means that the study did not report how the metabolite changed with air pollution exposure level.

Overview of the Biological Pathways Associated with Air Pollution

Figure 3 showed biological pathways reported in eligible studies, and biological pathways with more than five associations reported were included. There were 32 eligible studies reporting biological pathways enriched by metabolic features that were associated with an air pollutant exposure (Figure 3). The top five biological pathways consistently associated with the air pollutant categories and reported in majority of the eligible studies were glycerophospholipid metabolism, pyrimidine metabolism, methionine and cysteine metabolism, tyrosine metabolism, and tryptophan metabolism. To test if a particular laboratory or use of a certain pathway enrichment algorithm would have a predominant impact on the pathway analysis results, we further stratified the biological pathways by laboratory/institute and by pathway analysis algorithm. As shown in Figure S3A, the rank of the top five biological pathways remained the same by laboratory/institute.

As expected, because MetaboAnalyst and mummichog were the most widely used pathway analysis packages, most biological pathways were detected using these two algorithms (Figure S3B). Nevertheless, several pathways were consistently detected across more than two pathway analysis algorithms and included, for example, glycerophospholipid metabolism, tyrosine metabolism, and tryptophan metabolism. In addition, we found that metabolic features detected in different biospecimens were reported to contribute to distinctly different expressed biological pathways (Figure S4A–C). The top five biological pathways detected in plasma, for example, were different from those detected in serum, with only two overlapping pathways (i.e., the pyrimidine metabolism and the methionine and cysteine metabolism). In addition, the top five biological pathways detected in saliva were distinct from those detected in blood and included vitamin E metabolism, vitamin B9 metabolism, saturated fatty acids beta-oxidation, polyunsaturated fatty acid biosynthesis, and omega-3 fatty acid metabolism. Only two eligible studies used saliva samples and reported the associated biological pathways.[24,59]

Discussion

Over the past decade, considerable progress has been made in high-throughput omics technologies, including the development of untargeted metabolomics approaches, using high-resolution analytical platforms. With the accompanying technological advancements, untargeted HRM has emerged as a promising tool to improve internal exposure estimation to complex environmental mixtures along with their corresponding biological response. As highlighted in the present state-of-the-science review, although environmental applications of the field are methodologically inconsistent, there is growing evidence that exposure to air pollution is linked to changes in the human metabolome, as reported in a considerable number of recent studies (Table 1). Despite the continuing interest in using untargeted HRM in air pollution health research, we identified gaps, inconsistencies, and uncertainties in

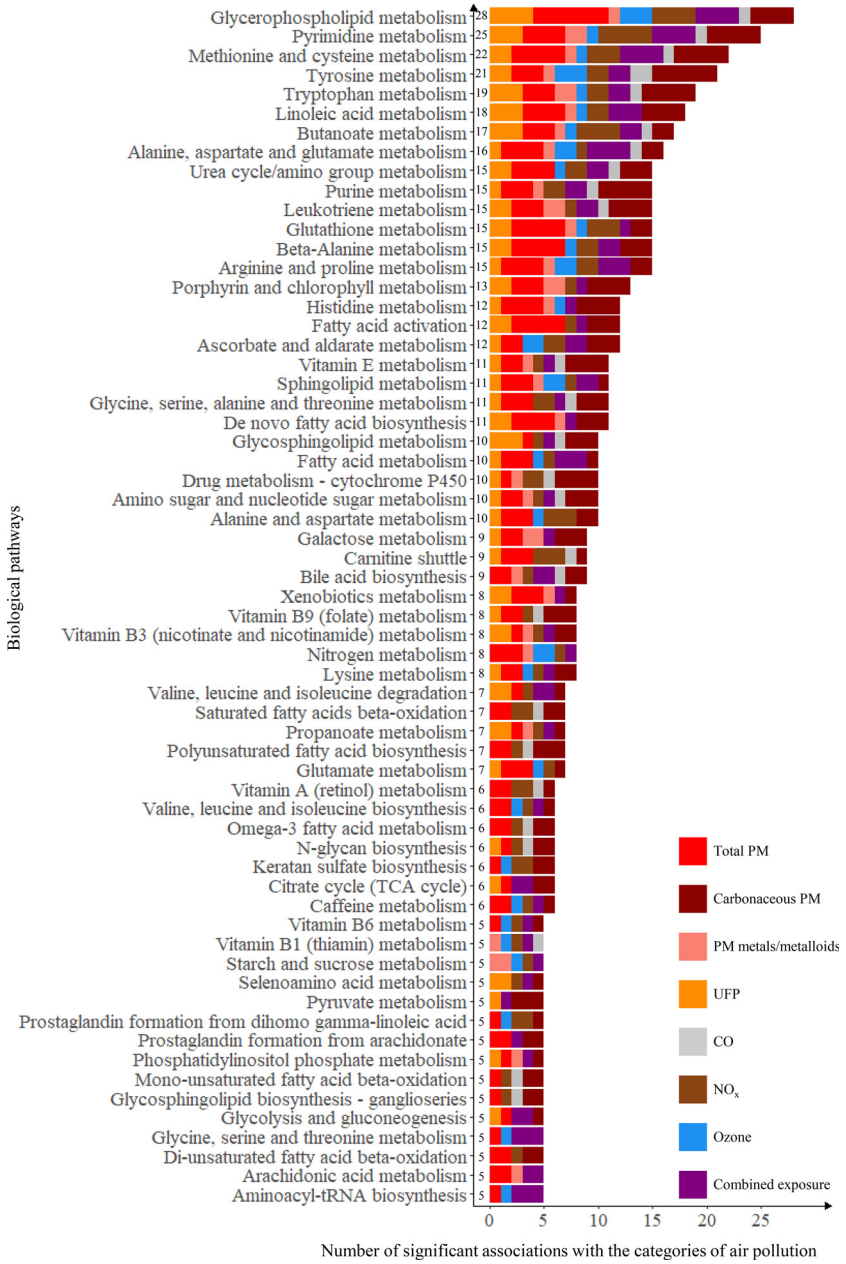

Figure 3. Biological pathways significantly associated with the exposure to air pollution reported in all eligible studies ($N = 32$). The x-axis represents the number of associations between the biological pathway and the category of air pollution, and the same pair of pathway and category was counted multiple times when the association was reported by different studies. Biological pathways with less than five associations reported were excluded. The data can be found in Excel Table S4. Note: CO, carbon monoxide; NO_x, nitrogen oxides; PM, particulate matter; TCA, tricarboxylic acid (cycle); UFP, ultrafine particle.

the existing literature, raising concerns related to the coherence and generalizability of findings across different air pollution metabolomics studies. By carefully comparing methodologies and thoroughly synthesizing findings, our aim was to conduct this first, to our knowledge, state-of-the-science review on the use of untargeted HRM in air pollution research, with the goal of providing a necessary and timely guide into the existing evidence, gaps, and recommendations for possible future next steps. Compared with the previous review on applying metabolomics to study polycyclic aromatic hydrocarbons (PAHs),[75] we cover a much more extensive range of air pollutants. Notably, we specifically limited this review to air pollution health effects studies employing untargeted metabolomic workflows, with a goal of elucidating whether these specific studies were consistent with or added to the considerable body of literature involving the use and identification of a limited suite of targeted metabolites.

