Practice of Epidemiology

Use of the "Exposome" in the Practice of Epidemiology: A Primer on -Omic Technologies

D. Gayle DeBord*, Tania Carreón, Thomas J. Lentz, Paul J. Middendorf, Mark D. Hoover, and Paul A. Schulte

* Correspondence to Dr. D. Gayle DeBord, Robert A. Taft Laboratories, Mailstop R2, National Institute for Occupational Safety and Health, 1090 Tusculum Avenue, Cincinnati, OH 45226 (e-mail: ded4@cdc.gov).

The exposome has been defined as the totality of exposures individuals experience over the course of their lives and how those exposures affect health. Three domains of the exposome have been identified: internal, specific external, and general external. Internal factors are those that are unique to the individual, and specific external factors include occupational exposures and lifestyle factors. The general external domain includes sociodemographic factors such as educational level and financial status. Eliciting information on the exposome is daunting and not feasible at present; the undertaking may never be fully realized. A variety of tools have been identified to measure the exposome. Biomarker measurements will be one of the major tools in exposomic studies. However, exposure data can also be obtained from other sources such as sensors, geographic information systems, and conventional tools such as survey instruments. Proof-of-concept studies are being conducted that show the promise of exposomic investigation and the integration of different kinds of data. The inherent value of exposomic data in epidemiologic studies is that they can provide greater understanding of the relationships among a broad range of chemical and other risk factors and health conditions and ultimately lead to more effective and efficient disease prevention and control.

biomarkers; environmental exposures; epidemiologic methods; occupational exposures

Abbreviations: EWAS, environment-wide association studies; GIS, geographic information systems; GWAS, genome-wide association studies; HEALS, Health and Environment-Wide Associations Based on Large Population Surveys; HELIX, Human Early-Life Exposome; HERCULES, Health and Exposome Research Center: Understanding Lifetime Exposures; NHANES, National Health and Nutrition Examination Survey.

WHAT IS THE EXPOSOME?

In 2005, Wild (1) defined the *exposome* as the totality of exposures individuals experience from conception until death and its impact on risk of chronic diseases. (A glossary of terms and definitions used in this article is provided in Appendix Table 1.) Exposures can include toxicants in the general environment and in workplaces, diet, lifestyle choices, and even socioeconomic status (Figure 1). People have unique characteristics that might make them more or less susceptible to stressors in their environment. A person's genetics, epigenetics, health status, and physiology, as well as changes in these personal components caused by previous exposures,

can influence the effects of new or present exposures. For example, metabolic pathways can be disrupted, which changes susceptibility to an insult or a disease.

The premise envisioned with the exposome concept was that the exposome is complementary to the genome and that an integrated understanding of both the genome and the exposome would contribute to synergistically addressing chronic human health issues. The science of epidemiology is the primary means of understanding the exposome and its interaction with health status. Research suggests that environmental exposures have a much greater impact on health and disease than genetic factors alone (2). The inherent value of exposomic approaches and data in epidemiologic studies is
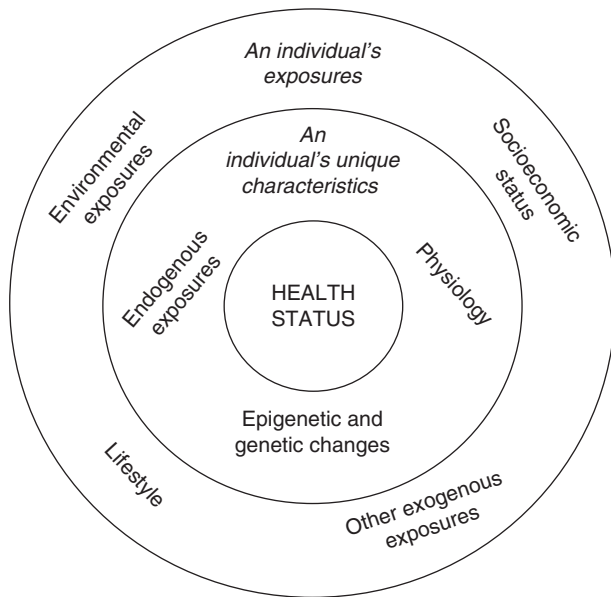
Figure 1. The concept of the exposome. An individual has many sources of exposure over the course of a lifetime. How those exposures are modulated and affect health depends on the individual's unique characteristics.

The exposome concept was further refined by Wild (3) to include 3 broad domains: internal, specific external, and general external. Internal factors are those that are specific to the individual, such as physiology, age, body morphology, and the individual's genome. Specific external factors include diet and occupational and environmental exposures, as well as physical, biological, and physiological exposures. The third domain, general external factors, includes broader social constructs such as home location, educational level, and socioeconomic status. Wild noted that the domains can be viewed as both overlapping and intertwining, and that it is sometimes difficult to place a particular exposure into one domain or another (3). For example, he observed that one can debate whether physical activity should be in the internal domain or the specific external domain. A comprehensive and informative assessment of exposure can be achieved by combining aspects of the 3 domains in ways that can be used to guide the design, conduct, and interpretation of epidemiologic studies.

An individual's exposome is dynamic, which makes measuring the exposome challenging. Several critical life stages have been identified for which some exposures may have a greater impact with respect to future diseases. For example, the fetus or young child has rapidly growing cells and immature cell repair processes (4). Therefore, exposures experienced at young ages may have significant influence over future health. Embryos or infants are rapidly maturing and may not have all the protective mechanisms in place to repair damage incurred from an exposure. For example, in the 1950s and 1960s, some women were given diethylstilbestrol to prevent miscarriage (5). Their offspring, who were exposed in utero, have increased risk of reproductive tract cancers, decreased fertility, and difficult pregnancies (5, 6). The susceptibility of children and teenagers may differ from that of adults because they are still developing, have immature cell repair processes, and have different hormone levels than

