The role of gene–environment interactions in lung disease: the urgent need for the exposome

Craig E. Wheelock [1] and Stephen M. Rappaport [2]

Affiliations: [1]Division of Physiological Chemistry 2, Dept of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. [2]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA.

Correspondence: Craig E. Wheelock, Division of Physiological Chemistry 2, Dept of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. E-mail: craig.wheelock@ki.se

@ERSpublications

Genetic susceptibility can alter the initiation of lung diseases, but environmental triggers are vital determinants. There is an urgent need to study the exposome – the sum total of environmental exposures – to understand the aetiology of lung diseases. http://bit.ly/2YG5XpP

Defining the exposome

Since the human genome was sequenced, extensive effort has been placed into mapping the role of genes in the onset of disease. It was expected that we would be able to explain the cause of disease and understand the genetic basis of health. However, we have found that while the genetic contribution to individual diseases varies, non-genetic factors have far greater attributable risks, often in the range of 80–90%. The dominance of non-genetic components highlights the importance of the environment to chronic disease risks and has led to the advent of the nascent field of exposome science. In its broadest sense, the exposome can be defined as the totality of all exposures from conception onwards [1]. This all-encompassing description includes multiple exposures ranging from pollution, allergens, diet, lifestyle factors and infections, to human and microbial metabolism (figure 1). A more specific proposed definition of the exposome is the cumulative effects of environmental exposure and the associated biological response [2].

The concept of the exposome was originally introduced to acknowledge the importance of both genes and the environment in cancer aetiology [1]. Genetics has been found to exert a relatively modest effect in the pathogenesis of respiratory disease. Data from monozygotic twins point to a genome-driven population attributable fraction (PAF) of 48.6% for asthma, which represents the proportion of cases that would theoretically be prevented if entire genotypes (plus shared exposures) could be removed [3]. The PAF for asthma in monozygotic twins was the largest for any of 28 chronic diseases and syndromes, including COPD (PAF 18.5%) and lung cancer (PAF 9.89%) [3]. These attributable risks indicate that, whereas genetic susceptibility can alter the initiation and trajectory of lung diseases, environmental triggers are necessary determinants.