Broadly, we observed some general consistencies in the results and findings across the 47 reviewed studies and the 14 HRM laboratories. Starting with the pathway enrichment analysis, at least 5 of 14 laboratories reported perturbations in eight common metabolic pathways, most of which were detected by at least two different algorithms. Specifically, multiple pathways—including glycerophospholipid metabolism, pyrimidine metabolism, methionine and cysteine metabolism, tyrosine metabolism, and tryptophan metabolism—were consistently detected in different studies using various laboratory platforms and bioinformatic tools (Figure S3A,B). A similar degree of consistency was observed among the reported metabolites associated with air

pollution exposures, which were all confirmed with level-1 and -2 evidence. Specifically, 35 metabolites—including, for example, hypoxanthine, histidine, serine, aspartate, glutamate, glutamine, glycerol 3-phosohate, phenylalanine, lactate, and azelate—were repeatedly shown to be associated with air pollution exposures in >5 independent studies. Furthermore, the coherence of repeatable pathways and metabolites provided additional confidence that the collective findings were not random. Specifically, hypoxanthine, adenosine 5′-monophosphate, glycine, and urate were involved in purine metabolism; creatine, serine, threonine, alanine, aspartate, glycine, glycerate, and tryptophan were involved in glycine, serine, alanine, and threonine metabolism; serine and glycerol 3-phosphate were involved in glycerophospholipid metabolism (Figures 2 and 3). Interestingly, most of these reproducible metabolic pathways and metabolites were those closely linked to acute inflammation and oxidative stress, the commonly accepted mechanistic pathway associated with air pollution exposure.[76–79]

Unlike untargeted HRM, few targeted metabolomics analyses have been conducted in human air pollution health research. Although a formal comparison of results between these untargeted air pollution metabolomics with other targeted analyses is beyond the scope of this review, it is worth highlighting several consistent findings with a recent targeted quantitative analysis examining metabolic response from exposure to air pollution, which suggested that air pollution exposure leads to acute perturbation in amino acid metabolism profile in plasma.[80] Consistent with summaries in our present state-of-the-science review, *a*) increases in short-term O_3 exposure were found to be significantly associated with increases in taurine and aspartate, *b*) increases in $PM_{2.5}$ exposure were statistically associated with decreases in taurine and aspartate, and *c*) increases in NO_2 exposure were significantly associated with decreases in phenylalanine. We also observed some minor inconsistencies between this targeted study[80] and our review on the untargeted HRM applications. Specifically, in this targeted analysis, short-term O_3 exposure but not $PM_{2.5}$ exposure was found to perturb urea cycle activity. Conversely, urea cycle/amino acid metabolism was a significant, repeatable pathway found to be associated with several air pollutants, including total PM and O_3 in our present review on the untargeted metabolomics studies. Nevertheless, an important finding from the targeted analysis result was that the effects of air pollution on amino acid metabolism profile varied by exposure durations and air pollutant, which also aligns well with our initial observations from the untargeted metabolomics analysis. The observed good concordance between the targeted and untargeted metabolomics analyses also necessitates the need to combine both the targeted and untargeted metabolomics approaches in future air pollution health studies. Indeed, as we see it, bridging both targeted and untargeted approaches can lead to novel hybrid approaches to enable full characterization of the metabolome and further expand the great potential of metabolomic studies in environmental risk assessment and health research.

Despite the overall consistencies, which suggest an encouraging potential for air pollution metabolomic applications, we also identified several major inconsistencies and critical questions inherent in the existing findings. Next, we discuss challenges and offer 10 recommendations for possible future steps.

1. **Need to standardize protocols.** Generally, most of the inconsistencies in the findings we reviewed likely resulted from the vast heterogeneities in analytical platforms, statistical approaches, pathway enrichment analysis, and chemical annotation and confirmation. Indeed, given that environmental metabolomics is still in its infancy, there is no established protocol to unify each key step in the air pollution-HRM application, which hinders the comparison, integration, and generalization of findings across different studies. The sources of heterogeneity remain critical factors to consider when interpreting the pooled results. For example, according to our previous work, the metabolic features and biological pathways associated with air pollution largely varied by the choice of chromatography columns (i.e., hydrophilic vs. lipophilic columns).[24,38] As shown in Figure S3B, the detection of biological pathways was impacted by the reference database used for the pathway analysis algorithms, which possibly resulted in several biological pathways being solely reported by mummichog.

   To address these sources of heterogeneity, we recommend the development and implementation of a guideline that provides key considerations in study design and minimum standards in study reporting, which would ultimately improve the quality of research findings and study reproducibility and facilitate evidence comparing across environmental epidemiological studies using an untargeted metabolomics approach. It is important to note that using identical procedures across studies also has limitations, given that the specific goals, design, and context of each study all vary. As such, a particular workflow that may be suitable for a specific study design/population may not be appropriate for another. For example, in a crossover design where participants are randomly exposed to two or more exposure conditions in a sequential order and serve as their own control, univariate and dimension reduction approaches are often used as the main statistical methods. In observational studies, statistical approaches that can rigorously adjust for confounders are more desirable. In general, analytical considerations may differ depending on study design, exposure assessment strategy, study population, and analytical platform, and the challenges are multifaceted and include QA/QC approaches, data preprocessing (software for feature extraction, batch effect correction, feature filtering criteria, missing imputation, and data normalization), statistical approaches (choices of model types, choices of dimension reduction techniques, selection of confounders, and correction for multiple comparison), pathway and network analysis, metabolite identification and annotation, and reporting standards. These are all important aspects to be considered when developing the harmonizing operational guidelines for metabolomics applications in environmental epidemiological studies. Currently, efforts are underway to overcome the complexity and lack of standard protocols for metabolomics application in the broader field of epidemiological studies.[81–84]
2. **Time window of exposure and effects.** Although untargeted metabolomics was capable of characterizing both acute and chronic impact of exposures on metabolic perturbations,[85] short- and long-term changes could reveal different metabolic patterns associated with air pollution exposures. For example, metabolites with a short half-life may only reflect transient changes, whereas more stable metabolites can reflect a long-lasting effect. As a result, varying patterns and effect sizes may be observed on the same metabolite–air pollutant association when comparing studies investigating long-term impact to those studying short-term effects. From our perspective, future study using kinetic models will help better understand the observed metabolic responses in observational literature and complex dynamics underlying these reactions.