that they provide a greater understanding of the relationships among exposures and health conditions and ultimately could lead to prevention of chronic diseases. Epidemiologic research both can utilize exposomic data in health and disease research and can be a means of understanding the exposome (Figure 2).
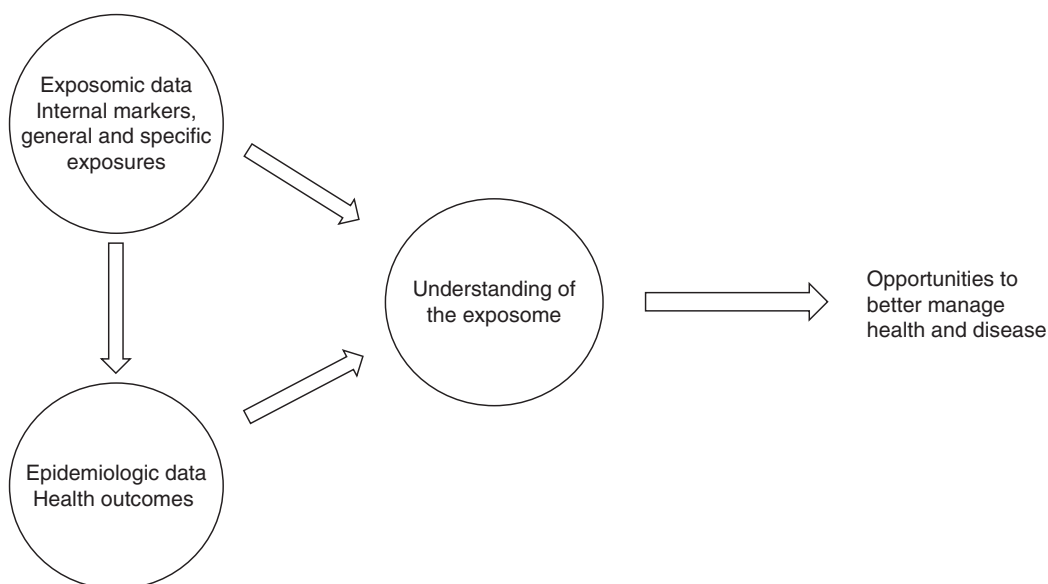

Figure 2. Exposome-informed epidemiologic research.

adults. As a result of these early-life exposures, susceptibility to disease caused by later exposures may be increased. Additionally, certain exposures can occur throughout different life stages that are critically important as well (7). In utero exposures would mainly be due to maternal diet, medication use, or environmental/occupational exposures. Throughout their lifetime, individuals would have a steady-state exposure to some ambient agents such as allergens (7). Some xenobiotics may bioaccumulate, resulting in a higher body burden with age. Occupational exposures would occur mainly during the working years, and as people age, exposure to pharmaceuticals tends to increase.

Most epidemiologic studies get a snapshot look at exposures that affect health. Although the exposomic approach may allow for a broader view of exposure, the ability to measure past exposures to any great extent is limited; and measuring each agent to which a person may have been exposed at any given time is not feasible at present or in the foreseeable future. Depending upon the study hypotheses, measurement of specific agents will be important, along with a holistic exposure assessment approach that includes aspects of the 3 areas of the exposome described by Wild (3).

The challenges of measuring the complete exposome are daunting. Newer technologies such as -omics, sensors, and geographic or spatial information systems are allowing for a more comprehensive understanding of the exposome. While the exposome is more likely to be useful in epidemiologic studies, it may also have other clinical or public health utility (8, 9). That utility could range from personalized medicine to improved risk assessment for regional exposures to chemicals, and will necessarily be dependent on the extent to which exposomic indicators are validated. Doing that will require additional epidemiologic studies.

While Wild's definition of the exposome was developed to draw attention to the critical need for more complete environmental exposure assessment in epidemiologic studies, other authors have broadened the definition of the exposome to include additional factors such as behavior. Miller and Jones defined the exposome as "the cumulative measure of environmental influences and associated biological responses throughout the lifespan, including exposures from the environment, diet, behavior, and endogenous processes" (10, p. 2). Their definition expands on that of Wild to consider cumulative biological responses as well as endogenous processes. Ultimately, health status is influenced by the interaction of an individual's environmental and genetic factors from conception to death (Figure 3). The explanation of both the exposome and the genome in a manner that can protect and promote health is important, and there is need to pursue both.
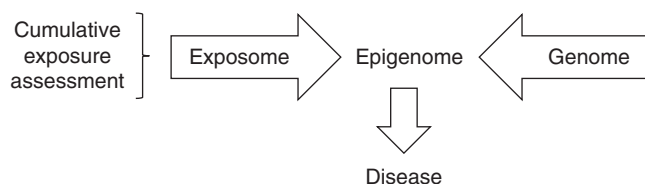

Figure 3. Example of an exposome-genome interaction.

WHAT HAS BEEN DONE IN THE FIELD?

Utilization of exposomic data in epidemiologic studies and surveillance efforts has occurred because of advancements in exposure science. Yet these advances also indicate the complexity of factors associated with total exposure assessment and the development of appropriate and effective risk management strategies. More importantly, though, they help to determine what questions are most pertinent to performing comprehensive exposure assessments. Among the questions to be considered, the growing scientific literature on this topic suggests the following (11):

- Which mixtures are most important from a public or occupational health perspective?
- What is the nature (i.e., duration, frequency, and timing) and magnitude (e.g., exposure concentration and dose) of relevant exposures for the population of interest?
- What is the mechanism (e.g., toxicokinetic or toxicodynamic) and consequence (e.g., additive, less than additive, more than additive) of the mixture's interactive effects on exposed populations?

Exposure science is the discipline that studies the conditions for contact with toxicants, characterizing the quality and quantity of the toxicant from its sources to its transport to and receipt by or interaction with the human body (12, 13). Regarding the evolution of exposure science (defined below) as the means for bridging the disciplines of environmental science and environmental health science, Lioy (12) poses the following:

- What does one do with such exposure information (i.e., understanding variables that define contact with environmental stressors and the factors that influence the contact)?
- What role does exposure science play in situations beyond observational analyses and interpretation?

