Title

Utilizing a biology-driven approach to map the exposome in health and disease: an essential investment to drive the next generation of environmental discovery

Authors:

Ming Kei Chung^{1*}, Stephen M. Rappaport², Craig E. Wheelock³, Vy Kim Nguyen⁴, Thomas P. van der Meer⁵, Gary W. Miller⁶, Roel Vermeulen⁷, Chirag J. Patel^{1*}

¹Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115 USA

²Program in Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, California, 94720 USA. E-mail: srappaport@berkeley.edu

³Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. E-mail: craig.wheelock@ki.se

⁴Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI, USA; Department of Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI, USA E-mail: nguyenvy@umich.edu

⁵Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. E-mail: t.p.van.der.meer@umcg.nl

⁶Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY. E-mail: gary.miller@columbia.edu

⁷Utrecht University & Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. E-mail: r.c.h.vermeulen@uu.nl

Corresponding authors*:

Ming Kei (Jake) Chung

Department of Biomedical Informatics

Harvard Medical School

Harvard University

10 Shattuck Street, Room 302

Boston, MA 02115

jake_chung@hms.harvard.edu

jakemk@berkeley.edu

Chirag J. Patel

Department of Biomedical Informatics

Harvard Medical School

Harvard University

10 Shattuck Street, Room 302

Boston, MA 02115

1-617-432-1195

chirag_patel@hms.harvard.edu

Acknowledgements

The authors declare no competing financial interests. This study was supported by grants from the U.S. National Institutes of Health through the National Institute for Environmental Health Sciences (S.M.R.; P01ES018172, P50ES018172 & P42ES004705; V.K.N.; R01ES028802), the Swedish Heart-Lung Foundation (C.E.W.; Hjärt-Lungfonden 20170736; 20180290), the U.S. National Institutes of Health (G.W.M.; U2CES030163, RC2 DK118619), the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research through the EXPOSOME-NL (R.V.; 024.004.017) and European Union's Horizon 2020 research and innovation programme (R.V.; 874627), the National Institutes of Health through the National Institute for Allergy and Infectious Diseases (C.J.P.; AI127250), and National Institute of Mental Health and National Institute of Allergy and Infectious Diseases (M.K.C; R01MH107205, R01AI127250). Certain icons in figures were made by Freepik from www.flaticon.com.

Abstract

Background:

Recent developments in technologies have offered the opportunities to measure the exposome with unprecedented accuracy and scale. However, because most investigations have targeted only a few exposures at a time, the majority of the environmental determinants of chronic diseases remain unknown.

Objectives:

We describe a “functional exposome” concept and explain how it can leverage existing bioassays and high-resolution mass spectrometry for exploratory study. We discuss how such an approach can address well-known barriers to interpret exposures and present a vision of next-generation exposomics.

Discussion:

The exposome is vast. Instead of trying to capture all exposures, we can reduce the complexity by measuring the functional exposome — the totality of the biologically-active exposures relevant to disease development — through coupling biochemical receptor-binding assays with affinity purification–mass spectrometry. We claim the idea of capturing exposures with functional biomolecules opens new opportunities to solve critical problems in exposomics, including low-dose detection, unknown annotations, and complex mixtures of exposures. Though novel, biology-based measurement can make use of the existing data processing and bioinformatics pipelines. The functional exposome concept also complements traditional and emerging research frameworks for understanding the exposure-disease relationships.

Conclusions:

While measurement technology has advanced, there are critical technological, analytical, and inferential barriers impeding the detection of many environmental exposures relevant to chronic-disease etiology. Through biology-driven exposomics, it is possible to simultaneously scale up discovery of these causal environmental factors.

Keywords:

exposome, high-resolution mass spectrometry, risk assessment, chemical exposures, mixture, effect-directed analysis, environmental health

Glossary

Environmental exposure: a non-host, exogenous exposure.

Biologically-active exposure: a molecular exposure capable of inducing a biological change at the observed level in the population.

Ligand: a substance that binds to a biomolecule such as DNA and protein thereby forming a complex that alters the biological activity. A ligand can be a small molecule, an ion, or a large molecule such as a peptide or protein.

Feature: a distinctive signal defined by a measurement technology. For example, with liquid chromatography–mass spectrometry a feature can be defined by its retention time and mass-to-charge ratio.

Exposome: an individual's life-course non-genetic exposures that could be classified as internal, specific external, or general external (Wild 2012).

Functional exposome: a subset of the exposome that consists of the biologically-active exposures present in an individual.

Biology-driven exposomics: a study approach that utilizes functional biomolecules to capture biologically-active exposures and generate hypotheses for follow-up.

Biology-based measurement (BBM): an umbrella term for measurement *approaches and technologies*, notably mass spectrometry, that employ functional biomolecules to capture biologically-active exposures.

Functional-exposome measurement (FEM): a measurement *process*, that is powered by a few targeted, a particular class of, or ostensibly all the functional biomolecules of an individual, intended to characterize the functional exposome.

Introduction

Environment in health and disease

Genes (G), environmental factors (E), and their interplay (G x E) are thought to contribute to development of chronic diseases. Extensive analyses from twin and health-record studies have revealed that — except for a handful of rare disorders — genetics on average explains 30-50% of disease risk (Czene et al. 2002; Lakhani et al. 2019; Polubriaginof et al. 2018). On the other hand, common environmental factors, including diet, smoking, and air pollution together may contribute about 46 percent of global deaths (Lim et al. 2012; Rappaport 2016). The specific environmental drivers remain elusive for many complex chronic diseases and it is difficult to disentangle the interactions among the elements of mixture exposures (Chung et al. 2018; Patel and Ioannidis 2014; Patel and Manrai 2015)

The exposome encompasses the life-course of exposures received by an individual (Wild 2005) and provides a unified framework for discovering environmental determinants of diseases (Rappaport and Smith 2010). The exposome has been embraced by both governmental and research communities (Birnbaum 2012; Lippmann 2013; Wheelock and Rappaport 2020) and has motivated initiatives focusing on early-life exposures and life-course approaches investigating the impacts of complex environmental exposures (Vineis et al. 2017; Vrijheid Martine et al. 2014). It has also stimulated development of new technologies to study exposures based upon measurements obtained both

outside (Jiang et al. 2018; Vineis et al. 2017) and inside the body (Dennis et al. 2016) such as remote sensing and portable sensors. Data derived from such investigations not only characterize the exposome but also offer clues on potentially causal exposures.