Furthermore, in our opinion several factors need to be carefully considered when investigating the impact of air pollution exposures on human metabolome. For studies that focus on short-term impact, additional exposure factors might contribute substantially to the short-term variability of blood metabolome. These factors include fasting status, dietary intake and nutrition, diurnal pattern of the metabolic variation, and meteorological factors.[86] These factors may, in turn, act as confounders for the association between air pollution and metabolic changes. As for long-term exposure, time-independent factors, such as age, sex, race, socioeconomic status, and meteorological factors (e.g., temperature, relative humidity) might need to be considered, any of which could impact an individual's metabolic health status. Notably, most of previous studies focused on the effect of individual air pollutant on human metabolome. In reality, air pollution is a complex entirety,[87] and its composition can be altered by factors such as photochemical reactions.[88] Instead of simply teasing out the individual effect, future research should take the integrity of air pollutants into account.

3. **Multipollutant analyses.** Because humans are exposed to the mixture of air pollutants simultaneously, employing a multipollutant approach in air pollution research has been encouraged with the goal of characterizing more fully the complexity of the health impacts of air pollution and identifying the most harmful emission sources.[89] However, like most existing air pollution health research, the majority of the existing air pollution metabolomics analysis employ only single-pollutant approaches or consider the air mixture as a single predictor. Although most studies investigated the impact of multiple air pollutants on the human metabolome, only four studies used a multipollutant approach with consideration of the correlated nature of air pollutants, which included a main exposure of interests and adjusted for exposures to its constituents or secondary pollutants.[39,40,58,63] Currently, the use of a multipollutant statistical framework remains difficult to apply in high multidimensional data settings, including HRM. From our perspective, future advancements in statistical algorithms and mixture analysis methods may facilitate greater use of a multipollutant analyses approach in air pollution metabolomics research.

   In addition, heterogeneity existed when characterizing air pollution exposures as continuous or categorical variables. Among the eligible studies that used PLS-DA, the exposure was often dichotomized to fit better into analyzing the high-throughput data. However, this would introduce several potential problems—including loss of statistical power, misclassification, and obscuration of nonlinearity—in the relation between the exposure and outcome,[90] given that air pollution has no health threshold.

4. **Study population and susceptibility.** We found that several studies recruited study populations with wide age ranges and that some eligible studies did not clearly specify and define the study population. These factors would make the altered metabolic features and biological pathways less generalizable and less comparable across studies, as demographics and underlying health conditions can contribute substantially to heterogeneities across studies. Patel et al. revealed that significant race and sex differences existed in small-molecule metabolites and metabolic hormones among overweight/obese adults,[91] and Johnson et al. demonstrated that circulating plasma metabolomic profiles were associated with the rate of biological aging.[92] Working with participants with and without asthma, we also found asthmatic status could modify the patterns and magnitude of metabolic perturbations associated with traffic-related air pollution exposures.[23] In sum, key demographic characteristics, including age, race, and sex, and underlying health conditions all potentially play critical roles in modifying the metabolic and health responses to air pollution exposure and thus need to be considered and adjusted in the metabolomics analysis.

5. **Sample size and statistical power.** Given that the environmental metabolomics application only started to emerge in recent years, many of the existing analyses were designed as pilot studies or proof of concept research efforts aimed to demonstrate the feasibility of metabolomics in the field of air pollution health research. As highlighted in the present review, most studies had relatively small sample sizes, ranging from a handful to <100 participants. Owing to the limited statistical power resulting from the small sample size, many of the current studies failed to find significant signals after adjusting for multiple testing correction.[21,26,53] Consequentially, the application of untargeted metabolomics was inevitably vulnerable to a high risk of false-negative and false-positive findings, which limits the potential to discover reliable biomarkers and remains a major concern in the current field of environmental metabolomics. From our perspective, expanding the sample size will be the most straightforward and effective way to obtain sufficient statistical power to identify robust and meaningful metabolic signals associated with air pollution, as demonstrated in the field of genomics and epigenomics. Budget constraints and limited resources and infrastructure, however, are still common concerns shared by many research groups. Future advancement in large-scale metabolic profiling and standardization in analytical protocols may solve this bottleneck.

6. **Metabolic profiling approach.** LC/MS remains the most used analytical platform for metabolic profiling among the existing studies. The heavy reliance on a single metabolic profiling approach may, however, fail to capture important metabolic signals during the profiling. For instance, many air pollutants are hydrophobic and semivolatile and do not ionize well with popular LC/MS methods, whereas GC/MS can sensitively capture these exogenous signals.[93] Hence, a minimum sufficient set of multiple chromatography columns may help separate the complex mixture of biospecimens and enhance the coverage of feature detection and extraction in the future. In addition, three eligible studies used another most common technique in metabolomics studies, H-NMR.[45,46,58] Unlike MS spectrometry, H-NMR requires minimal sample preparation and tissue extraction and has a higher degree of reproducibility.[94] Although the combined application of two techniques can increase the metabolome coverage, this may not be practical due to the limited funding and quantity of samples.

7. **Pathway enrichment analysis.** Pathway enrichment analysis is an important step of air pollution metabolomics, commonly used for the functional interpretation of metabolomics data sets. As shown in this review, most studies performed pathway enrichment analysis using ad hoc software tools, which map significant metabolites to known biochemical pathways on the basis of the information contained in public databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG). Although providing an insightful view of the biological pathways and

interactions involved, the pathway enrichment analysis is inherent to risks of false-positive discoveries, resulting from the incompleteness of databases and the uncertainty of metabolite identification (i.e., sole *m/z* input), especially for those methods using the *m/z* value as input without *a priori* knowledge of feature identity. As such, the predominant reliance of the existing studies on a single ad hoc tool (e.g., mummichog or MetaboAnalyst) is problematic. Moreover, many of the consistent metabolic pathways found across different air pollution metabolomics studies, whereas biologically plausible, do not appear to be air pollutant specific, given that they were also commonly found in other environmental metabolomics applications on non–air pollution exposures.[95–97] Expanding the metabolic pathway database to include list of air toxicants and metabolites is also critical because most of the existing pathway enrichment tools use databases that focus only on endogenous physiological processes, with little information available on endogenous environmental contaminants.