The exposome is often interpreted to represent an omics-scale characterisation of non-genetic contributions to an individual's phenome, including products of the epigenome as well as the proteome, metabolome and foreign DNA/RNA. When coupled with metadata related to population characteristics, the exposome motivates investigations for determining the aetiology of a disease. We emphasise in the
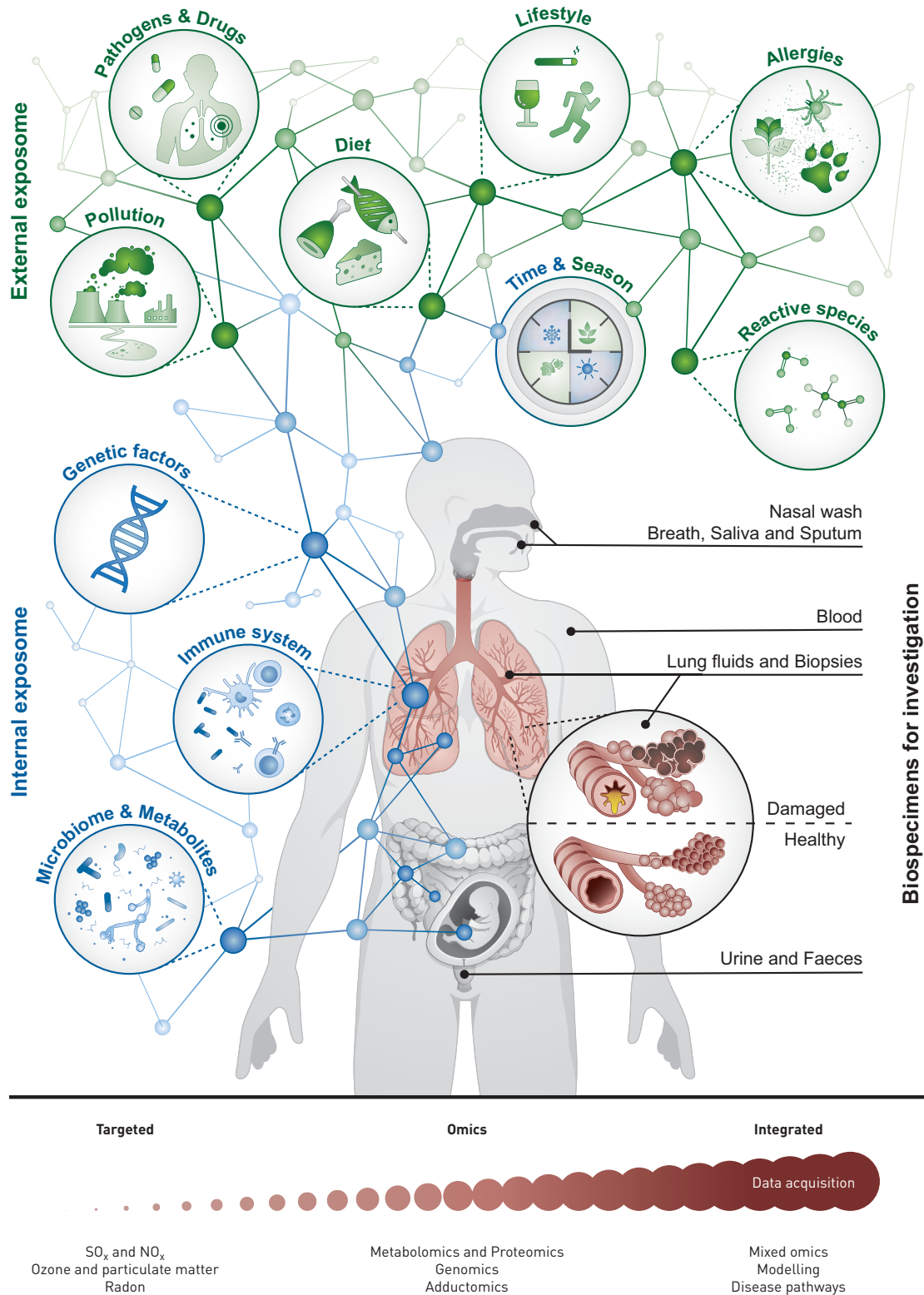
FIGURE 1 The exposome concept. The exposome encompasses the cumulative exposures from conception onwards, integrating internal and external exposures. Multiple biospecimens can be sampled to acquire exposome-based data, with analytical methods ranging from single targeted acquisitions to omics-based profiling and integrative modelling. In its truest sense, the exposome includes the full complement of the processes displayed in this figure, representing a significant analytical and logistical challenge in our efforts to understand the role of the environment in the aetiology of respiratory disease.

context of lung diseases that the exposome constitutes more than the collection of air pollutants that are recognised triggers for lung injury, including reactive gases, particulate matter and environmental tobacco smoke. Indeed, airborne particulate matter contains diverse populations of bacteria, viruses and fungi that affect respiratory health through infections and modulation of the immune system. Instead, the concept of the exposome is an explicit acknowledgement of the fact that exposures to these airborne pollutants, as well as other environmental contaminants, can interact with the genome in combination with the microbiome, dietary and lifestyle factors to influence the incidence and severity of lung diseases. Accordingly, there is a need to understand the combined effects of these factors upon respiratory health and to identify combinations of stressors that require mitigation in sensitive populations.

Analysing the exposome

Characterisation of the exposome in aetiological studies of respiratory disease represents a significant analytical challenge. It has been suggested that it would be more practical to focus on the chemical signature inside the body to identify traces of past exposures. For example, chemical signatures in blood, urine, teeth and hair can contain evidence of previous exposures. Recent investigations of disease aetiology have employed mass spectrometry-based metabolomics to discover circulating small molecules that differ in abundance between incident disease cases and controls. This strategy has the advantage of covering small molecules derived from both exogenous sources, including the diet, and endogenous sources, including microbial metabolism. This approach enables estimates of the full complement of the personalised internal exposome, including 1) measurement of exogenous compounds (*e.g.* quantification of environmental chemicals), 2) metabolic imprints of exposure or exposure–metabolite associations (*e.g.* total arsenic in urine associates with the gut bacteria metabolite TMAO in pregnant women [4]), 3) the biological impact of exposures (*e.g.* acetylcholinesterase activity is a biomarker of organophosphate exposure) and 4) processes that reflect wear and tear in the individual (*e.g.* telomere length and the epigenetic ageing clock). However, with the exception of some occupational investigations, untargeted metabolomics with 10–50 μL of blood cannot robustly detect pollutants that are typically present in normal populations at blood concentrations 1000-fold lower than those arising from the diet and endogenous sources [5]. Thus, a combination of approaches should be considered to integrate omics-level untargeted analyses with panels of known environmental chemicals. This could be achieved through efforts such as the National Health and Nutrition Examination Survey (NHANES) or by simply concatenating datasets when constructing statistical models. Given the generally low concentrations of pollutants in blood, this will be challenging with archived biospecimens from cohort studies that have limited volumes available for analysis. To address this point, there is a pressing need for analytical methods with both increased sensitivity and specificity as well as biobanking of increased sample volumes.