Data-driven exposomics

Almost all epidemiological studies of environmental factors are hypothesis-driven. That is, they target selected pollutants or dietary and lifestyle factors to identify associations with a particular disease and, ultimately, to quantify exposure-response relationships. In contrast, data-driven studies are being increasingly applied to discover risk factors for disease, particularly for genetic epidemiology. Indeed, genome-wide association studies (GWASs) employ a top-down strategy to simultaneously investigate associations between millions of genetic variants and a disease outcome (Tam et al. 2019). To establish an equivalent top-down strategy for environmental epidemiology, exposome-wide association studies (EWASs) are being designed to simultaneously detect many thousands of small-molecule features in biospecimens and compare their abundances between healthy and diseased subjects, preferably from a longitudinal cohort (Rappaport 2012, 2018; Rappaport and Smith 2010). The approach is hypothesized to overcome challenges in multiplicity and fragmented literature of associations that occur due to non-systematic testing (Ioannidis et al. 2009)

Researchers can seek out environmental risk or protective factors with broad-scale targeted measurement (Balik-Meisner Michele et al. 2018; Patel et al. 2010; Sipes et al. 2013) or screen for relevant exposures with semi-targeted approaches such as suspect screening, which search for a predefined and typically by-class chemical space (Andra et al. 2017; Wang Aolin et al. 2018). Currently, untargeted measurement of the blood exposome is conducted with high-resolution mass spectrometry (HRMS) because of its exceptional capability to distinguish peaks in mass spectra with high mass accuracy (less than 1 ppm level). Several studies have demonstrated the capability of HRMS to detect a large number of both endogenous and exogenous exposures in observational studies (Board on Life Sciences et al. 2016; Go et al. 2015a; Jamin et al. 2014; Roca et al. 2014). Although HRMS has largely been used to discover changes in biological pathways associated with known exposures (Turner et al. 2018; Vrijheid Martine et al. 2014; Walker et al. 2016b; Yuan et al. 2016) or to find biomarkers of diseases (Lu et al. 2016; Osborn et al. 2013; Roede et al. 2013; Wang et al. 2011), there is now an impetus to apply this technology to identify environmental risk factors in biomedical and epidemiological studies (Rappaport 2018).

Challenges of data-driven exposomics in observational human research

We identify two main challenges for wide application of data-driven exposomics. First, the environmental exposures relevant to chronic diseases are likely to be low dose. Summarizing from two databases, concentrations of many exogenous compounds in blood, e.g., dietary micronutrients,

phytoestrogens, microbial metabolites, drugs, consumer chemicals, and pollutants, are lower than 5 μM (Rappaport et al. 2014). One of the major distinctions between targeted and untargeted MS is the trade-off between sensitivity and coverage — the former involves optimized methodology for selected compounds for better sensitivity and specificity. This means that a sizable fraction of the exposome of relevance to chronic diseases may not be well measured by untargeted HRMS. Second, annotation of features is often incomplete in HRMS. During analysis, a large proportion of the interesting and outcome-associating signals are uninterpretable because they are “dark matter” of the exposome — lacking annotation or having ambiguous identity (Vermeulen et al. 2020). Putative annotation is generally performed by matching the accurate mass and spectrum of the features to the known records in reference databases, which can be specialized for endogenous metabolites (Kanehisa and Goto 2000; Wishart et al. 2018) or xenobiotic compounds (Neveu et al. 2017; Wishart et al. 2015, 2008). This will lead to biased reporting because the latter contains only about 10 000 common chemicals. In contrast, PubChem contains over 60 million records (Kim et al. 2016) but only a small fraction of this database comes with information related to sources and biological activities. This puts a premium on follow-up investigations to validate potentially causative features with independent biospecimens and chemically identify the most promising hits.

So far, rapid development in exposure assessment technology and infrastructure have transformed the exposome concept to a research paradigm. While significant progress has been made to advance our

understanding on *how* candidate exposures could influence human health, we still lack the capacity for answering one of the most fundamental questions in biomedical research: *what* are the environmental factors, in aggregate and in their totality, that can cause or prevent a disease?

Discussion

Biology-driven exposomics

We are likely decades away from measuring the life-course exposures of an individual. A top-down strategy provides a way to explore the blood exposome, but challenges from miniscule and complex exposures have become a major hurdle in data-driven exposomics. If the ultimate goal is to find the environmental causes of diseases, we do not need to measure all the entities. We believe that a pragmatic view is to focus on those chemicals that 1) we are exposed to and 2) are biologically active — capable of inducing a biological change at the observed level in the population. From a molecular view, almost all exposures, including chemicals entering the body and initial signaling molecules synthesized by the body due to non-chemical stimuli (e.g., mental stress and light), exert their biological effects through interacting with functional biomolecules, such as DNA, mRNA (riboswitch), and protein (receptor and enzyme) (Peters et al. 2021; Vermeulen et al. 2020).

We define the *functional exposome* as encompassing the life-course exposures to both endogenous and exogenous biologically-active molecules. It is a reduced set of exposures at the interface of environment (nurture) and biology (nature) that is important for explaining the variation in disease

among individuals. Indeed, the essence of the functional exposome has been suggested by early exposome advocates (Miller and Jones 2014; Rappaport and Smith 2010; Wild 2012), and finding the determinants of health and diseases through the functional exposome is analogous to using expressed sequence tags to speed up gene discovery in the early days of genomics (Adams et al. 1991). We define biology-based measurement (BBM, Figure 1) as an umbrella term for the collection of analytical methods or technologies that can characterize the molecular interaction between an exposure and a functional biomolecule, and thereby perform functional-exposome measurement (FEM).

Types of molecular binding

Broadly speaking, molecular binding can be classified into two types (Klaassen 2018). First, some interactions result in irreversible non-specific binding of reactive electrophiles to functional macromolecules. These reactions form adducts that can change the biological activity of the biomolecules or initiate mutations that can induce cancers. A second class of interactions includes reversible (non-covalent) and specific binding caused by ligands, which can be small molecules, ions, peptides or proteins that modulate the biological function of a target.

Approaches to search for biologically-active exposures

Effect-directed analysis shares the same goal of BBM and is an effective yet tedious approach for finding compounds of interest in a mixture (Brack et al. 2016). It involves an iterative sample

fractionation process by chromatography with effect assessment spanning from molecular bindings to whole-organism assays. This approach has been successfully applied to identify 6PPD-quinone, an ozonation product of a common tire chemical 6PPD, in stormwater that is responsible for the sudden death of coho salmon for decades (Tian et al. 2021).

Because reactive chemicals have short lifespans *in vivo*, their adducts are used as proxies for exposures. Chemoproteomic platform has been developed to conduct proteome-wide mapping of targeted environmental electrophiles (Medina-Cleghorn et al. 2015) and untargeted adductomics has also demonstrated its potential to discover exposures to reactive electrophiles (Carlsson et al. 2019; Grigoryan et al. 2016; Rappaport et al. 2012).