8. **Chemical confirmation and reporting standards.** As the crucial link between untargeted metabolomics data and meaningful biological interpretation, chemical annotation and confirmation remain the most prominent bottleneck in the metabolomics application, as demonstrated by the very limited number of metabolites validated with high-level annotation confidence (i.e., level-1 or -2 MSI evidence) in each study included in this review. Specifically, of the tens of thousands of features detected in metabolomics profiling, only a tiny fraction of metabolites was annotated and eventually confirmed. Slightly more than half of the eligible studies were able to confirm metabolite with level-1 evidence by using MS/MS and co-elution with authentic standards (27/47). Moreover, the number of features that can be validated with level-1 evidence is limited by the number of chemical standards available in the internal library, which also varies across different analytical platforms and laboratories. Furthermore, a vast majority of chemicals do not have commercial standards available, and the synthesis of chemical standards is both time consuming and costly, adding tremendous challenges to the metabolite confirmation processes. Meanwhile, despite the advancement in development of bioinformatics tools for chemical annotation, most of these tools still heavily rely on searching the experimental MS 1 or MS/MS data through publicly available databases [e.g., ChemSpider (http://www.chemspider.com),[98] METLIN (https://metlin.scripps.edu),[99] HMDB (https://hmdb.ca),[100] MassBank (https://massbank.eu/MassBank),[101] mzCloud (https://www.mzcloud.org)[102]]. These public databases focus mainly on endogenous metabolites, with very few exogenous factors included.[103] These preliminary identifications, built on matching precursor *m/z* to a metabolite database (level 4) would inevitably suffer from increased risks of false positives and false negatives as a result of low-quality spectra and incomplete databases.[104] Finally, although the untargeted metabolomics is capable of detecting novel metabolic signals associated with environmental exposures, the majority of the existing chemical annotation and confirmation approaches are heavily reliant on existing databases, limiting the ability to unravel the chemical identity of the real "unknown" unknowns.[103,105,106] Hopefully, these challenges will be overcome with the future development and refinement of analytical platforms, experimental MS/MS data, and bioinformatics algorithms. More importantly, standardizing and leveraging methodologies, and sharing data and knowledge will all greatly benefit the entire community.
9. **Validation.** Validation is a critical step for biomarker discovery, development, and translation in air pollution metabolomics. Owing to the hypothesis-generating nature of most untargeted metabolomics applications, the findings on any significant metabolite–air pollutant association require replication and validation in future hypothesis-testing settings. Specifically, replicating air pollution metabolomics study findings in a targeted manner in different populations will be critical for ultimately translating the results into clinical applications and in understanding in detail the exposure–response associations.[107] Conducting pooled and meta-analysis across multiple cohorts can serve as an alternative, efficient means of external validation. However, such analysis is challenging to conduct given the discrepancies in analytical protocols among the existing studies. Based on the molecular links found in the air pollution metabolomics, using *in vitro* and animal-based studies will help validate the findings and further elucidate the underlying mechanisms of air pollution toxicity on human metabolome and disease etiology.[108] Future air pollution metabolomics applications can also consider using a more comprehensive multistage metabolomics analysis strategy (e.g., using untargeted metabolomics for discovery phase, followed by a confirmatory step using targeted metabolomics) to move from semiquantitative measurements toward fully quantitative measurements of identified markers.[107] In addition, given that most of the existing studies are observational, there is a need for theoretical and experimental validation approaches, which will undoubtedly assist in validation and strengthening causal inference.
10. **Quantification.** Currently, it remains unfeasible to use untargeted metabolomics to quantify metabolites owing to the technical bottleneck, and only targeted metabolomics is capable of conducting an absolute quantification of endogenous and exogenous metabolites by comparison with calibration curves for the predetermined compounds.[109] Therefore, untargeted metabolomics is considered a hypothesis-generating process, in which the detected significant metabolites need validation and absolute quantification in targeted analysis to be used for interlaboratory comparison and establishment of diagnostic standards.[65,66] In the eligible studies, internal standards and QC samples were employed to perform relative quantification of metabolic features, but differences in experimental design, such as sample extraction and LC-MS conditions, made the comparison across laboratories almost impossible, compromising the feasibility of using untargeted metabolomics for metabolite quantification.[65,109]

Conclusions

The current studies demonstrate a core feasibility for using untargeted HRM as a platform linking air pollution exposure to internal dose and biological response. Our review of the existing untargeted HRM–air pollution findings points to an underlying coherence and consistency across studies, using a range of sample analytical quantitation methods, extraction algorithms, and MWAS modeling approaches. Metabolic perturbations involved in oxidative stress, acute inflammation, and nucleic acid damage and repair were associated with both short- and long-term air pollution exposure across multiple studies and included altered glycerophospholipid metabolism, pyrimidine metabolism, methionine and cysteine metabolism, tyrosine metabolism, and tryptophan metabolism, as well as changes in hypoxanthine, histidine, glutamate, arginine, and tryptophan. Future directions should focus on validating these findings via hypothesis-driven protocols and technical advances in metabolic annotation and quantification.

Acknowledgments

We thank D. Walker and L. Mu for providing additional information on their individual air pollution metabolomics study. We also thank all members from the Environmental Metabolomics and Exposomics Research Group at Emory (EMERGE) for their valuable input and feedback on this project.

Support for this project was provided by the National Institutes of Health (NIH) research grant (R21ES032117, to D.L.). We also acknowledge the tremendous support from the HERCULES Exposome Research Center, supported by the National Institute of Environmental Health Sciences of the NIH (P30ES019776, to D.L., D.J., and J.S.). R.V. and J.V. received funding from EXPOSOME-NL (the Netherlands Organization for Scientific Research; project 024.004.017, to R.V. and J.V.) and EXPANSE (EU-H2020 grant 874627, to R.V. and J.V.).