In addition to the small molecule profile (*e.g.* metabolomics), there is a need to acquire additional biochemical signatures of both the external and internal exposome (*e.g.* proteins, miRNAs, viral DNA/RNA). Omics-based measurements of these chemical and biochemical signatures will need to be complemented with estimations of lifestyle, sociological and psychological exposures (*e.g.* green space, noise level, stress). Mobile technology and smart wearables can be components of exposome-based data collection and represent feasible approaches to the collection of personalised exposure data. While the analytical complexity is daunting, this wealth of information can be evaluated in a data-driven agnostic approach to identify potential targets for focused causative studies as part of a "next-generation exposure assessment" paradigm. These approaches will require new data analysis strategies, including machine and deep learning artificial intelligence-based methods for large-scale data interrogation. For example, a recent exposome study reported 70 billion different variables over the ~2.4 years of the study, which only included 15 individuals [6]. Applying these approaches at the cohort or population level will require extensive computing power, most likely involving access to supercomputer facilities.

Adductomics for the exposome

Metabolomics-based methods have been used to measure chemical components of the exposome as well as assess the biological and metabolic/toxicological manifestations of exposures (*i.e.* the internal environment). While effective, traditional metabolomics is not able to capture chemical species that are unstable or reactive (*e.g.* 4-hydroxynonenal, acrolein, malondialdehyde). Metabolic processes generate a constellation of reactive electrophiles that are inherently bioactive because they react with nucleophilic loci in the body, including DNA and functional proteins, to form macromolecular adducts. Unlike DNA adducts, protein adducts are not repaired and tend to form at relatively high levels. Because these reactive metabolites represent important components of the exposome that cannot be measured directly, due to their transient nature *in vivo*, their measurement has motivated a field called "adductomics" that measures adducts at particular loci of circulating proteins or DNA. One such current untargeted method focuses on adducts to the Cys34 of human serum albumin (HSA). Adductomics of HSA are particularly relevant to

investigations of lung diseases because airborne chemicals have been associated with formation of Cys34 adducts following exposure to tobacco smoke [7], indoor combustion products [8], and atmospheric pollutants [9], and specific adducts have been associated with COPD and heart disease in a pilot study [9]. This approach has also identified protein adducts associated with dietary lipids (*e.g.* linoleic acid) [10], demonstrating the ability to acquire integrative exposome profiles of diet and exposure. Adducts associated with reactive carbonyls could be of particular relevance for monitoring in the lung because they can provide links between environmental exposure, inflammatory processes, autoimmunity and lung disease. It would also be of interest to analyse HSA adducts in lung aspirates and to screen for uncommon lipids in lung fluid (*e.g.* unique surfactants) and novel oxidative products that could serve as molecular markers of direct pulmonary exposure. Protein adducts, unlike many other biomarkers, can quantitatively reflect exposures that occurred during the previous weeks or months (depending upon the half-life of the protein), and the long half-life of some adducts could render them useful dosimeters of chemical exposures. Further work should scale-up adductomics methods using high-resolution mass spectrometry to develop global screens of protein adducts, which could serve as sensitive metrics for environmental and dietary exposures.