For studying reversible binding, affinity selection–mass spectrometry is one option. Briefly, a selection of proteins or enzymes is incubated with the mixtures and the bound complexes are separated with size exclusion chromatography and characterization of the active compounds by mass spectrometry (Annis et al. 2007; Yang et al. 2020). Affinity purification–mass spectrometry (AP-MS) is a similar technology; however, active compounds in the mixtures are pulled down by immobilized biomolecules (Dunham et al. 2012).

Affinity-based methods have been applied to proteomics and molecular biology to understand interaction and activity of molecules (Morris et al. 2014; Rinschen et al. 2019). Yet, there is no known attempt to apply an untargeted platform for systematic identification of biologically active molecules in a human observational study. Since receptor binding is inherently relevant to the development of complex chronic diseases, we focus our remaining discussion on ligand-receptor binding and the adaptation of binding assay into affinity purification format, which is one of the BBM approaches for FEM. We then discuss how BBM, exemplified by AP-MS, can address well-known bottlenecks in data-driven studies and our view on next generation exposomics.

Binding assays to assess effects of exposures

Traditional molecular studies typically characterize the receptor activity of known ligands using either binding or functional reporter assays (Seethala and Zhang 2016). The former is a type of biochemical-based assay utilizing a purified recombinant protein and engineered ligands, which are radiolabeled or fluorophore-tagged for generating signals proportional to the binding level. The latter is cell-based and uses genetically modified cells to produce a detectable signal upon receptor activation. These assays are also used in *in vitro* high-throughput screening for drug discovery in the pharmaceutical industry with over 1 million compounds (Carnero 2006; Mayr and Bojanic 2009) and for compound toxicity in the U.S. EPA ToxCast project with more than 1858 chemicals and 821 assays (Filer et al. 2014; Sipes et al. 2013).

Measuring the functional exposome

When both ligand and receptor are known, an optical system is a simple and sensitive choice for measurement (Seethala and Zhang 2016). In the context of data-driven exposomics, exposures are unknowns that require structural characterization and thus cannot be detected by existing binding assays. However, biochemistry-based assays can be adapted to screen for classes of ligands with AP-MS (Dunham et al. 2012). As an example, tagged recombinant receptors (“bait”) can be immobilized on a microarray which bind to specific ligands (“prey”) during incubation. After removing the unbound sample, the intensity and mass spectra of the recovered ligands (putative exposures that were bound to a receptor) are measured with HRMS (Figure 2). Here, we only describe one of the many approaches that leverages existing methodologies for BBM. The workflow is modular and substitutable. For example, baits can be immobilized on microbeads and ligands can be detected by various HRMS platforms (Lanucara et al. 2014; Morris et al. 2014; Wieser et al. 2012).

The entire data preprocessing pipeline for HRMS can be applied to the raw data generated by AP-MS, including feature detection, filtering, alignment, and normalization (Dunn et al. 2011). There are online resources (Tautenhahn et al. 2012; Xia et al. 2009) and offline software (Uppal et al. 2013; Yu et al. 2009) for data and bioinformatic processing, and data can be exported for additional visualization and statistical analyses, e.g., correlation globe (Chung et al. 2018; Patel and Manrai 2015). Depending on

the research questions, the signals can be summarized into three ways: subject-by-feature, subject-by-receptor, and receptor-by-feature (Figure 2). The first data representation will fit directly into an EWAS framework in which chemical features are used as the input variables.

Opportunity and new areas

While BBM quantifies chemical exposures akin to current-day HRMS, it has three innovative aspects that provide solutions to solve some of the well-known barriers in data-driven exposomics in real-world scenarios (Appendix 1, Figure 3).

1. Enrichment and high-throughput screening

Much of the blood exposome consists of low-abundance features that are poorly characterized by untargeted HRMS that sacrifices sensitivity for wider coverage (Rappaport et al. 2014). The pull-down format of AP-MS is equivalent to enriching the samples — keeping potentially interesting ligands while removing the non-binding species — hence increasing the signal-to-noise ratio and sensitivity of the measurement (Chung et al. 2014; Florentinus-Mefailoski et al. 2014).

Another benefit of AP-MS is the optional use of liquid chromatography to separate complex components. Unlike many optical detection systems capable of massive parallel measurement, liquid chromatography-HRMS operates in serial fashion and a chromatography gradient profile typically takes

30 to 60 min per sample (Vos et al. 2007; Yuan et al. 2012). Enriched samples have a clear matrix and therefore can be analyzed directly by novel commercialized high-throughput HRMS that take only 10 to 60 seconds per sample (E. Kempa et al. 2019).

2. Annotating the “dark matter”

The Metabolomics Standards Initiative has set identification standards of features from MS that have four levels of confidence (Sumner et al. 2007): 1) unknown compounds that are differentiable and quantified; 2) putative characterization by compound class; 3) putative identification by matching the accurate mass and spectrum to the records in libraries; and 4) compound confirmation by comparing properties with those of an authentic reference standard. Untargeted HRMS can detect 10 000 to 20 000 features in blood samples, but even though there are accurate mass measures and retention time, up to 80% cannot be identified (Uppal et al. 2016), and of those putatively identified, only a small fraction is reported with higher confidence (Go et al. 2015b; Walker et al. 2016a, 2016b). Further, sources and biological effects of the identified compounds may not be available in the literature, making meaningful inference of their potential roles in disease development difficult. This is an important drawback for using conventional untargeted HRMS to investigate biological plausibility of potential health effects arising from thousands of features.

In contrast, each feature detected by BBM is a molecular entity that is physically bound to a receptor modulating a known biological function and is therefore intrinsically interesting and functionally interpretable without *a priori* feature identification. To accelerate the pace of discovery, functional annotation complements existing approaches to eventually identify unknown features. For example, binding information could enhance the interpretability of correlation-based network analysis of features (Quinn et al. 2017; Uppal et al. 2016).

3. Analyzing mixtures

Humans are exposed to complex chemical mixtures from endogenous and exogenous sources that are common to everyday life (Rappaport and Smith 2010). Past epidemiological studies have tended to analyze only one or a few exposures and confounders at a time. This practice can inflate false positives and lead to publication bias because the effects of multiple comparisons are typically not considered (Ioannidis 2005, 2008) or may be confounded by variables not assayed in the study. Also, it eliminates the opportunity to study how multiple exposures can interact in disease processes. In its 2018–2023 strategic plan, NIEHS highlighted this problem as a focus area for future research (Birnbaum Linda S. 2018).