References

1. World Health Organization. 2022. Air pollution data portal. https://www.who.int/data/gho/data/themes/air-pollution [accessed 18 February 2022].
2. Bourdrel T, Bind MA, Béjot Y, Morel O, Argacha JF. 2017. Cardiovascular effects of air pollution. Arch Cardiovasc Dis 110(11):634–642, PMID: 28735838, https://doi.org/10.1016/j.acvd.2017.05.003.
3. Guarnieri M, Balmes JR. 2014. Outdoor air pollution and asthma. Lancet 383(9928):1581–1592, PMID: 24792855, https://doi.org/10.1016/S0140-6736(14)60617-6.
4. Kurt OK, Zhang J, Pinkerton KE. 2016. Pulmonary health effects of air pollution. Curr Opin Pulm Med 22(2):138–143, PMID: 26761628, https://doi.org/10.1097/MCP.0000000000000248.
5. Weisskopf MG, Kioumourtzoglou MA, Roberts AL. 2015. Air pollution and autism spectrum disorders: causal or confounded? Curr Environ Health Rep 2(4):430–439, PMID: 26399256, https://doi.org/10.1007/s40572-015-0073-9.
6. Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. 2020. Environmental and health impacts of air pollution: a review. Front Public Health 8:14, PMID: 32154200, https://doi.org/10.3389/fpubh.2020.00014.
7. Yang D, Yang X, Deng F, Guo X. 2017. Ambient Air Pollution and Biomarkers of Health Effect. *Adv Exp Med Biol*. 1017:59–102.
8. Feng S, Gao D, Liao F, Zhou F, Wang X. 2016. The health effects of ambient $PM_{2.5}$ and potential mechanisms. Ecotoxicol Environ Saf 128:67–74, PMID: 26896893, https://doi.org/10.1016/j.ecoenv.2016.01.030.
9. Health Effects Institute. 2010. *Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects*. Special Report 17. https://www.healtheffects.org/publication/traffic-related-air-pollution-critical-review-literature-emissions-exposure-and-health [accessed 12 July 2022].
10. Suhaimi NF, Jalaludin J. 2015. Biomarker as a research tool in linking exposure to air particles and respiratory health. Biomed Res Int 2015:962853, PMID: 25984536, https://doi.org/10.1155/2015/962853.
11. Wild CP. 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 14(8):1847–1850, PMID: 16103423, https://doi.org/10.1158/1055-9965.EPI-05-0456.
12. Miller GW, Jones DP. 2014. The nature of nurture: refining the definition of the exposome. Toxicol Sci 137(1):1–2, PMID: 24213143, https://doi.org/10.1093/toxsci/kft251.
13. Vrijheid M. 2014. The exposome: a new paradigm to study the impact of environment on health. Thorax 69(9):876–878, PMID: 24906490, https://doi.org/10.1136/thoraxjnl-2013-204949.
14. Vermeulen R, Schymanski EL, Barabási AL, Miller GW. 2020. The exposome and health: where chemistry meets biology. Science 367(6476):392–396, PMID: 31974245, https://doi.org/10.1126/science.aay3164.
15. Bundy JG, Davey MP, Viant MR. 2009. Environmental metabolomics: a critical review and future perspectives. Metabolomics 5:3–21, https://doi.org/10.1007/s11306-008-0152-0.
16. Lankadurai BP, Nagato EG, Simpson MJ. 2013. Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. Environ Rev 21:180–205, https://doi.org/10.1139/er-2013-0011.
17. Baker MG, Simpson CD, Lin YS, Shireman LM, Seixas N. 2017. The use of metabolomics to identify biological signatures of manganese exposure. Ann Work Expo Health 61(4):406–415, PMID: 28355443, https://doi.org/10.1093/annweh/wxw032.
18. Chen C, Li H, Niu Y, Liu C, Lin Z, Cai J, et al. 2019. Impact of short-term exposure to fine particulate matter air pollution on urinary metabolome: a randomized, double-blind, crossover trial. Environ Int 130:104878, PMID: 31200160, https://doi.org/10.1016/j.envint.2019.05.072.
19. Cheng W, Duncan KE, Ghio AJ, Ward-Caviness C, Karoly ED, Diaz-Sanchez D, et al. 2018. Changes in metabolites present in lung-lining fluid following exposure of humans to ozone. Toxicol Sci 163(2):430–439, PMID: 29471466, https://doi.org/10.1093/toxsci/kfy043.
20. Ladva CN, Golan R, Greenwald R, Yu T, Sarnat SE, Flanders WD, et al. 2017. Metabolomic profiles of plasma, exhaled breath condensate, and saliva are correlated with potential for air toxics detection. J Breath Res 12(1):016008, PMID: 28808178, https://doi.org/10.1088/1752-7163/aa863c.
21. Ladva CN, Golan R, Liang D, Greenwald R, Walker DI, Uppal K, et al. 2018. Particulate metal exposures induce plasma metabolome changes in a commuter panel study. PLoS One 13(9):e0203468, PMID: 30231074, https://doi.org/10.1371/journal.pone.0203468.
22. Li H, Cai J, Chen R, Zhao Z, Ying Z, Wang L, et al. 2017. Particulate matter exposure and stress hormone levels: a randomized, double-blind, crossover trial of air purification. Circulation 136(7):618–627, PMID: 28808144, https://doi.org/10.1161/CIRCULATIONAHA.116.026796.
23. Liang D, Ladva CN, Golan R, Yu T, Walker DI, Sarnat SE, et al. 2019. Perturbations of the arginine metabolome following exposures to traffic-related air pollution in a panel of commuters with and without asthma. Environ Int 127:503–513, PMID: 30981021, https://doi.org/10.1016/j.envint.2019.04.003.
24. Liang D, Moutinho JL, Golan R, Yu T, Ladva CN, Niedzwiecki M, et al. 2018. Use of high-resolution metabolomics for the identification of metabolic signals associated with traffic-related air pollution. Environ Int 120:145–154, PMID: 30092452, https://doi.org/10.1016/j.envint.2018.07.044.
25. Menni C, Metrustry SJ, Mohney RP, Beevers S, Barratt B, Spector TD, et al. 2015. Circulating levels of antioxidant vitamins correlate with better lung function and reduced exposure to ambient pollution. Am J Respir Crit Care Med 191(10):1203–1207, PMID: 25978575, https://doi.org/10.1164/rccm.201411-2059LE.
26. Miller DB, Ghio AJ, Karoly ED, Bell LN, Snow SJ, Madden MC, et al. 2016. Ozone exposure increases circulating stress hormones and lipid metabolites in humans. Am J Respir Crit Care Med 193(12):1382–1391, PMID: 26745856, https://doi.org/10.1164/rccm.201508-1599OC.
27. Mu L, Niu Z, Blair RH, Yu H, Browne RW, Bonner MR, et al. 2019. Metabolomics profiling before, during, and after the Beijing Olympics: a panel study of within-individual differences during periods of high and low air pollution. Environ Health Perspect 127(5):57010, PMID: 31140880, https://doi.org/10.1289/EHP3705.
28. Surowiec I, Karimpour M, Gouveia-Figueira S, Wu J, Unosson J, Bosson JA, et al. 2016. Multi-platform metabolomics assays for human lung lavage fluids in an air pollution exposure study. Anal Bioanal Chem 408(17):4751–4764, PMID: 27113461, https://doi.org/10.1007/s00216-016-9566-0.
29. van Veldhoven K, Kiss A, Keski-Rahkonen P, Robinot N, Scalbert A, Cullinan P, et al. 2019. Impact of short-term traffic-related air pollution on the metabolome—results from two metabolome-wide experimental studies. Environ Int 123:124–131, PMID: 30522001, https://doi.org/10.1016/j.envint.2018.11.034.
30. Vlaanderen JJ, Janssen NA, Hoek G, Keski-Rahkonen P, Barupal DK, Cassee FR, et al. 2017. The impact of ambient air pollution on the human blood metabolome. Environ Res 156:341–348, PMID: 28391173, https://doi.org/10.1016/j.envres.2017.03.042.
31. Walker DI, Lane KJ, Liu K, Uppal K, Patton AP, Durant JL, et al. 2019. Metabolomic assessment of exposure to near-highway ultrafine particles. J Expo Sci Environ Epidemiol 29(4):469–483, PMID: 30518795, https://doi.org/10.1038/s41370-018-0102-5.
32. Wei Y, Wang Z, Chang CY, Fan T, Su L, Chen F, et al. 2013. Global metabolomic profiling reveals an association of metal fume exposure and plasma unsaturated fatty acids. PLoS One 8(10):e77413, PMID: 24143234, https://doi.org/10.1371/journal.pone.0077413.
33. Xia Y, Niu Y, Cai J, Lin Z, Liu C, Li H, et al. 2018. Effects of personal short-term exposure to ambient ozone on blood pressure and vascular endothelial function: a mechanistic study based on DNA methylation and metabolomics. Environ Sci Technol 52(21):12774–12782, PMID: 30259740, https://doi.org/10.1021/acs.est.8b03044.
34. Yan Q, Liew Z, Uppal K, Cui X, Ling C, Heck JE, et al. 2019. Maternal serum metabolome and traffic-related air pollution exposure in pregnancy. Environ Int 130:104872, PMID: 31228787, https://doi.org/10.1016/j.envint.2019.05.066.
35. Zhang Y, Chu M, Zhang J, Duan J, Hu D, Zhang W, et al. 2019. Urine metabolites associated with cardiovascular effects from exposure of size-fractioned particulate matter in a subway environment: a randomized crossover study. Environ Int 130:104920, PMID: 31228782, https://doi.org/10.1016/j.envint.2019.104920.
36. Loo RL, Lu Q, Carter EM, Liu S, Clark S, Wang Y, et al. 2021. A feasibility study of metabolic phenotyping of dried blood spot specimens in rural Chinese women exposed to household air pollution. J Expo Sci Environ Epidemiol 31(2):328–344, PMID: 32709935, https://doi.org/10.1038/s41370-020-0252-0.
37. Liu S, Huang Q, Wu Y, Song Y, Dong W, Chu M, et al. 2020. Metabolic linkages between indoor negative air ions, particulate matter and cardiorespiratory function: a randomized, double-blind crossover study among children. Environ Int 138:105663, PMID: 32203810, https://doi.org/10.1016/j.envint.2020.105663.