Addressing these questions through exposome-informed advances in exposure science and risk assessment methods will provide a foundation for the development of improved tools for total exposure assessment and risk management.

In exploring implications for exposure science focusing on the exposome, Rappaport (2) advocates utilizing biomonitoring (e.g., blood sampling and other internal measures of dose) rather than focusing primarily on sampling exposures in food, water, and air. He also suggests the importance of better integration of these biomonitoring measures with environmental exposure measures to advance the field of exposure science and the understanding of the exposome as a means of characterizing and controlling detrimental exposures. Pleil has suggested that biomarkers can be used to "assess the sustainability of the environmental conditions with respect to human health" (14, p. 264).

There are 4 conventions in the literature for characterizing environmental biomarkers and how they fit into categories of grouping schemes. These 4 conventions are origin, function, kinetics, and medium (14). From the categories of biomarkers, a sequence of data management strategies might be applied to recognize patterns and statistics within the exposome which provide insight into the exposure–dose-response relationship involving environmental stressors.

Exposomic data may contribute to determining why some people will develop a disease while others with the same or greater exposure will not. A key factor in describing the exposome is the ability to accurately measure exposures and their effects on human health. However, the evolution and maturity of the science of exposomics can be viewed as a practical extension of the principles of total exposure assessment, which ultimately seeks to inform risk management strategies.

The US Environmental Protection Agency has been at the forefront of research in total exposure assessment. Led by the agency's Office of Research and Development, the Cumulative Communities Research Program "focuses on exposure tools for advancing the science and understanding of cumulative risk to communities and individuals" (15, p. 353). There are multiple factors driving this approach for community (i.e., nonoccupational) exposures, but most can be tied to the motivation that people want to know about the multiple stressors (e.g., pollutants) to which they are exposed, what the associated health risks are, and how these exposures and related risks can be prevented or reduced (16).

In a similar national-level surveillance effort, the French National Network for Monitoring and Prevention of Occupational Diseases (RNV3P) established a database for recording and tracking occupational health exposures and related adverse health effects (17, 18). A goal of the network and database is to provide better characterization of occupational disease-exposure relationships, thereby exploring a theoretical framework of the occupational exposome.

FUNDING INITIATIVES AND RESOURCES

National Institute of Environmental Health Sciences

The National Institute of Environmental Health Sciences has long-supported the concept of the exposome, funding studies that are now being defined as exposomic in nature. The Institute recently established a strategic goal of transforming exposure science and has identified exposomics as a possible approach (19). It plans to advance characterizations of environmental exposure assessment at both the individual and population levels, which will be accomplished through tools and technologies for multiscale measurements.

Recently, the Institute funded the Health and Exposome Research Center: Understanding Lifetime Exposures (HERCULES) Project at Emory University (http://emoryhercules.com/) with the stated goal of understanding lifetime exposures. The main aims of HERCULES are: 1) to provide greater access to exposome-related approaches such as systems biology, metabolomics, high-throughput toxicological screening, and spatial and temporal statistical models; 2) to facilitate communication of the importance of environmental factors in disease using exposomic principles; and 3) to expedite translation of novel scientific findings to develop new sustainability, prevention, or treatment strategies in humans.

Human Early-Life Exposome

The Human Early-Life Exposome (HELIX) Project (http://www.projecthelix.eu/) is a European collaboration that has been established as a proof-of-concept study to characterize children's exposomes as they progress through early life (20).

The project involves 13 partner institutions and will use data from 6 ongoing, prospective European birth cohort studies of mothers and children living in Spain, France, the United Kingdom, Norway, Greece, and Lithuania. Traditional methods are being used for exposure assessment, as well as biomarker and -omic measures to assess the exposome. The HELIX investigators plan to measure external environmental exposures to food, water, air pollution, pesticides, noise, and ultraviolet radiation in up to 32,000 mother-child pairs and track the growth, development, and health of the children, including birth outcomes, postnatal growth and body mass index, asthma and lung function, and neurodevelopment.

The intended outcome from this project is an exploration of the relationships between the early-life exposome and -omic markers and health in childhood. An important long-term goal is to estimate health impacts for the European population based on exposure levels and dose-response relationships developed from HELIX.

Health and Environment-Wide Associations Based on Large Population Surveys

The Health and Environment-Wide Associations Based on Large Population Surveys (HEALS) Project (http://www.heals-eu.eu/) is a project funded by the 7th Framework Programme of the European Commission (21). The general objective of HEALS is to refine a methodology that integrates and applies analytical and computational tools for performing environment-wide association studies (EWAS) in support of European Union-wide environment and health assessments. HEALS links activities that focus on aspects of individual exposure assessment to conventional and emerging environmental stressors and the prediction of associated health outcomes. The overall approach will be verified and refined in a series of population studies across Europe, including studies of twin cohorts, tackling different levels of environmental exposure, age windows of exposure, and socioeconomic and genetic variability.

The external exposome will be estimated by integrating environmental, occupational, and dietary data into exposure models. HEALS proposes to characterize the internal exposome at the individual level by integrating -omics-derived data and biomonitoring data. The HEALS approach and tools will be tested by applying them in a number of population studies (including twin studies) across different exposure settings, tackling key health endpoints for both children and the elderly. The overall size of the population involved in these studies to date is approximately 335,000 individuals covering different age, sex, and socioeconomic status groups. A high-level goal of HEALS is to develop scientific guidance on exposome-based risk assessment.

EXPOsOMICS

EXPOsOMICS (http://www.exposomicsproject.eu/) is a program created by the European Union that aims to develop a new approach to assessment of environmental exposures (22). EXPOsOMICS investigators will focus primarily on air pollution and water contaminants by developing a personal exposure monitoring system (including sensors, smartphones,

georeferencing, and satellites) to collect data on individuals' external exposome, as well as analyzing biological samples (internal markers of external exposures) using -omic technologies. The program will search for relationships between external exposures measured using the personal exposure monitoring system—which has not previously been used in large-scale studies—and profiles of chemical exposures, measured via -omics, in the same individuals. Using -omic techniques, the collected exposure data can be linked to biochemical and molecular changes, and the results will help to improve understanding on how xenobiotics influence the risk of developing chronic diseases.