To address this problem, we can analyze multiple exposures together with statistical methods that are robust to model assumptions or account for the correlation among predictors. These methods

generally involve data reduction to enhance interpretation, for examples, (sparse) partial least squares regression, elastic net, and Bayesian kernel machine regression (Agier et al. 2016; Lazarevic Nina et al. 2019; Taylor et al. 2016), and had been compared in terms of computational performance and strength and weakness, such as bias-variance tradeoff and in epidemiological settings (James et al. 2017; Patel 2017), to understand the best-use scenarios.

In toxicology, the mixture problem is addressed with an effect-based approach. Exposures with structural similarity (e.g., PCBs and dioxins; PAHs; xenoestrogens) generally act together because they share the same mode of action (i.e., acting on the same receptor). The joint effects can reasonably be assumed additive and hence their effects can be expressed in a relative scale as toxic equivalency factor (TEF). The total toxicity of a mixture, called toxic equivalents, is the weighted sum of individual TEFs by concentration (Nisbet and LaGoy 1992; Van den Berg M et al. 1998).

BBM presents mixture exposure data in two ways (Figure 2). For features as input variables (i.e., integrating intensity across receptors), we can apply off-the-shelf regression approaches to identify associations and co-exposure patterns. On the other hand, a complex mixture contains thousands of compounds and those with the same mode of action mostly have additive effects (Kortenkamp 2014). The rest of the compounds act independently since synergy generally does not occur at low dose

(Escher et al. 2020). Assuming receptor activation is proportional to ligand binding affinity, the overall effect of a mixture on a receptor can be estimated by integrating the intensity across features.

To summarize, BBM provides a way to bypass well-known bottlenecks in annotation while tackling the challenges in low dose and mixture exposures. Additionally, it also fits into emerging research frameworks to speed up discovery in biomedical and etiological research (Appendix 2).

Envisioning the next generation exposomics

From another perspective, BBM is essentially an approach to map the functional exposome to loci for expression in the genome. This mapping strengthens functional analysis of exposures because biological inference is both statistical and mechanistic — in this case involves biophysical interactions with receptors of known biological function.

HRMS generates accurate mass for library matching. Even so, multiple compound matches for a single feature is common and the ambiguity makes sharing the masses of features hardly useful (Uppal et al. 2016). Because binding is structurally specific to each receptor, in addition to mass and spectrum, BBM can use mapping to receptors as an orthogonal parameter suitable for assigning unique “functional exposome IDs” (similar to reference SNP IDs in GWASs). We claim that creating a catalog of distinct and

non-redundant entries of known chemicals and unknown features in a central sharing platform opens new ways to interpret the exposures and permit systematic comparisons *across* studies (Figure 3D).

Crowdsourcing the unknown exposures in the catalog

The exposome is a growing entity. Since 1800, new chemical compounds have been synthesized at a stable 4.4% annual growth rate (Llanos et al. 2019). Still, it is possible to gain additional insights on the unknown exposures with an open-access and crowdsourced catalog. First, unknown exposures could be systematically analyzed across multiple diseases. Assuming exposure *A* binds to receptor *K*. If *A* was associated with two diseases in the same category (e.g., both are neurological), then the exposure-receptor pair could be mechanistically related to disease development. The exposures with broader biological significance could be priority exposures while those with high specificity to a particular disease could play a causal role. Second, it is possible to glean the spatial patterns and temporal variability of the exposures from multiple studies and establish background level for comparison (Vermeulen et al. 2020). We can ask questions such as: is it a long term exposure? Is there seasonality? Does the prevalence of an unknown exposure match that of a disease? Reproducibility of the association between an unknown and a disease could also be assessed. Third, research community can coordinate their resources to conduct structural identification of unknown features. Priority and importance of features can be assigned by voting, and the catalog can be updated with latest identification results. In short, the catalog could reinforce causal inference (Ioannidis 2016) and

coalesce knowledge gained on the roles of known and unknown exposures from field-specific sub-exposome studies.

Fine mapping of exposures to mixtures

We foresee that in the coming decade, conventional targeted and untargeted measurement approaches will continue to be harnessed in observational studies because they will still be cheaper than BBM. Nevertheless, conventional studies could, we predict, indirectly benefit from BBM when the significant signals or targeted interests overlap with the records in the catalog.

Fine mapping of associated or selected exposures to meaningful mixtures may have three implications. First, it provides biological interpretation of the statistical associations of EWASs. For example, are the exposures associated with a disease mostly the ligands of a particular receptor? Second, the approach may guide the design of hypothesis-driven or experimental mixture study. Conventional targeted studies test a predefined exposure-disease relationship. By using the predefined exposure as a marker, the catalog may provide a set of meaningful mixtures for further investigation. Third, the approach may facilitate rare exposure analysis. Most of the EWASs include a data filtering step to remove rare exposures prior to statistical analysis. Assuming the exposures binding to the same receptor or same category of receptor have similar biological effects, their signal could be combined and thus increasing the statistical power to detect a potential effect.

Technology development

There are many different functional protein targets for BBM, such as plasma proteins, transport proteins, chemoreceptors, and enzymes. At present, there are 48 nuclear receptors (Mangelsdorf et al. 1995) and roughly 750 G protein-coupled receptors (Vassilatis et al. 2003). A small subset of receptors (e.g. estrogen and androgen receptors) (Lee et al. 2013) relevant to some of the ubiquitous synthetic or natural xenoestrogens such as genistein, phthalates, and bisphenol A could be selected for a proof of concept study and to identify challenges for follow-up (Appendix 1). In the future, successful development of a “bio-bait” microarray will lower the requirements for large-scale and routine profiling of the functional exposome, such as in the National Health and Nutrition Examination Survey (Sobus Jon R. et al. 2015). Although the measurement technologies for BBM can be borrowed from other fields, significant investment and concerted efforts are still needed for optimizing sensitivity, scaling up capacity, setting up data collection and sharing standards, and assessing reproducibility.

Summary

We acknowledge the monumental advancements made in exposomics in the past 10 years, which focused largely on answering "how" questions, such as how to get more accurate and convenient measurements and how a molecular or complex exposures can exert effects on the development of chronic diseases. In contrast, biology-driven exposomics focuses on "what" questions to help scale the

search for causal exposures. In this essay, we articulate how conceptually AP-MS can address well-known barriers in data-driven studies. Creating a biophysical map of the exposome to guide mixture studies will help to lower the barrier of entry for exposome research between disciplines in biomedical sciences.

In the era of next generation exposomics, health researchers will have accessible tools to make a balance between answering "how" and "what" questions. In our view, through tighter integration with the genome and other subcellular networks, we will have better capability to study the totality of exposures and decipher the known unknowns and unknown unknowns in the exposome. We need to invest and embrace a new technological approach that can bring us closer to causality. It is time to revisit a fundamental question: what are the environmental causes of complex diseases?