38. Li Z, Liang D, Ye D, Chang HH, Ziegler TR, Jones DP, et al. 2021. Application of high-resolution metabolomics to identify biological pathways perturbed by traffic-related air pollution. Environ Res 193:110506, PMID: 33245887, https://doi.org/10.1016/j.envres.2020.110506.
39. Nassan FL, Wang C, Kelly RS, Lasky-Su JA, Vokonas PS, Koutrakis P, et al. 2021. Ambient $PM_{2.5}$ species and ultrafine particle exposure and their differential metabolomic signatures. Environ Int 151:106447, PMID: 33639346, https://doi.org/10.1016/j.envint.2021.106447.
40. Nassan FL, Kelly RS, Kosheleva A, Koutrakis P, Vokonas PS, Lasky-Su JA, et al. 2021. Metabolomic signatures of the long-term exposure to air pollution and temperature. Environ Health 20(1):3, PMID: 33413450, https://doi.org/10.1186/s12940-020-00683-x.
41. Chen CHS, Yuan TH, Shie RH, Wu KY, Chan CC. 2017. Linking sources to early effects by profiling urine metabolome of residents living near oil refineries and coal-fired power plants. Environ Int 102:87–96, PMID: 28238459, https://doi.org/10.1016/j.envint.2017.02.003.
42. Huang Q, Hu D, Wang X, Chen Y, Wu Y, Pan L, et al. 2018. The modification of indoor $PM_{2.5}$ exposure to chronic obstructive pulmonary disease in Chinese elderly people: a meet-in-metabolite analysis. Environ Int 121(pt 2):1243–1252, PMID: 30389378, https://doi.org/10.1016/j.envint.2018.10.046.
43. Jeong A, Fiorito G, Keski-Rahkonen P, Imboden M, Kiss A, Robinot N, et al. 2018. Perturbation of metabolic pathways mediates the association of air pollutants with asthma and cardiovascular diseases. Environ Int 119:334–345, PMID: 29990954, https://doi.org/10.1016/j.envint.2018.06.025.
44. Chu H, Huang FQ, Yuan Q, Fan Y, Xin J, Du M, et al. 2021. Metabolomics identifying biomarkers of $PM_{2.5}$ exposure for vulnerable population: based on a prospective cohort study. Environ Sci Pollut Res Int 28(12):14586–14596, PMID: 33215280, https://doi.org/10.1007/s11356-020-11677-0.
45. Cruz R, Koch S, Matsuda M, Marquezini M, Sforça ML, Lima-Silva AE, et al. 2022. Air pollution and high-intensity interval exercise: implications to anti-inflammatory balance, metabolome and cardiovascular responses. Sci Total Environ 809:151094, PMID: 34688752, https://doi.org/10.1016/j.scitotenv.2021.151094.
46. Cruz R, Pasqua L, Silveira A, Damasceno M, Matsuda M, Martins M, et al. 2021. Traffic-related air pollution and endurance exercise: characterizing non-targeted serum metabolomics profiling. Environ Pollut 291:118204, PMID: 34560574, https://doi.org/10.1016/j.envpol.2021.118204.
47. Du X, Zhang Q, Jiang Y, Li H, Zhu X, Zhang Y, et al. 2022. Dynamic molecular choreography induced by traffic exposure: a randomized, crossover trial using multi-omics profiling. J Hazard Mater 424(pt A):127359, PMID: 34601410, https://doi.org/10.1016/j.jhazmat.2021.127359.
48. Feng B, Liu C, Yi T, Song X, Wang Y, Liu S, et al. 2021. Perturbation of amino acid metabolism mediates air pollution associated vascular dysfunction in healthy adults. Environ Res 201:111512, PMID: 34166659, https://doi.org/10.1016/j.envres.2021.111512.
49. Gao S, Zhuo Z, Hutchinson J, Su L, Christiani DC. 2021. Metabolomic profiling identifies plasma sphingosine 1-phosphate levels associated with welding exposures. Occup Environ Med 78(4):255–261, PMID: 33106349, https://doi.org/10.1136/oemed-2020-106918.
50. Gaskins AJ, Tang Z, Hood RB, Ford J, Schwartz JD, Jones DP, et al. 2021. Periconception air pollution, metabolomic biomarkers, and fertility among women undergoing assisted reproduction. Environ Int 155:106666, PMID: 34116378, https://doi.org/10.1016/j.envint.2021.106666.
51. Hood RB, Liang D, Tang Z, Kloog I, Schwartz J, Laden F, et al. 2022. Length of $PM_{2.5}$ exposure and alterations in the serum metabolome among women undergoing infertility treatment. Environ Epidemiol 6(1):e191, PMID: 35169669, https://doi.org/10.1097/EE9.0000000000000191.
52. Huan S, Jin S, Liu H, Xia W, Liang G, Xu S, et al. 2021. Fine particulate matter exposure and perturbation of serum metabolome: a longitudinal study in Baoding, China. Chemosphere 276:130102, PMID: 33684857, https://doi.org/10.1016/j.chemosphere.2021.130102.
53. Liao J, Gheissari R, Thomas DC, Gilliland FD, Lurmann F, Islam KT, et al. 2022. Transcriptomic and metabolomic associations with exposures to air pollutants among young adults with childhood asthma history. Environ Pollut 299:118903, PMID: 35091019, https://doi.org/10.1016/j.envpol.2022.118903.
54. Liu F, Chen X, Liu Y, Niu Z, Tang H, Mao S, et al. 2021. Serum cardiovascular-related metabolites disturbance exposed to different heavy metal exposure scenarios. J Hazard Mater 415:125590, PMID: 33740723, https://doi.org/10.1016/j.jhazmat.2021.125590.
55. Liu S, Huang Q, Zhang X, Dong W, Zhang W, Wu S, et al. 2021. Cardiorespiratory effects of indoor ozone exposure associated with changes in metabolic profiles among children: a repeated-measure panel study. Innovation (Camb) 2(1):100087, PMID: 34557741, https://doi.org/10.1016/j.xinn.2021.100087.