TOOLS USED IN THE STUDY OF THE EXPOSOME

Different sets of tools (Table 1) are needed to measure exposures in the 3 exposomic domains outlined by Wild (3). Specific exposures found in an individual's environment may be evaluated in a variety of ways, including biomarker-based metrics (e.g., urinary metabolites of xenobiotics) and the use of sensors (either personal or remote monitoring) to detect contaminants in the individual environment and to evaluate prior occupational/environmental exposures. Personal monitoring might use sensors to determine physical activity and other exposures. Survey instruments or biomarkers can measure stress. Databases, geographic information systems (GIS), and surveys will be helpful in elucidating general external exposures, such as educational level or urbanicity (living in a more urban or rural environment), in the third domain identified by Wild (3). Biomarker measurements will play a large role in developing the exposome, especially given the associated advantages of -omics: a small sample size, high-throughput techniques, and the relatively low cost of obtaining a wealth of usable information.

If the concept of the exposome is to be realized, epidemiologists will need to incorporate exposure and health information from traditional sources as well as consider information from nontraditional sources. In 2012, the US National Research Council released a report identifying 21st century techniques that are likely to provide more information on a wide variety of exposures and the biological effects of those exposures (23). The exposome is a paradigm shift from a single-exposure-to-disease concept to a recognition that health is impacted by multiple exposures. Different

sources of exposure and health information that may be useful to capture are discussed below to provide epidemiologists with greater insight into the tools that may be available to them.

-Omic technologies

-Omic biomarkers are a class of biomarkers of current scientific interest (Table 2). The discovery and use of these biomarkers have increased rapidly, due to the advent of high-throughput technologies and innovations that include improved sample preparation, robotic sample-delivery systems, automated data processing, and use of multivariate statistical methods, with associated reductions in cost. Investigators have begun to use these biomarkers in larger-scale population studies of the exposome.

One of the greatest challenges with these studies is applying -omic technologies to generate meaningful results. Epidemiologists must strive to understand the principles of -omics and determine when it is appropriate to include biomarkers identified using these technologies. In addition, no single -omic approach will suffice to characterize the exposome, and integration of -omic outcomes and other sources of exposure information will be needed to deepen our understanding of the causes of disease. For example, metabolomic studies have shown links between gut flora, diet, and cardiovascular disease (24, 25). -Omic technologies have shown utility in determining toxicity and mode of action in risk assessments and in assessing the health impact of an exposure using analysis-of-variance approaches together with pairwise comparisons between dose groups and their corresponding controls (26). Thus, what might be considered an advance in exposomics can equally be considered a natural combination and extension of traditional investigative tools.

Molecular epidemiology

Many of the techniques for assessing exposure and the exposome involve molecular epidemiologic studies. Molecular epidemiology is the use of biological markers (exposure, effect, susceptibility) in epidemiologic research (27, 28). In assessing the exposome, biomarkers of exposure may be the most useful. The development of adductomics, which measures the full complement of protein adducts, might be useful in improving exposure assessment in epidemiologic studies because of the ability to reflect extended exposure (28, 29). As some -omic applications mature and systems biology becomes more incorporated into molecular epidemiology, the understanding of the exposome will be increased, and it should be possible to more broadly explore exposure-disease relationships, to study effect modifiers, and to obtain insight into temporal and multilevel factors in health and disease.

Sensor technologies

Remote sensing is a key innovation in exposure science and is defined as the measurement of some property of a phenomenon, object, or material that is not in direct physical contact with the population being studied (23). The use of sensors for not only remote monitoring but also personal monitoring

Table 1. Potential Tools With Which to Measure the Exposome

Tool	Domain[a]
Biomonitoring and biomarker data from -omics or other techniques	Internal
Sensors for environmental or personal monitoring	Specific external
Geographic information systems	General external
Conventional methods such as survey instruments or job-exposure matrices	Specific external
Reality mining from social networks or other sources	General external

[a] As described by Wild (3).

Table 2. Types of -Omic Technologies That May Be Useful for Study of the Exposome

Field	Definition	Description
Epigenetics	The study of the totality of all heritable changes in gene expression and chromatin organization that are independent of the DNA sequence itself and that can be inherited in a stable manner over the course of cell division (65)	The epigenome is comprised of a number of posttranslational modifications such as DNA methylation, histone modification, and specifically positioned nucleosomes. The epigenome can be dynamic, influenced by environmental factors and extracellular stimuli, and can change in response to these factors, and it is instrumental in developing the phenome (54, 66). It has been suggested that the epigenome may function as an interface between environmental factors and the genome, and that its deregulation by environmental stressors is likely to disrupt different cellular processes and contribute to disease risk (65).
Transcriptomics	The study of the genetic sequence of messenger RNA (mRNA). By analyzing the transcriptome, researchers can determine when and where each gene is actively expressed at any given moment. Gene expression microarrays measure packaged mRNA (mRNA with the introns spliced out) as a summary of gene activity (67).	The transcriptome is the complete set of RNA transcripts produced by the genome at any one time. Studies of transcriptome profiles may provide insights into how they are affected by development, disease, or environmental factors (54). Some studies have shown that the transcriptome is highly responsive to environmental exposures. For example, specific environmental exposures have been shown to alter the expression of as much as 30% of the transcriptome in specific blood cells (54), which may also affect the phenome.
Proteomics	The study of the full set of proteins encoded by a genome is known as proteomics and involves the identification, characterization, and quantitation of expressed proteins in biological samples (68).	Protein profiles may be useful in understanding the underlying causes of disease and may or may not be useful in identifying exposures. Protein profiles can be useful for describing the phenome and may be helpful in delineating health outcome.
Metabolomics	The study of low-molecular-weight metabolites present in a biological sample (69)	The metabolome is the sum of all low-molecular-weight metabolites present in a biological sample. Metabolic profiles obtained using metabolomic techniques constitute the measurable part of the metabolic phenotypes (or metabotypes). These metabotypes show large interindividual differences and are characteristic of an individual at a particular time of his/her life; moreover, differences can be seen at the population level when comparing populations from different regions (54). Metabotypes are both genetically and environmentally determined, and it is possible to establish metabolic profiles from samples collected under different conditions (e.g., exposures, treatments, or physiological states) and evaluate the effect of other environmental exposures on these profiles (54).
Adductomics	The study of macromolecular adducts in the context of an entire genome (29)	DNA adducts are compounds that bind to DNA, causing damage and mutations. These mutations can result in cancer and birth defects in multicellular organisms. The science of adductomics seeks to identify all DNA adducts and the target sequence of each adduct. It is possible to measure adducts of electrophiles that result from reactions with DNA, with glutathione, and with blood proteins such as hemoglobin and human serum albumin (29).
Genomics	The study of genes or their function (68)	While the exposome is complementary to the genome, some identification of genes that may impact exposure or the effect of exposure may be useful. For example, genetic variants that code for metabolizing enzymes may increase or decrease the level of internal dose.