Figure

Figure 1. Approaches to study the exposure and response relationship. Exposure can be broadly divided into four distinct domains, including living environment, socioeconomic factors, lifestyle, and physical-chemical exposures. Effects of exposures can be studied in ecological, *in vivo*, *in vitro*, molecular, and *in silico* levels. Although exposure and response are heterogeneous, almost all causal biological effect from an exposure is explainable by the underlying molecular connections. To investigate a relationship, traditional study takes a forward approach — from exposure to response. For example, a hypothesis-driven design with a few predefined exposures or a semi/fully data-driven design to conduct exploratory studies. In contrast, studies using biology-based measurement (BBM) approach work differently — from response to exposure — at the molecular level. The “bait” to capture the exposures can be a few predefined, a particular class, or a broad-scale selection of functional biomolecules.

Figure 2. Using a biology-based measurement approach to characterize the functional exposome. The assay is implemented in the form of affinity purification–mass spectrometry. 1) A group of individuals (n) is sampled. 2) Biofluids such as plasma is isolated from blood and 3) loaded to the microplate (each well is coated with one type of receptor) to capture circulating ligands. 4) After incubation, ligands are extracted and 5) characterized by mass spectrometry. 6) Data can be presented in three different matrices (n by feature, n by receptor, and receptor by feature) with signals integrating across features or receptors for downstream analyses. A key advantage of this approach is that functional analysis of the exposome is possible without upfront feature identification because molecular exposures are biophysically linked to receptors with known functions.

Figure 3. Key advantages of using a biology-based measurement approach to study the functional exposome (FE), where nature meets nurture. A) Enrichment is not required as an extra step but intrinsically provided by the pull-down assay format. B) Each detected feature is functionally annotated by the corresponding binding receptors. C) Health impacts can be assessed in two ways: co-exposure correlation or shared mode of action. D) Each feature can be assigned to a unique reference FE ID, which is characterized by an accurate mass, the functional molecules (e.g., binding receptors), and other attributes such as mass spectra and detection frequency in the population. The mapping, cataloging, and sharing of exposure information allow unambiguous communication of knowns and unknowns across studies.

Appendix

Appendix 1. Opportunities and limitations of using biology-based measurement approach in exposomics studies

Opportunities	Limitations and uncertainties
1. The functional exposome is mapped to the genome by biophysical interactions in the data generation step, providing mechanistic insights for statistically associated relationships	1. Not all receptors can be directly used without further assay development (e.g., transmembrane receptors)
2. Workflow is modular and capable of switching to different types of functional biomolecules (e.g., DNA, RNA, protein) and different omics platforms to measure the ligands (e.g., proteomics for peptide ligands)	2. Compared to the gold standard of measurement (liquid chromatography tandem mass spectrometry), detection rate in the general population is still an unknown
3. Approach leverages existing biochemical and mass spectrometry technologies. Minimal development is needed	3. Exact sample volume requirement per person is uncertain but likely in the range of 300 to 600 μ L
4. Approach is tailored to discovery study in exposomics; Data is used to drive exposome-wide association study for finding new exposures associating with diseases.	4. Compound confirmation of the unknowns is still needed to link sources and additional information from the literature
	5. Non-host compounds are expected to be dominating in the measurement but awaits characterization and confirmation

Appendix 2. BBM and emerging research frameworks

Traditional epidemiological studies link observable exposure and disease and BBM generates data in the center of the exposure-disease continuum. There are four emerging research frameworks that make use of the molecular evidence to assist causal inference and prediction of biological effects. 1) Adverse outcome pathways: a concept used in risk assessment to enhance the utility of pathway-based molecular information for assessing hazards to human health (Vinken et al. 2017). 2) Developmental origins of health and disease: a hypothesis to illuminate potential non-genetic developmental factors that occur in early life and lead to the development of chronic diseases (Gluckman et al. 2016). 3) Meet-in-the-middle approach: a study design that aims to reinforce causality in observational studies (Vineis et al. 2013). The goal is to search for biomarkers that connect retrospectively to an external exposure and prospectively to a disease outcome. 4) Multi-omics in observational cohort studies: complex diseases are a manifestation of minuscule effects caused by multiple factors. Integrating multiple omics datasets and modeling the intricate biological networks can obtain mechanistic insights into the pathophysiology of diseases (Rose et al. 2019; Zhou et al. 2019).

References

- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651–1656; doi:10.1126/science.2047873.
- Agier L, Portengen L, Chadeau-Hyam M, Basagaña X, Giorgis-Allemand L, Siroux V, et al. 2016. A Systematic Comparison of Linear Regression-Based Statistical Methods to Assess Exposome-Health Associations. *Environ Health Perspect* 124:1848–1856; doi:10.1289/EHP172.
- Andra SS, Austin C, Patel D, Dolios G, Awawda M, Arora M. 2017. Trends in the application of high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the environmental chemical space of the human exposome. *Environ Int* 100:32–61; doi:10.1016/j.envint.2016.11.026.
- Annis DA, Nickbarg E, Yang X, Ziebell MR, Whitehurst CE. 2007. Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Curr Opin Chem Biol* 11:518–526; doi:10.1016/j.cbpa.2007.07.011.
- Balik-Meisner Michele, Truong Lisa, Scholl Elizabeth H., La Du Jane K., Tanguay Robert L., Reif David M. 2018. Elucidating Gene-by-Environment Interactions Associated with Differential Susceptibility to Chemical Exposure. *Environ Health Perspect* 126:067010; doi:10.1289/EHP2662.
- Birnbaum Linda S. 2018. Updating the NIEHS Strategic Plan. *Environ Health Perspect* 125:071001; doi:10.1289/EHP2502.
- Birnbaum LS. 2012. NIEHS's New Strategic Plan. *Environ Health Perspect* 120:a298–a298; doi:10.1289/ehp.1205642.
- Board on Life Sciences, Division on Earth and Life Studies, National Academies of Sciences, Engineering, and Medicine. 2016. *Use of Metabolomics to Advance Research on Environmental Exposures and the Human Exposome: Workshop in Brief*. National Academies Press (US):Washington (DC).
- Brack W, Ait-Aissa S, Burgess RM, Busch W, Creusot N, Di Paolo C, et al. 2016. Effect-directed analysis supporting monitoring of aquatic environments — An in-depth overview. *Sci Total Environ* 544:1073–1118; doi:10.1016/j.scitotenv.2015.11.102.
- Carlsson H, Rappaport SM, Törnqvist M. 2019. Protein Adductomics: Methodologies for Untargeted Screening of Adducts to Serum Albumin and Hemoglobin in Human Blood Samples. *High-Throughput* 8:6; doi:10.3390/ht8010006.
- Carnero A. 2006. High throughput screening in drug discovery. *Clin Transl Oncol Off Publ Fed Span Oncol Soc Natl Cancer Inst Mex* 8:482–490; doi:10.1007/s12094-006-0048-2.
- Chung MK, Grigoryan H, Iavarone AT, Rappaport SM. 2014. Antibody enrichment and mass spectrometry of albumin-cys34 adducts. *Chem Res Toxicol* 27:400–407; doi:10.1021/tx400337k.
- Chung MK, Kannan K, Louis GM, Patel CJ. 2018. Toward capturing the exposome: exposure biomarker variability and coexposure patterns in the shared environment. *Environ Sci Technol* 52:8801–8810; doi:10.1021/acs.est.8b01467.
- Czene K, Lichtenstein P, Hemminki K. 2002. Environmental and heritable causes of cancer among 9.6