The microbiome, diet and the exposome

The microbiome (including bacteria, viruses, fungi and helminths) is an emerging contributor to the exposome that can interact directly with the host *via* release of microbial metabolites (*e.g.* microbial volatile organic compounds, short chain fatty acids, extracellular vesicles and DNA/RNA). The interplay between the microbiome, environmental factors, lifestyle and diet creates diverse sets of human and microbial metabolites that can contribute to health and disease in interesting and unanticipated ways. For example, consumption of omega-6 and omega-3 fatty acids is hypothesised to exert a role in the pathophysiology of lung disease. While omega-6 fatty acids (*e.g.* linoleic acid) have been linked to onset of asthma and allergy, omega-3 fatty acids (*e.g.* docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA)) are proposed to promote respiratory health, particularly in response to pollution and smoking. For instance, the linoleic acid pathway has been reported to associate with air pollution and adult onset asthma [11], while higher omega-3 intake was associated with reduced effect of indoor $PM_{2.5}$ (particles with a 50% cut-off aerodynamic diameter of 2.5 μm) on paediatric asthma morbidity [12]. In addition, the diet acting on the gut microbiota has been shown to influence airway responses, and suggested as an approach to prevent asthma [13].

Fatty acids are metabolised into downstream lipid mediators with potent biological functions in lung disease (*e.g.* eicosanoids). It is most likely that these lipid mediator metabolic products are responsible for the observed health effects of fatty acids. For example, linoleic acid products (*e.g.* leukotoxin, leukotoxin-diol) have been reported in association with acute respiratory distress syndrome and COPD, and the EPA- and DHA-derived specialised pro-resolving mediators (*e.g.* resolvins, protectins) reduce pulmonary inflammation and enhance microbial clearance. However, recent work has demonstrated that these processes can be driven by the gut microbiome, which converts linoleic acid to bioactive lipids (*e.g.* 12,13-DiHOME, isoleukotoxin-diol) with links to asthma [14] and metabolic disease [15]. The microbiome in turn can be influenced by environmental exposures, with organophosphate pesticides, heavy metals and traffic-related air pollution (TRAP) reported to be associated with microbial dysbiosis [16]. Accordingly, there is significant cross-talk between the diet, the microbiome and environmental exposure, which can modulate respiratory health. This demonstrates that environmental interactions within the lung–gut axis represent fruitful areas for investigation. These complex interactions argue strongly for the necessity of an exposome approach to investigating the aetiology and pathophysiology of lung disease. It is simply not feasible to understand the underlying mechanisms in isolation; instead a holistic systems-based approach is required to achieve a full understanding of the origins of lung disease.

The temporal nature of the exposome

The exposome is highly dynamic, with spatiotemporal variability, and the timing of an exposure can influence the associated health effects. The interaction of individuals and their environment at various stages of growth, development, and ageing trajectories needs to be considered. The dynamic nature of the exposome has led to the concept of windows of susceptibility in which the timing of exposure can exert strong effects upon disease onset, raising questions regarding the relative importance of pre- and post-natal environmental triggers of lung disease. For example, increased $PM_{2.5}$ levels during a specific prenatal window in the third trimester were found to associate with decreased mtDNA content in cord blood, a marker of oxidative stress [17]. Infants at risk of asthma were reported to exhibit transient gut microbial dysbiosis during the first 100 days of life, suggesting a window in which microbial-based therapeutics may be efficacious in preventing the development of asthma [18]. Recent work found that black carbon particles accumulate on the fetal side of the placenta, which suggests a potential mechanism

explaining the negative health effects of *in utero* exposure [19]. There is accordingly a need to examine life-stage and time dependency of environmental exposures, and this requires specific study designs. For example, while prenatal exposure to perfluorooctanoate and perfluorononanoate was associated with lower forced expiratory volume in 1 s in childhood, it is unclear if there were trimester-specific effects [20]. It should also be stressed that the dynamic nature of the exposome extends to adulthood and should be factored into cohort design when working with adult populations. It is a significant challenge to incorporate life-long dynamic fluctuations in exposures into exposome-based studies; however, there is a clear need. One approach is simply to collect additional samples at more time points. However, this will naturally result in increased study costs as well as collection and analysis time. An alternative approach is to focus on the analysis of matrices such as teeth, that offer sufficient temporal resolution to assign exposures within specific time intervals, or protein adducts, with extended half-lives.