is growing. Sensors are now used to measure clinical parameters such as blood pressure and glucose levels, and new sensors are being developed to measure biomarkers, such as portable adhesive sweat analyzers (30). The devices—along with remote-sensing-based spatial referencing technologies and modeling—enable the continuous, real-time, real-world assessment of exposures that vary according to an individual's location, activity, and lifestyle. The use of smartphones and computer tablet technology is only likely to grow and provide enhanced opportunities to collect exposure information (23, 31). Apple (Apple Inc., Cupertino, California) has created ResearchKit (https://developer.apple.com/researchkit/), a framework for Apple platforms that can be used in medical research on diseases such as Parkinson disease, asthma, diabetes, heart disease, and breast cancer.

A mobile application was created for the HELIX Project that uses a global positioning system and a built-in accelerometer to track a person's location and measure physical activity every 10 seconds (32). Participants in the HELIX Project will track their exposures using their smartphones for a week, along with air pollution and ultraviolet radiation monitors, and the data will be used to calculate the amount of air inhaled and the individual's exposure to air pollutants. A possible challenge is that persons with access to information from such applications may alter their behavior and thereby inject biases into the study data.

Geographic information systems

GIS are informative for tracking the acquisition, editing, analysis, storage, and visualization of geographic data (33). GIS allow for mapping of a variety of data (e.g., environmental, topographical, or health-related) to understand trends and patterns. Information from GIS would fall into and provide clarity on exposures in the general external domain of the exposome as described by Wild (3). GIS have been reported to enhance exposure assessment in epidemiologic studies, because such systems can provide information on broad

environmental contaminant levels that can be mapped, which can be used to define an exposed population (34). For example, the Research Center on Health Disparities, Equity, and the Exposome at the University of Tennessee Health Science Center (http://rchdee.uthsc.edu/) maintains the Public Health Exposome Data Information System, a longitudinal, 30-year data set that integrates over 20,000 environmental and health data records (9). The database contains data indicators of the atmosphere, climate, water, land cover, and land use that have been geocoded. In addition, a considerable amount of other data are geocoded and included in this database. These data, derived from ground stations and field samples of emissions, heavy metals, toxic dump and storage sites, and brownfields, are collected by the Environmental Protection Agency in conjunction with other federal, state, and local agencies.

Conventional measurements

Survey instruments (e.g., questionnaires) gather information for which there is no way to quantitatively measure exposures and have long been used in epidemiologic studies. These tools may be important in documenting retrospective exposures, which currently is not feasible with other exposomic tools. Modeling the data to improve exposure matrices will be an important aspect of using exposomic data.

Exposome informatics

With the advances in molecular medicine and development of -omic technologies, the field of biomedical informatics has evolved as a discipline. Genome-wide association studies (GWAS) and studies combining genomic and phenotypic data have required the development of new biostatistical methods for quality control, imputation, and analysis issues such as multiple hypothesis testing. Developments in exposome science have revealed the need for evaluating the interrelationships among phenotype, genotype, and exposure data. As shown in the different approaches discussed here, these data will be heterogeneous, wide-ranging, and massive; will frequently involve time series; and will require high-velocity processing (35). Some efforts have been made for the analysis of EWAS, but new research is required to develop approaches for data management, analysis, and visualization.

Reality mining

The term "reality mining" refers to the analysis of behavioral and self-reported data extracted from social network programs and other software applications developed for portable electronic devices (36). Because these programs continually log and time-stamp information about a subject's activity, location, and proximity to other users, it is currently possible to identify patterns in the data and translate them into maps of social relationships. By definition of the exposome, all exposures are to be taken into account. Logistically, that is impossible, but by using social media networks or citizen scientists' efforts, such as HabitatMap (HabitatMap Inc., Brooklyn, New York; http://habitatmap.org/), additional exposure information could be obtained. This relational information

could have much broader implications, including the improvement of existing computational models of exposure, the disease status of an individual, and disease spread (37). A recent report demonstrated how the mobile application Yelp (Yelp Inc., San Francisco, California) was used to identify patterns of foodborne illnesses in New York City restaurants (38).

Environment-wide association studies

In an EWAS employing techniques partially adapted from GWAS, Patel et al. (39, 40) used chemical, clinical, and questionnaire data from the 1999–2006 cycles of the National Health and Nutrition Examination Survey (NHANES) to evaluate associations between multiple environmental factors and type 2 diabetes mellitus and serum lipid levels, respectively. The study followed 2 methodological steps analogous to those in a GWAS. First, the authors considered a panel of environmental assays and identified those significantly associated or correlated with diabetes or serum lipids while controlling for multiple confounders. Second, they validated the associations by testing significant findings in other cohorts in the NHANES I Epidemiologic Follow-up Study. The authors identified environmental factors, including select chemicals, that may play a role in serum lipid levels and the development of diabetes, corroborating earlier findings. Additionally, Patel et al. (41) screened for gene-environment interactions by integrating results from GWAS and EWAS. Properly designed, EWAS can lead to the discovery of biomarkers for exposure and disease and establish a molecular basis for the cause of environmental diseases (42). Patel and Manrai (43) described the "exposome globe" (e.g., see Figure 4), which is a visual depiction of the network of replicated correlations between individual exposures in the exposome. The exposome globe allows visualization of clusters of exposure.