- million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 99:260–266; doi:10.1002/ijc.10332.
- Dennis KK, Auerbach SS, Balshaw DM, Cui Y, Fallin MD, Smith MT, et al. 2016. The Importance of the Biological Impact of Exposure to the Concept of the Exposome. *Environ Health Perspect* 124:1504–1510; doi:10.1289/EHP140.
- Dunham WH, Mullin M, Gingras A-C. 2012. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *Proteomics* 12:1576–1590; doi:10.1002/pmic.201100523.
- Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 6:1060–1083; doi:10.1038/nprot.2011.335.
- E. Kempa E, A. Hollywood K, A. Smith C, E. Barran P. 2019. High throughput screening of complex biological samples with mass spectrometry – from bulk measurements to single cell analysis. *Analyst* 144:872–891; doi:10.1039/C8AN01448E.
- Escher BI, Stapleton HM, Schymanski EL. 2020. Tracking complex mixtures of chemicals in our changing environment. *Science* 367:388–392; doi:10.1126/science.aay6636.
- Filer D, Patisaul HB, Schug T, Reif D, Thayer K. 2014. Test driving ToxCast: endocrine profiling for 1858 chemicals included in phase II. *Curr Opin Pharmacol* 19:145–152; doi:10.1016/j.coph.2014.09.021.
- Florentinus-Mefailoski A, Safi F, Marshall JG. 2014. Enzyme Linked Immuno Mass Spectrometric Assay (ELIMSA). *J Proteomics* 96:343–352; doi:10.1016/j.jprot.2013.11.022.
- Gluckman PD, Buklijas T, Hanson MA. 2016. Chapter 1 - The Developmental Origins of Health and Disease (DOHaD) Concept: Past, Present, and Future. In: *The Epigenome and Developmental Origins of Health and Disease* (C.S. Rosenfeld, ed). Academic Press:Boston. 1–15.
- Go Y-M, Walker DI, Liang Y, Uppal K, Soltow QA, Tran V, et al. 2015a. Reference Standardization for Mass Spectrometry and High-resolution Metabolomics Applications to Exposome Research. *Toxicol Sci* 148:531–543; doi:10.1093/toxsci/kfv198.
- Go Y-M, Walker DI, Soltow QA, Uppal K, Wachtman LM, Strobel FH, et al. 2015b. Metabolome-wide association study of phenylalanine in plasma of common marmosets. *Amino Acids* 47:589–601; doi:10.1007/s00726-014-1893-x.
- Grigoryan H, Edmands W, Lu SS, Yano Y, Regazzoni L, Iavarone AT, et al. 2016. Adductomics pipeline for untargeted analysis of modifications to Cys34 of human serum albumin. *Anal Chem* 88:10504–10512; doi:10.1021/acs.analchem.6b02553.
- Ioannidis JPA. 2016. Exposure-wide epidemiology: revisiting Bradford Hill: Bradford Hill criteria revisited. *Stat Med* 35:1749–1762; doi:10.1002/sim.6825.
- Ioannidis JPA. 2008. Why Most Discovered True Associations Are Inflated: *Epidemiology* 19:640–648; doi:10.1097/EDE.0b013e31818131e7.
- Ioannidis JPA. 2005. Why Most Published Research Findings Are False. *PLOS Med* 2:e124;

- doi:10.1371/journal.pmed.0020124.
- Ioannidis JPA, Loy EY, Poulton R, Chia KS. 2009. Researching Genetic Versus Nongenetic Determinants of Disease: A Comparison and Proposed Unification. *Sci Transl Med* 1:7ps8-7ps8; doi:10.1126/scitranslmed.3000247.
- James G, Witten D, Hastie T, Tibshirani R. 2017. *An Introduction to Statistical Learning: with Applications in R*. 1st ed. 2013, Corr. 7th printing 2017 edition. Springer:New York.
- Jamin EL, Bonvallot N, Tremblay-Franco M, Cravedi J-P, Chevrier C, Cordier S, et al. 2014. Untargeted profiling of pesticide metabolites by LC–HRMS: an exposomics tool for human exposure evaluation. *Anal Bioanal Chem* 406:1149–1161; doi:10.1007/s00216-013-7136-2.
- Jiang C, Wang X, Li X, Inlora J, Wang T, Liu Q, et al. 2018. Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring. *Cell* 175:277-291.e31; doi:10.1016/j.cell.2018.08.060.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28:27–30; doi:10.1093/nar/28.1.27.
- Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. 2016. PubChem Substance and Compound databases. *Nucleic Acids Res* 44:D1202–D1213; doi:10.1093/nar/gkv951.
- Klaassen C. 2018. *Casarett & Doull's Toxicology: The Basic Science of Poisons, 9th Edition*. 9th edition. McGraw-Hill Education / Medical:New York.
- Kortenkamp A. 2014. Low dose mixture effects of endocrine disrupters and their implications for regulatory thresholds in chemical risk assessment. *Curr Opin Pharmacol* 19:105–111; doi:10.1016/j.coph.2014.08.006.
- Lakhani CM, Tierney BT, Manrai AK, Yang J, Visscher PM, Patel CJ. 2019. Repurposing large health insurance claims data to estimate genetic and environmental contributions in 560 phenotypes. *Nat Genet* 51:327; doi:10.1038/s41588-018-0313-7.
- Lanucara F, Holman SW, Gray CJ, Eyers CE. 2014. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* 6:281–294; doi:10.1038/nchem.1889.
- Lazarevic Nina, Barnett Adrian G., Sly Peter D., Knibbs Luke D. 2019. Statistical Methodology in Studies of Prenatal Exposure to Mixtures of Endocrine-Disrupting Chemicals: A Review of Existing Approaches and New Alternatives. *Environ Health Perspect* 127:026001; doi:10.1289/EHP2207.
- Lee H-R, Jeung E-B, Cho M-H, Kim T-H, Leung PCK, Choi K-C. 2013. Molecular mechanism(s) of endocrine-disrupting chemicals and their potent oestrogenicity in diverse cells and tissues that express oestrogen receptors. *J Cell Mol Med* 17:1–11; doi:10.1111/j.1582-4934.2012.01649.x.
- Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380:2224–2260; doi:10.1016/S0140-6736(12)61766-8.
- Lippmann M. 2013. Exposure science in the 21st century: a vision and a strategy. *J Expo Sci Environ*