56. Misra N, Clavaud C, Guinot F, Bourokba N, Nouveau S, Mezzache S, et al. 2021. Multi-omics analysis to decipher the molecular link between chronic exposure to pollution and human skin dysfunction. Sci Rep 11(1):18302, PMID: 34526566, https://doi.org/10.1038/s41598-021-97572-1.
57. Ritz B, Yan Q, He D, Wu J, Walker DI, Uppal K, et al. 2022. Child serum metabolome and traffic-related air pollution exposure in pregnancy. Environ Res 203:111907, PMID: 34419469, https://doi.org/10.1016/j.envres.2021.111907.
58. Selley L, Lammers A, Le Guennec A, Pirhadi M, Sioutas C, Janssen N, et al. 2021. Alterations to the urinary metabolome following semi-controlled short exposures to ultrafine particles at a major airport. Int J Hyg Environ Health 237:113803, PMID: 34517159, https://doi.org/10.1016/j.ijheh.2021.113803.
59. Tang Z, Sarnat JA, Weber RJ, Russell AG, Zhang X, Li Z, et al. 2022. The oxidative potential of fine particulate matter and biological perturbations in human plasma and saliva metabolome. Environ Sci Technol 56(11):7350–7361, PMID: 35075906, https://doi.org/10.1021/acs.est.1c04915.
60. Walker DI, Hart JE, Patel CJ, Rudel R, Chu JH, Garshick E, et al. 2022. Integrated molecular response of exposure to traffic-related pollutants in the US trucking industry. Environ Int 158:106957, PMID: 34737152, https://doi.org/10.1016/j.envint.2021.106957.
61. Wu S, Chen Z, Yang L, Zhang Y, Luo X, Guo J, et al. 2021. Particle-bound PAHs induced glucose metabolism disorders through HIF-1 pathway. Sci Total Environ 797:149132, PMID: 34311363, https://doi.org/10.1016/j.scitotenv.2021.149132.
62. Zhao L, Fang J, Tang S, Deng F, Liu X, Shen Y, et al. 2022. $PM_{2.5}$ and serum metabolome and insulin resistance, potential mediation by the gut microbiome: a population-based panel study of older adults in China. Environ Health Perspect 130(2):27007, PMID: 35157499, https://doi.org/10.1289/EHP9688.
63. Nassan FL, Kelly RS, Koutrakis P, Vokonas PS, Lasky-Su JA, Schwartz JD. 2021. Metabolomic signatures of the short-term exposure to air pollution and temperature. Environ Res 201:111553, PMID: 34171372, https://doi.org/10.1016/j.envres.2021.111553.
64. Jin L, Godri Pollitt KJ, Liew Z, Rosen Vollmar AK, Vasiliou V, Johnson CH, et al. 2021. Use of untargeted metabolomics to explore the air pollution-related disease continuum. Curr Environ Health Rep 8(1):7–22, PMID: 33420964, https://doi.org/10.1007/s40572-020-00298-x.
65. Gertsman I, Barshop BA. 2018. Promises and pitfalls of untargeted metabolomics. J Inherit Metab Dis 41(3):355–366, PMID: 29536203, https://doi.org/10.1007/s10545-017-0130-7.
66. Ribbenstedt A, Ziarrusta H, Benskin JP. 2018. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. PLoS One 13(11):e0207082, PMID: 30439966, https://doi.org/10.1371/journal.pone.0207082.
67. Babineau J. 2014. Product review: Covidence (systematic review software). J Can Health Libr Assoc 35(2):68, https://doi.org/10.5596/c14-016.
68. Health Effects Institute. 2022. *Systematic Review and Meta-Analysis of Selected Health Effects of Long-Term Exposure to Traffic-Related Air Pollution*. Special Report 23. https://www.healtheffects.org/publication/systematic-review-and-meta-analysis-selected-health-effects-long-term-exposure-traffic [accessed 12 July 2022].
69. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 48(4):2097–2098, PMID: 24476540, https://doi.org/10.1021/es5002105.
70. Armstrong RA. 2014. When to use the Bonferroni correction. Ophthalmic Physiol Opt 34(5):502–508, PMID: 24697967, https://doi.org/10.1111/opo.12131.
71. Thissen D, Steinberg L, Kuang D. 2002. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 27(1):77–83, https://doi.org/10.3102/10769986027001077.
72. Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100(16):9440–9445, PMID: 12883005, https://doi.org/10.1073/pnas.1530509100.
73. Marco-Ramell A, Palau-Rodriguez M, Alay A, Tulipani S, Urpi-Sarda M, Sanchez-Pla A, et al. 2018. Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data. BMC Bioinformatics 19(1):1, PMID: 29291722, https://doi.org/10.1186/s12859-017-2006-0.
74. Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, et al. 2013. Predicting network activity from high throughput metabolomics. PLoS Comput Biol 9(7):e1003123, PMID: 23861661, https://doi.org/10.1371/journal.pcbi.1003123.
75. Gao P, da Silva E, Hou L, Denslow ND, Xiang P, Ma LQ. 2018. Human exposure to polycyclic aromatic hydrocarbons: metabolomics perspective. Environ Int 119:466–477, PMID: 30032012, https://doi.org/10.1016/j.envint.2018.07.017.
76. Miller MR. 2020. Oxidative stress and the cardiovascular effects of air pollution. Free Radic Biol Med 151:69–87, PMID: 31923583, https://doi.org/10.1016/j.freeradbiomed.2020.01.004.
77. Araujo JA. 2010. Particulate air pollution, systemic oxidative stress, inflammation, and atherosclerosis. Air Qual Atmos Health 4(1):79–93, PMID: 21461032, https://doi.org/10.1007/s11869-010-0101-8.
78. Chin MT. 2015. Basic mechanisms for adverse cardiovascular events associated with air pollution. Heart 101(4):253–256, PMID: 25552258, https://doi.org/10.1136/heartjnl-2014-306379.