Although -omic technologies are at the forefront in studies of the exposome and associated biomarkers, -omic-based measurements do not always reflect exogenous exposures and may instead be products of normal cellular function (3). Proteomics, transcriptomics, and metabolomics are a few of the -omic technologies that have shown great promise for exposomic studies. As our understanding of basic biological pathways grows, perturbations in the pathways as measured by -omics result in improved interpretation with respect to how health is affected. Other types of exposure information of value to EWAS include databases containing exposure data or population demographic data that can affect exposures. Use of remote sensors to gather environmental data may help ascertain exposures to the population as a whole. Personal monitoring using sensors to determine exposures, physiological factors, and geographic location is beginning to be used in environmental health studies to assess the "total" exposure of study participants.

STUDY DESIGN IN EXPOSOMIC STUDIES

Ideally, the best approach in exposomics would be to use a longitudinal (prospective) cohort study design, as it allows follow-up of individuals and repeated sampling, as well as monitoring during windows of increased sensitivity. However, neither the biomarker data nor the environmental data
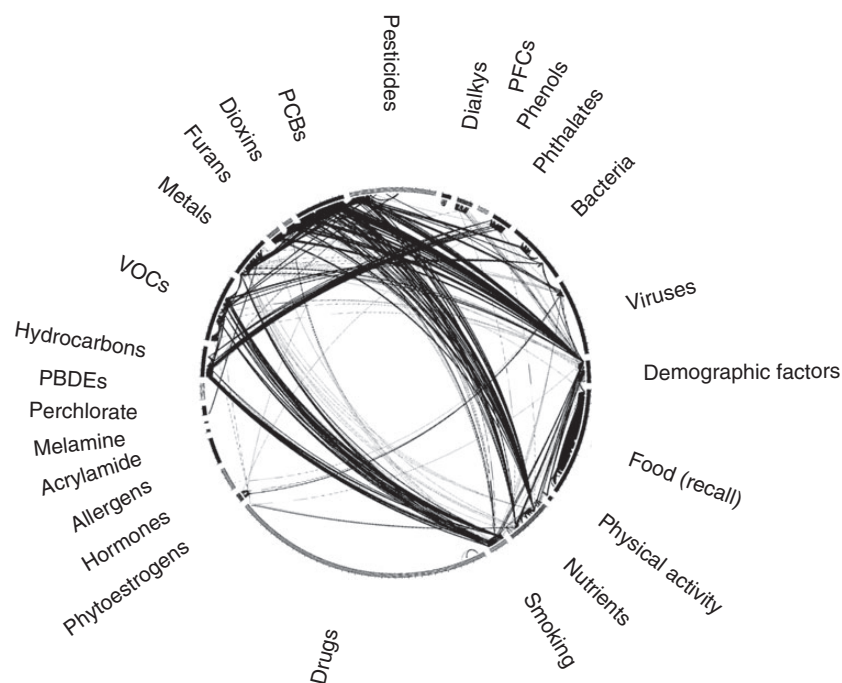
Figure 4. Example of an overall exposome correlation globe. A total of 575 exposures are grouped by a priori-defined environmental health categories. Line thickness is proportional to the size of the absolute value of the correlation coefficient. Adapted from Patel and Manrai (43) with permission from the Pacific Symposium on Biocomputing. © 2015. PBDEs, polybrominated diphenyl ethers; PCBs, polychlorinated biphenyls; PFCs, polyfluorinated compounds; VOCs, volatile organic compounds.

could be recorded continuously, and it would be necessary to carry out a series of cross-sectional investigations of the study population.

Candidate exposures vs. agnostic approaches

Two strategies have been identified for characterizing the exposome (2). One is a "bottom-up" strategy in which all of the exposures in a person's exposome are measured at set time points. Although this approach would have the advantage of identifying important exposures in the air, water, or diet, it is not currently feasible and would miss essential components of the internal chemical environment due to such factors as sex, obesity, inflammation, and stress (44). At the other extreme is the strategy of measuring a combination of -omic endpoints and legacy biomarkers in repeated blood specimens. This strategy has been referred to as "top-down exposomics" (2). This data-driven (or "agnostic") approach lacks specific hypotheses (2). According to Rappaport (2), the exposome would consist of a profile of exogenous and endogenous exposures but would not pinpoint their sources. Since it is currently not feasible to measure all of the chemicals in the blood, it has been proposed (44, 45) that researchers focus initially on the most prominent classes of toxicants known to cause disease—namely, reactive electrophiles, endocrine (hormone) disruptors, modulators of immune responses, agents that bind to cellular receptors, and metals. Exposures to these agents can be monitored in the blood either by direct measurement or by looking for their effects on

physiological processes (such as receptor-based signaling). These measurements could help to generate signatures or profiles for these exposures in the blood. These profiles could be used to help identify key exposures associated with a disease by comparing the profiles of disease cases and controls, preferably from longitudinal studies. Additionally, once important profiles of biomarkers have been identified, the sources of exposure could be determined and methods to reduce the exposures could be identified. Therefore, discovery-driven research and hypothesis-driven research should be considered complementary and synergistic.

Challenges in study design

While many epidemiologic studies measure multiple variables, the multiplicity of variables in exposomic studies can be daunting (46, 47) and may require approaches different from those of traditional epidemiology. From the epidemiologic point of view, the following challenges have been identified in exposomic studies.

Reverse causality. For an exposure to be a cause of disease, the exposure must precede the outcome. Reverse causation is a situation in which the outcome precedes and causes the exposure instead of the other way around. This could occur, for example, if a person moved or changed his/her address as a result of a condition in the domicile that was making him or her sick. Reverse causality is of particular concern in retrospective and cross-sectional studies. In prospective cohort studies, since exposure is determined in advance

of disease onset, the probability of reverse causation is greatly diminished (1). EWAS designs that can reduce the possibility of reverse causation have been developed (48).