Biosampling for measuring the exposome

Because the respiratory system is connected directly to the air environment, it is uniquely positioned to reflect airborne exposures as well as gene-environment interactions. Yet, with 40 different cell types, the complexity of the lung renders it challenging to determine the most appropriate strategy for examining the exposome in relation to respiratory disease, especially for mechanistic investigations. Focusing on the airway epithelium greatly reduces the number of cell types and is a reasonable approach given that it is the receptor of the vast majority of inhaled pollution. Accordingly, epithelial brushings may be particularly insightful for mechanistic investigations; however, this is clearly not feasible on the cohort scale. Current large-scale approaches to interrogate the exposome have focused primarily on blood, and circulating signatures of pulmonary biology have been observed for both asthma and COPD. However, biospecimens derived from lung fluids (*e.g.* bronchoalveolar lavage fluid, bronchial wash), or more distal fluids (*e.g.* exhaled breath condensate, sputum, saliva, nasal washes) may be more relevant to respiratory diseases (figure 1). Saliva in particular has been suggested to be well-suited for longitudinal sampling of the exposome. There is a need to develop a normalisation strategy (*e.g.* cell type/count or protein/metabolite marker that is homeostatically controlled) to compare and select the most appropriate biofluids(s) for evaluating the influence of the exposome and to archive biospecimens for investigations of lung diseases. Ideally, cohorts would include longitudinal sampling of entire populations. While this may not be feasible, repeat sampling can be performed for subgroups of the full cohort to investigate sources of exposure variability within and between subjects and their contributions to attenuation of exposure–response relationships. It would be exceedingly useful to perform a focused study in which multiple biospecimens were collected from different locations in order to determine which compartment provided the most relevant molecular signature of lung disease, and the associated exposures. This information would be highly informative for study/cohort design to ensure long-term sample suitability.

Actionising the exposome

Global trends including climate change, intensified urbanisation, increased malnutrition and microbial dysbiosis will coalesce to increase the incidence of lung disease. Understanding the intense crosstalk and interplay between these factors requires an exposome approach. An important component of these efforts will be to capture the biological impact of these complex exposures. Documenting the exposures is insufficient; there is a need to demonstrate the associated biological and health effects as part of the effort to actionise the exposome and move beyond associations to understanding casual pathways of exposure. However, in order to measure the exposome, we will need to augment traditional analytical chemistry and environmental health approaches to include untargeted measures of exposure. While analytical advances including high-resolution mass spectrometry will play an important role, we will also need to incorporate big data approaches to acquire comprehensive exposure profiles. There remains a fundamental need for statistical methods capable of fusing the multidimensional and longitudinal data streams emerging from measurements of the external and internal exposome. Extensive collaboration will be required in order to achieve the necessary scale of data acquisition. As analytical methods for interrogation of the exposome continue to advance, it will be possible to investigate both multiple genetic factors (GWAS) and exposures (EWAS) with high resolution. This could be performed with a genome-wide inferred study (GWIS) approach, which provides an approximation of GWAS summary statistics for a phenotype that is a function of other phenotypes. Without such parity, we are unlikely to understand the impacts of genes, environment and their interactions on the incidence and mitigation of respiratory diseases.

The question remains: how do we actionise the exposome? While the utility of an integrated signature of lifelong exposure in understanding disease processes is clear, there is a need to translate this information to the clinic and, ultimately, to the patient. Ideally, the exposome would enable individuals to make specific lifestyle choices to alter their susceptibility to given exposures and, particularly, mixtures of exposures. While we are still far from this scenario, the concept is not in the realm of science fiction.

The combination of personal wearables with clinical information and genetic history will eventually enable individuals to monitor their real-time exposures and correlate these data with their current disease status (*e.g.* home spirometry kits). The advent of scalable reality mining of machine-based environmental data reflecting social behaviours will provide a wealth of information for exposome-based mapping. While daunting in its scale, initial steps can already be taken now. Cohort studies can be designed to incorporate as much exposure data as possible (*e.g.* NHANES), which will enable legacy analyses. These data can come from smartphone sensors, GIS modelling and portable monitors, as well as targeted analyses of known chemicals of concern (*e.g.* perfluorinated alkylated substances). Omics data, in particular high-resolution mass spectrometry data, can be collected and used to search exposure databases (*e.g.* Exposome-Explorer, Blood Exposome DB, Toxic Exposome DB) to provide comprehensive chemical signatures. These data can be placed in repositories (*e.g.* MetaboLights, Metabolomics Workbench) to enable both reanalysis as new data processing workflows are developed as well as meta-analyses. The potential of this approach has already been made clear by resources such as the ExpoApp [21], which enables smart phone-based real-time monitoring of individualised exposome profiles. The coupling of smart technology with exposure data and direct patient engagement will be a powerful advance in interrogating the complex relationships between environmental exposure, diet, the microbiome and lung disease. The future of pulmonary medicine needs to include this exposome-based approach to understanding disease, which will enable individuals to actionise their real-time exposome.