79. Peden DB. 2001. Air pollution in asthma: effect of pollutants on airway inflammation. Ann Allergy Asthma Immunol 87(6 suppl 3):12–17, PMID: 11770676, https://doi.org/10.1016/s1081-1206(10)62334-4.
80. Hu X, Yan M, He L, Qiu X, Zhang J, Zhang Y, et al. 2021. Associations between time-weighted personal air pollution exposure and amino acid metabolism in healthy adults. Environ Int 156:106623, PMID: 33993003, https://doi.org/10.1016/j.envint.2021.106623.
81. Playdon MC, Joshi AD, Tabung FK, Cheng S, Henglin M, Kim A, et al. 2019. Metabolomics analytics workflow for epidemiological research: perspectives from the Consortium of Metabolomics Studies (COMETS). Metabolites 9(7):145, PMID: 31319517, https://doi.org/10.3390/metabo9070145.
82. Antonelli J, Claggett BL, Henglin M, Kim A, Ovsak G, Kim N, et al. 2019. Statistical workflow for feature selection in human metabolomics data. Metabolites 9(7):143, PMID: 31336989, https://doi.org/10.3390/metabo9070143.
83. Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. 2016. Computational metabolomics: a framework for the million metabolome. Chem Res Toxicol 29(12):1956–1975, PMID: 27629808, https://doi.org/10.1021/acs.chemrestox.6b00179.
84. Tzoulaki I, Ebbels TMD, Valdes A, Elliott P, Ioannidis JPA. 2014. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. Am J Epidemiol 180(2):129–139, PMID: 24966222, https://doi.org/10.1093/aje/kwu143.
85. Narath SH, Mautner SI, Svehlikova E, Schultes B, Pieber TR, Sinner FM, et al. 2016. An untargeted metabolomics approach to characterize short-term and long-term metabolic changes after bariatric surgery. PLoS One 11(9):e0161425, PMID: 27584017, https://doi.org/10.1371/journal.pone.0161425.
86. Dolcini J, Kioumourtzoglou MA, Cayir A, Sanchez-Guerra M, Brennan KJ, Dereix AE, et al. 2020. Age and mitochondrial DNA copy number influence the association between outdoor temperature and cognitive function: insights from the VA Normative Aging Study. Environ Epidemiol 4(4):e0108, PMID: 32832843, https://doi.org/10.1097/EE9.0000000000000108.
87. Gao P, Shen X, Zhang X, Jiang C, Zhang S, Zhou X, et al. 2022. Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. Genome Res 32(6):1199–1214, PMID: 35667843, https://doi.org/10.1101/gr.276521.121.
88. Xu J, Hu W, Liang D, Gao P. 2022. Photochemical impacts on the toxicity of $PM_{2.5}$. Crit Rev Environ Sci Technol 52(1):130–156, https://doi.org/10.1080/10643389.2020.1816126.
89. Dominici F, Peng RD, Barr CD, Bell ML. 2010. Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. Epidemiology 21(2):187–194, PMID: 20160561, https://doi.org/10.1097/EDE.0b013e3181cc86e8.
90. Altman DG, Royston P. 2006. The cost of dichotomising continuous variables. BMJ 332(7549):1080, PMID: 16675816, https://doi.org/10.1136/bmj.332.7549.1080.
91. Patel MJ, Batch BC, Svetkey LP, Bain JR, Turer CB, Haynes C, et al. 2013. Race and sex differences in small-molecule metabolites and metabolic hormones in overweight and obese adults. OMICS 17(12):627–635, PMID: 24117402, https://doi.org/10.1089/omi.2013.0031.
92. Johnson LC, Parker K, Aguirre BF, Nemkov TG, D'Alessandro A, Johnson SA, et al. 2019. The plasma metabolome as a predictor of biological aging in humans. Geroscience 41(6):895–906, PMID: 31707594, https://doi.org/10.1007/s11357-019-00123-w.
93. Hu X, Walker DI, Liang Y, Smith MR, Orr ML, Juran BD, et al. 2021. A scalable workflow to characterize the human exposome. Nat Commun 12(1):5575, PMID: 34552080, https://doi.org/10.1038/s41467-021-25840-9.
94. Emwas AH. 2015. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods Mol Biol 1277:161–193, PMID: 25677154, https://doi.org/10.1007/978-1-4939-2377-9_13.
95. Chang CJ, Barr DB, Ryan PB, Panuwet P, Smarr MM, Liu K, et al. 2022. Per- and polyfluoroalkyl substance (PFAS) exposure, maternal metabolomic perturbation, and fetal growth in African American women: a meet-in-the-middle approach. Environ Int 158:106964, PMID: 34735953, https://doi.org/10.1016/j.envint.2021.106964.
96. Zhang X, Barr DB, Dunlop AL, Panuwet P, Sarnat JA, Lee GE, et al. 2022. Assessment of metabolic perturbations associated with exposure to phthalates among pregnant African American women. Sci Total Environ 818:151689, PMID: 34793805, https://doi.org/10.1016/j.scitotenv.2021.151689.
97. Tan Y, Barr DB, Ryan PB, Fedirko V, Sarnat JA, Gaskins AJ, et al. 2022. High-resolution metabolomics of exposure to tobacco smoke during pregnancy and adverse birth outcomes in the Atlanta African American maternal-child cohort. Environ Pollut 292(pt A):118361, PMID: 34655695, https://doi.org/10.1016/j.envpol.2021.118361.
98. Pence HE, Williams A. 2010. ChemSpider: an online chemical information resource. J Chem Educ 87(11):1123–1124, https://doi.org/10.1021/ed100697w.
99. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, et al. 2018. METLIN: a technology platform for identifying knowns and unknowns. Anal Chem 90(5):3156–3164, PMID: 29381867, https://doi.org/10.1021/acs.analchem.7b04424.
100. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. 2018. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 46(D1):D608–D617, PMID: 29140435, https://doi.org/10.1093/nar/gkx1089.
101. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 45(7):703–714, PMID: 20623627, https://doi.org/10.1002/jms.1777.
102. HighChem LLC Slovakia. mzCloud™ [database]. https://www.mzcloud.org/ [accessed 15 December 2021].
103. Meijer J, Lamoree M, Hamers T, Antignac JP, Hutinet S, Debrauwer L, et al. 2021. An annotation database for chemicals of emerging concern in exposome research. Environ Int 152:106511, PMID: 33773387, https://doi.org/10.1016/j.envint.2021.106511.
104. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. 2016. Untargeted metabolomics strategies–challenges and emerging directions. J Am Soc Mass Spectrom 27(12):1897–1905, PMID: 27624161, https://doi.org/10.1007/s13361-016-1469-y.
105. Tachibana C. 2014. What's next in 'omics: the metabolome. Science 345(6203):1519–1521, https://doi.org/10.1126/science.345.6203.1519.
106. Sherrod SD, McLean JA. 2016. Systems-wide high-dimensional data acquisition and informatics using structural mass spectrometry strategies. Clin Chem 62(1):77–83, PMID: 26453699, https://doi.org/10.1373/clinchem.2015.238261.
107. Nemet I, Saha PP, Gupta N, Zhu W, Romano KA, Skye SM, et al. 2020. A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors. Cell 180(5):862–877.e22, PMID: 32142679, https://doi.org/10.1016/j.cell.2020.02.016.
108. Johnson CH, Ivanisevic J, Siuzdak G. 2016. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17(7):451–459, PMID: 26979502, https://doi.org/10.1038/nrm.2016.25.
109. Rochat B. 2016. From targeted quantification to untargeted metabolomics: why LC-high-resolution-MS will become a key instrument in clinical labs. Trends Analyt Chem 84(pt B):151–164, https://doi.org/10.1016/j.trac.2016.02.009.