Testing multiple variables for associations. Statistical inference problems with the use of -omic methodologies involve the simultaneous testing of thousands of null hypotheses. The multitude of comparisons made in these studies will result in both false-positive findings (type 1 errors) and, if the correction for multiple comparisons is overly conservative or statistical power is inadequate, false-negative findings (type 2 errors) (49).

Univariate models consider each predictive variable separately, and the association of each variable with the outcome of interest is tested using the same statistical model. The familywise error rate and the false discovery rate have been used to characterize the number of associations that could falsely be declared statistically significant (50, 51). The familywise error rate is the probability of making one or more false discoveries, or type 1 errors, among all the hypotheses. Bonferroni correction and other correction methods are commonly used; however, these often produce exceedingly conservative thresholds (50).

With an exploratory approach ("top-down exposomics" as described above), it may be preferable to use a less conservative correction strategy. The false discovery rate is the expected proportion of errors among all associations declared statistically significant. To control for the false discovery rate, one must define the proportion of positive findings that are allowed to be false (usually 5%). Methods have been developed for both approaches and are described in detail elsewhere (50).

Correlation among variables. While univariate methods are useful for uncovering simple relationships between predictors and responses, they are also likely to overlook relationships involving combinations of factors. Different multivariate methods have been developed for this purpose. Multivariate analyses aim to summarize the information contained in large data sets into a few synthetic variables (the principal components) that capture the latent structure of the data. Because these methods effectively reduce the number of dimensions necessary to represent the data, they are often referred to as dimension-reduction methods. These methods have been described in detail elsewhere (50). Due to the density of the data in exposomic studies, identification of independent associations will be challenging. Patel and Ioannidis (46, 47) have proposed some agnostic approaches to aid in the analysis of the large number of variables found in exposomic studies. Smith et al. (52) proposed the use of genetic traits and Mendelian randomization techniques to study highly confounded risk factors and disease causation. Mendelian randomization may be useful in exposomic studies to investigate the effects of modifiable risk factors for diseases that are too heavily confounded to be studied by conventional approaches.

Variability over time and between subjects. Variability over time and between subjects is associated with a multitude of intrinsic and extrinsic factors, some known and some unknown. Unlike the genome, -omic endpoints are dynamic and are likely to show variability in different cells and tissues and throughout the lifetime of an individual (53). Panel studies in small population samples have been proposed to measure the effect of short-term variability on exposure and -omic biomarkers, on individual behaviors (physical activity, mobility, time activity), and on personal and indoor exposures (20).

Variability of exposure data. For exposures with a short biological half-life and little constancy in the underlying exposure behavior, temporal variability may be particularly high. For such exposures, intraindividual variability as compared with interindividual variability is known to be high, and only repeat measurements taken over time provide improved exposure estimates (20). Studies designed to measure daily repeat biomarkers of nonpersistent chemicals (phthalates, phenols, organophosphate pesticides) in urine have been proposed to characterize intra- and interindividual variability in these urinary biomarkers and, where possible, correct for the uncertainties in a larger cohort.

Analytical measurement error. With the development of stable and cost-effective high-throughput techniques, large amounts of experimental data are being generated. However, data obtained from these platforms are highly sensitive to experimental conditions and can therefore include considerable noise in the form of measurement error. Statistical methods should be able to estimate technical nuisance variation (50).

Measurements. Many challenges in measuring the exposome exist. The sheer diversity and variability of exposures that individuals experience throughout their lives is impossible to quantify. Exposures constantly change in a person's environment over time (54). Environmental contaminants appear and disappear, and individuals may make changes in their lifestyle choices that can increase or decrease exposures. Additionally, the relevance of past exposures remains unknown, and the inability to measure past exposures hampers research on the exposome (54). People have been identified as being more susceptible to environmental exposures in certain life stages (i.e., life cycle "windows of susceptibility"), so relevancy across time may also be an issue.

Biomarkers, one of the primary ways to measure the exposome, pose numerous challenges. Their relevance to exposure or their predictive value in nontarget tissues may not have been established. Full validation of the biomarker both in the laboratory and in the population may not always have been performed, making the interpretation or relevance of the measurement difficult. -Omic technologies have multiple advantages that make them well-suited for exposomic studies, but they also have limitations in interpretation and validation for exposures and effects of exposures. The management of large data sets is still a challenge for these technologies, and these technologies also require major investments in equipment and expertise.

Other techniques that may be important for measuring the exposome also have limitations, including uncertainty about the relevance of the measurement, variability over time, and the inability to measure past exposures. While environmental monitoring (personal or remote), geospatial information, and reality mining may tell us what exposures a person may have had, the actual dose that an individual internalized and the relevance of that exposure cannot always be determined. The inherent problem of connecting these external exposures to internal biomarker measurements has always been a challenge, and that challenge is further exacerbated with exposomic research. Newer approaches such as those used by Patel

et al. (55) may be helpful in this regard. Better models that integrate exposure data from multiple sources will be more useful for determining health impact. Additionally, surveys or questionnaires can be improved to assess current or past exposures in greater depth and with greater accuracy, facilitating better research.

Multilevel analysis

Since the exposome can involve various types of data, such as biological, economic, behavioral, and social data, there will be challenges in modeling data from these disparate areas and from the individual, group, and ecological levels in regards to their roles as determinants of disease. Difficulties in inference from ecological data can impede epidemiologic research concerning the effects of exposures on individual-level health behaviors and disease risks. Multilevel analysis seeks to explain relationships involving both individual-level and aggregate-level variables (56). Population and group-level factors may modify the relationship between individual-level risk factors and risk of disease (57). There are various approaches available to epidemiologists for performing multilevel modeling that utilize random-effects models and generalized estimating equations. Multilevel models can be implemented using SAS Proc MIXED and SAS Proc GLIMMIX (58, 59).