- Epidemiol 23:1–1; doi:10.1038/jes.2012.109.
- Llanos EJ, Leal W, Luu DH, Jost J, Stadler PF, Restrepo G. 2019. Exploration of the chemical space and its three historical regimes. *Proc Natl Acad Sci* 116:12660–12665; doi:10.1073/pnas.1816039116.
- Lu Y, Wang Y, Ong C-N, Subramaniam T, Choi HW, Yuan J-M, et al. 2016. Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. *Diabetologia* 59:2349–2359; doi:10.1007/s00125-016-4069-2.
- Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, et al. 1995. The nuclear receptor superfamily: The second decade. *Cell* 83:835–839; doi:10.1016/0092-8674(95)90199-X.
- Mayr LM, Bojanic D. 2009. Novel trends in high-throughput screening. *Curr Opin Pharmacol* 9:580–588; doi:10.1016/j.coph.2009.08.004.
- Medina-Cleghorn D, Bateman LA, Ford B, Heslin A, Fisher KJ, Dalvie ED, et al. 2015. Mapping Proteome-Wide Targets of Environmental Chemicals Using Reactivity-Based Chemoproteomic Platforms. *Chem Biol* 22:1394–1405; doi:10.1016/j.chembiol.2015.09.008.
- Miller GW, Jones DP. 2014. The Nature of Nurture: Refining the Definition of the Exposome. *Toxicol Sci* 137:1–2; doi:10.1093/toxsci/kft251.
- Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermanic P, Greninger AL, et al. 2014. Affinity purification–mass spectrometry and network analysis to understand protein-protein interactions. *Nat Protoc* 9:2539–2554; doi:10.1038/nprot.2014.164.
- Neveu V, Moussy A, Rouaix H, Wedekind R, Pon A, Knox C, et al. 2017. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res* 45:D979–D984; doi:10.1093/nar/gkw980.
- Nisbet ICT, LaGoy PK. 1992. Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regul Toxicol Pharmacol* 16:290–300; doi:10.1016/0273-2300(92)90009-X.
- Osborn MP, Park Y, Parks MB, Burgess LG, Uppal K, Lee K, et al. 2013. Metabolome-wide association study of neovascular age-related macular degeneration. *PloS One* 8:e72737; doi:10.1371/journal.pone.0072737.
- Patel CJ. 2017. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. *Curr Epidemiol Rep* 4:22–30; doi:10.1007/s40471-017-0100-5.
- Patel CJ, Bhattacharya J, Butte AJ. 2010. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLOS ONE* 5:e10746; doi:10.1371/journal.pone.0010746.
- Patel CJ, Ioannidis JPA. 2014. Studying the elusive environment in large scale. *JAMA* 311:2173–2174; doi:10.1001/jama.2014.4129.
- Patel CJ, Manrai AK. 2015. Development of exposome correlation globes to map out environment-wide associations. *Pac Symp Biocomput Pac Symp Biocomput* 231–242.
- Peters A, Nawrot TS, Baccarelli AA. 2021. Hallmarks of environmental insults. *Cell* 0; doi:10.1016/j.cell.2021.01.043.

- Polubriaginof FCG, Vanguri R, Quinnes K, Belbin GM, Yahi A, Salmasian H, et al. 2018. Disease Heritability Inferred from Familial Relationships Reported in Medical Records. *Cell* 173:1692–1704.e11; doi:10.1016/j.cell.2018.04.032.
- Quinn RA, Nothias L-F, Vining O, Meehan M, Esquenazi E, Dorrestein PC. 2017. Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol Sci* 38:143–154; doi:10.1016/j.tips.2016.10.011.
- Rappaport SM. 2012. Biomarkers intersect with the exposome. *Biomark Biochem Indic Expo Response Susceptibility Chem* 17:483–489; doi:10.3109/1354750X.2012.691553.
- Rappaport SM. 2016. Genetic Factors Are Not the Major Causes of Chronic Diseases. *PLOS ONE* 11:e0154387; doi:10.1371/journal.pone.0154387.
- Rappaport SM. 2018. Redefining environmental exposure for disease etiology. *Npj Syst Biol Appl* 4:1–6; doi:10.1038/s41540-018-0065-0.
- Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. 2014. The blood exposome and its role in discovering causes of disease. *Environ Health Perspect* 122:769–774; doi:10.1289/ehp.1308015.
- Rappaport SM, Li H, Grigoryan H, Funk WE, Williams ER. 2012. Adductomics: Characterizing exposures to reactive electrophiles. *Toxicol Lett* 213:83–90; doi:10.1016/j.toxlet.2011.04.002.
- Rappaport SM, Smith MT. 2010. Environment and Disease Risks. *Science* 330:460–461; doi:10.1126/science.1192603.
- Rinschen MM, Ivanisevic J, Giera M, Siuzdak G. 2019. Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol* 20:353–367; doi:10.1038/s41580-019-0108-4.
- Roca M, Leon N, Pastor A, Yusà V. 2014. Comprehensive analytical strategy for biomonitoring of pesticides in urine by liquid chromatography–orbitrap high resolution mass spectrometry. *J Chromatogr A* 1374:66–76; doi:10.1016/j.chroma.2014.11.010.
- Roede JR, Uppal K, Park Y, Lee K, Tran V, Walker D, et al. 2013. Serum Metabolomics of Slow vs. Rapid Motor Progression Parkinson's Disease: a Pilot Study. *PLOS ONE* 8:e77629; doi:10.1371/journal.pone.0077629.
- Rose SMS-F, Contrepois K, Moneghetti KJ, Zhou W, Mishra T, Mataraso S, et al. 2019. A longitudinal big data approach for precision health. *Nat Med* 25:792–804; doi:10.1038/s41591-019-0414-6.
- Seethala R, Zhang L, eds. 2016. *Handbook of Drug Screening*. 2 edition. CRC Press:New York, USA.
- Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM, et al. 2013. Profiling 976 ToxCast Chemicals across 331 Enzymatic and Receptor Signaling Assays. *Chem Res Toxicol* 26:878–895; doi:10.1021/tx400021f.
- Sobus Jon R., DeWoskin Robert S., Tan Yu-Mei, Pleil Joachim D., Phillips Martin Blake, George Barbara Jane, et al. 2015. Uses of NHANES Biomarker Data for Chemical Risk Assessment: Trends, Challenges, and Opportunities. *Environ Health Perspect* 123:919–927; doi:10.1289/ehp.1409177.
- Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. 2007. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG)