References

1 Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 1847–1850.

2 Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. *Toxicol Sci* 2014; 137: 1–2.

3 Rappaport SM. Genetic factors are not the major causes of chronic diseases. *PLoS One* 2016; 11: e0154387.

4 Maitre L, Robinson O, Martinez D, *et al.* Urine metabolic signatures of multiple environmental pollutants in pregnant women: an exposome approach. *Environ Sci Technol* 2018; 52: 13469–13480.

5 Rappaport SM, Barupal DK, Wishart D, *et al.* The blood exposome and its role in discovering causes of disease. *Environ Health Perspect* 2014; 122: 769–774.

6 Jiang C, Wang X, Li X, *et al.* Dynamic human environmental exposome revealed by longitudinal personal monitoring. *Cell* 2018; 175: 277–291. e231.

7 Grigoryan H, Edmands W, Lu SS, *et al.* Adductomics pipeline for untargeted analysis of modifications to Cys34 of human serum albumin. *Anal Chem* 2016; 88: 10504–10512.

8 Lu SS, Grigoryan H, Edmands WM, *et al.* Profiling the serum albumin Cys34 adductome of solid fuel users in Xuanwei and Fuyuan, China. *Environ Sci Technol* 2017; 51: 46–57.

9 Liu S, Grigoryan H, Edmands WMB, *et al.* Cys34 adductomes differ between patients with chronic lung or heart disease and healthy controls in central London. *Environ Sci Technol* 2018; 52: 2307–2313.

10 Shibata T, Uchida K. Protein adductomics: a comprehensive analysis of protein modifications by electrophiles. *Free Radic Biol Med* 2019; 144: 218–222.

11 Jeong A, Fiorito G, Keski-Rahkonen P, *et al.* Perturbation of metabolic pathways mediates the association of air pollutants with asthma and cardiovascular diseases. *Environ Int* 2018; 119: 334–345.

12 Brigham EP, Woo H, McCormack M, *et al.* Omega-3 and omega-6 intake modifies asthma severity and response to indoor air pollution in children. *Am J Respir Crit Care Med* 2019; 199: 1478–1486.

13 Thorburn AN, McKenzie CI, Shen S, *et al.* Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat Commun* 2015; 6: 7320.

14 Levan SR, Stamnes KA, Lin DL, *et al.* Elevated faecal 12,13-diHOME concentration in neonates at high risk for asthma is produced by gut bacteria and impedes immune tolerance. *Nat Microbiol* 2019; 4: 1851–1861.

15 Miyamoto J, Igarashi M, Watanabe K, *et al.* Gut microbiota confers host resistance to obesity by metabolizing dietary polyunsaturated fatty acids. *Nat Commun* 2019; 10: 4007.

16 National Academies of Sciences, Engineering, and Medicine; Division on Earth and Life Studies; Board on Life Sciences; Board on Environmental Studies and Toxicology; Committee on Advancing Understanding of the Implications of Environmental-Chemical Interactions with the Human Microbiome. *Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy.* Washington, National Academies Press, 2017.

17 Rosa MJ, Just AC, Guerra MS, *et al.* Identifying sensitive windows for prenatal particulate air pollution exposure and mitochondrial DNA content in cord blood. *Environ Int* 2017; 98: 198–203.

18 Arrieta MC, Stiemsma LT, Dimitriu PA, *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* 2015; 7: 307ra152.

19 Bove H, Bongaerts E, Slenders E, *et al.* Ambient black carbon particles reach the fetal side of human placenta. *Nat Commun* 2019; 10: 3866.

20 Agier L, Basagana X, Maitre L, *et al.* Early-life exposome and lung function in children in Europe: an analysis of data from the longitudinal, population-based HELIX cohort. *Lancet Planet Health* 2019; 3: e81–e92.

21 Donaire-Gonzalez D, Valentin A, van Nunen E, *et al.* ExpoApp: an integrated system to assess multiple personal environmental exposures. *Environ Int* 2019; 126: 494–503.