- Metabolomics Standards Initiative (MSI). *Metabolomics* 3:211–221; doi:10.1007/s11306-007-0082-2.
- Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. 2019. Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 20:467–484; doi:10.1038/s41576-019-0127-1.
- Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. 2012. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal Chem* 84:5035–5039; doi:10.1021/ac300698c.
- Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. 2016. Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop. *Environ Health Perspect* 124; doi:10.1289/EHP547.
- Tian Z, Zhao H, Peter KT, Gonzalez M, Wetzel J, Wu C, et al. 2021. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. *Science* 371:185–189; doi:10.1126/science.abd6951.
- Turner MC, Vineis P, Seleiro E, Dijmarescu M, Balshaw D, Bertollini R, et al. 2018. EXPOsOMICS: final policy workshop and stakeholder consultation. *BMC Public Health* 18:260; doi:10.1186/s12889-018-5160-z.
- Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, Yu T, et al. 2013. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics* 14:15; doi:10.1186/1471-2105-14-15.
- Uppal K, Walker DI, Liu K, Li S, Go Y-M, Jones DP. 2016. Computational Metabolomics: A Framework for the Million Metabolome. *Chem Res Toxicol* 29:1956–1975; doi:10.1021/acs.chemrestox.6b00179.
- Van den Berg M, Birnbaum L, Bosveld A T, Brunström B, Cook P, Feeley M, et al. 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ Health Perspect* 106:775–792; doi:10.1289/ehp.98106775.
- Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, et al. 2003. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci* 100:4903–4908; doi:10.1073/pnas.0230374100.
- Vermeulen R, Schymanski EL, Barabási A-L, Miller GW. 2020. The exposome and health: Where chemistry meets biology. *Science* 367:392–396; doi:10.1126/science.aay3164.
- Vineis P, Chadeau-Hyam M, Gmuender H, Gulliver J, Herceg Z, Kleijnans J, et al. 2017. The exposome in practice: Design of the EXPOsOMICS project. *Int J Hyg Environ Health* 220:142–151; doi:10.1016/j.ijheh.2016.08.001.
- Vineis P, van Veldhoven K, Chadeau-Hyam M, Athersuch TJ. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology: Application of Omics in Environmental Epidemiology. *Environ Mol Mutagen* 54:461–467; doi:10.1002/em.21764.
- Vinken M, Knapen D, Vergauwen L, Hengstler JG, Angrish M, Whelan M. 2017. Adverse outcome pathways: a concise introduction for toxicologists. *Arch Toxicol* 91:3697–3707; doi:10.1007/s00204-017-2020-z.

- Vos RCD, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2:778–791; doi:10.1038/nprot.2007.95.
- Vrijheid Martine, Slama Rémy, Robinson Oliver, Chatzi Leda, Coen Muireann, van den Hazel Peter, et al. 2014. The Human Early-Life Exposome (HELIX): Project Rationale and Design. *Environ Health Perspect* 122:535–544; doi:10.1289/ehp.1307204.
- Walker DI, Pennell KD, Uppal K, Xia X, Hopke PK, Utell MJ, et al. 2016a. Pilot Metabolome-Wide Association Study of Benzo(a)pyrene in Serum From Military Personnel: *J Occup Environ Med* 58:S44–S52; doi:10.1097/JOM.0000000000000772.
- Walker DI, Uppal K, Zhang L, Vermeulen R, Smith M, Hu W, et al. 2016b. High-resolution metabolomics of occupational exposure to trichloroethylene. *Int J Epidemiol* 45:1517–1527; doi:10.1093/ije/dyw218.
- Wang Aolin, Gerona Roy R., Schwartz Jackie M., Lin Thomas, Sirota Marina, Morello-Frosch Rachel, et al. 2018. A Suspect Screening Method for Characterizing Multiple Chemical Exposures among a Demographically Diverse Population of Pregnant Women in San Francisco. *Environ Health Perspect* 126:077009; doi:10.1289/EHP2920.
- Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, et al. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472:57–63; doi:10.1038/nature09922.
- Wheelock CE, Rappaport SM. 2020. The role of gene–environment interactions in lung disease: the urgent need for the exposome. *Eur Respir J* 55:1902064; doi:10.1183/13993003.02064-2019.
- Wieser A, Schneider L, Jung J, Schubert S. 2012. MALDI-TOF MS in microbiological diagnostics—identification of microorganisms and beyond (mini review). *Appl Microbiol Biotechnol* 93:965–974; doi:10.1007/s00253-011-3783-4.
- Wild CP. 2005. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Prev Biomark* 14:1847–1850; doi:10.1158/1055-9965.EPI-05-0456.
- Wild CP. 2012. The exposome: from concept to utility. *Int J Epidemiol* 41:24–32; doi:10.1093/ije/dyr236.
- Wishart D, Arndt D, Pon A, Sajed T, Guo AC, Djoumbou Y, et al. 2015. T3DB: the toxic exposome database. *Nucleic Acids Res* 43:D928–D934; doi:10.1093/nar/gku1004.
- 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:D608–D617; doi:10.1093/nar/gkx1089.
- Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 36:D901–D906; doi:10.1093/nar/gkm958.
- Xia J, Psychogios N, Young N, Wishart DS. 2009. MetaboAnalyst: a web server for metabolomic data

- analysis and interpretation. *Nucleic Acids Res* 37:W652–W660; doi:10.1093/nar/gkp356.
- Yang D, Han J, Hall DR, Sun J, Fu J, Kutarna S, et al. 2020. Nontarget Screening of Per- and Polyfluoroalkyl Substances Binding to Human Liver Fatty Acid Binding Protein. *Environ Sci Technol* 54:5676–5686; doi:10.1021/acs.est.0c00049.
- Yu T, Park Y, Johnson JM, Jones DP. 2009. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* 25:1930–1936; doi:10.1093/bioinformatics/btp291.
- Yuan M, Breitkopf SB, Yang X, Asara JM. 2012. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat Protoc* 7:872–881; doi:10.1038/nprot.2012.024.
- Yuan T-H, Chung M-K, Lin C-Y, Chen S-T, Wu K-Y, Chan C-C. 2016. Metabolic profiling of residents in the vicinity of a petrochemical complex. *Sci Total Environ* 548–549:260–269; doi:10.1016/j.scitotenv.2016.01.033.
- Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, et al. 2019. Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* 569:663; doi:10.1038/s41586-019-1236-x.