FUTURE PROSPECTS

Wild (1) originally developed the concept of the exposome to draw attention to the need for comprehensive exposure assessment in epidemiologic studies. Lichtenstein et al. (60) estimated that the attributable risk of chronic disease from the human genome was only somewhere between 10% and 30% and that the environment was responsible for the other 70%–90%. Measurement of the exposome will play a critical role in future understanding of chronic disease formation and progression. The exposome provides improved exposure assessment via the integration of different types of exposure information, but can the utility of this concept be expanded beyond exposure assessment?

Improved exposure assessment can feed into the systems-biology approach to help evaluate how various exposures disrupt normal biological processes. This type of approach could provide information used in delineating the mode of action of the body's response or toxicity and more accurately inform risk assessments and risk management. The study of early biological effects using a systems-biology approach and computational toxicology efforts offer great promise for improving risk assessment (26). The biomarker techniques such as the -omic approaches have shown promise and can provide information on mode of action and dose-response relationships (61). As these techniques evolve, estimation of internal dose and response markers will be a critical test of these new technologies for application in risk assessment strategies (26). However, it has been suggested that many research "findings" are false or overinflated (62, 63). Improving the accuracy of research findings will require better-powered studies, diminished bias, and improvement in our understanding of *R* values. Juarez and Hood (9) have explored the public health exposome as a

way to integrate environmental contextual data with measures of health outcomes. The exposome is a new way to conduct health disparities research, to better understand the social and environmental factors involved and their effects on health. Other authors have suggested that the exposome could play a major role in clinical care or serve as a basis for transformative research (8, 64). The benefits of the exposome for both individual and population health research are many. The exposome can aid in identifying modes of action of stressors, identifying unknown exposures, and improving our understanding of disease. This can lead to better risk assessments, better translation of science into practice, and ultimately disease prevention.

The concept of the exposome is a paradigm shift. While the concept is daunting—particularly the idea of measuring all exposures an individual has experienced in a lifetime and predicting their health impact—studies are already being conducted using exposomic principles. Epidemiologic studies can be improved by incorporation of exposomic principles, increasing the accuracy of estimated associations between exposures and numerous health outcomes and conditions.

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(8):1847–1850.

2. Rappaport SM. Implications of the exposome for exposure science. *J Expo Sci Environ Epidemiol.* 2011;21(1):5–9.
3. Wild CP. The exposome: from concept to utility. *Int J Epidemiol.* 2012;41(1):24–32.
4. Needham LL, Calafat AM, Barr DB. Assessing developmental toxicant exposures via biomonitoring. *Basic Clin Pharmacol Toxicol.* 2008;102(2):100–108.
5. Goodman A, Schorge J, Greene MF. The long-term effects of in utero exposures—the DES story. *N Engl J Med.* 2011;364(22): 2083–2084.
6. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med.* 1971;284(15): 878–881.
7. Athersuch TJ. The role of metabolomics in characterizing the human exposome. *Bioanalysis.* 2012;4(18):2207–2212.
8. Li-Pook-Than J, Snyder M. iPOP goes the world: integrated personalized omics profiling and the road toward improved health care. *Chem Biol.* 2013;20(5):660–666.
9. Juarez P, Hood DB. Sequencing the public health genome. *J Health Care Poor Underserved.* 2013;24(1 suppl):114–120.
10. Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. *Toxicol Sci.* 2014;137(1):1–2.
11. Sexton K, Hattis D. Assessing cumulative health risks from exposure to environmental mixtures—three fundamental questions. *Environ Health Perspect.* 2007;115(5):825–832.
12. Lioy PJ. Exposure science: a view of the past and milestones for the future. *Environ Health Perspect.* 2010;118(8): 1081–1090.
13. Cohen Hubal EA, Barr DB, Koch HM, et al. The promise of exposure science. *J Expo Sci Environ Epidemiol.* 2011;21(2): 121–122.
14. Pleil JD. Categorizing biomarkers of the human exposome and developing metrics for assessing environmental sustainability. *J Toxicol Environ Health B Crit Rev.* 2012;15(4):264–280.
15. Zartarian VG, Schultz BD. The EPA's human exposure research program for assessing cumulative risk in communities. *J Expo Sci Environ Epidemiol.* 2010;20(4):351–358.
16. Williams P, Dotson G, Maier A. Cumulative risk assessment (CRA): transforming the way we assess health risks. *Environ Sci Technol.* 2012;46(20):10868–10874.
17. Bonneterre V, Faisandier L, Bicout D, et al. Programmed health surveillance and detection of emerging diseases in occupational health: contribution of the French national occupational disease surveillance and prevention network (RNV3P). *Occup Environ Med.* 2010;67(3):178–186.
18. Faisandier L, Bonneterre V, De Gaudemaris R, et al. Occupational exposome: a network-based approach for characterizing occupational health problems. *J Biomed Inform.* 2011;44(4):545–552.
19. National Institute of Environmental Health Sciences. Exposure biology and the exposome. http://www.niehs.nih.gov/research/supported/dert/programs/exposure/index.cfm. Updated October 5, 2015. Accessed October 30, 2015.
20. Vrijheid M, Slama R, Robinson O, et al. The Human Early-Life Exposome (HELIX): project rationale and design. *Environ Health Perspect.* 2014;122(6):535–544.
21. Adekoya N, Myers JR. Fatal harmful substances or environmental exposures in agriculture, 1992 to 1996. *J Occup Environ Med.* 1996;41(8):699–705.
22. Adgate JL, Kukowski A, Stroebel C, et al. Pesticide storage and use patterns in Minnesota households with children. *J Expo Anal Environ Epidemiol.* 2000;10(2):159–167.
23. National Research Council. *Exposure Science in the 21st Century: a Vision and a Strategy.* Washington, DC: National Academies Press; 2012.
24. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med.* 2013;368(17):1575–1584.
25. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature.* 2011;472(7341):57–63.
26. DeBord DG, Burgoon L, Edwards SW, et al. Systems biology and biomarkers of early effects for occupational exposure limit setting. *J Occup Environ Hyg.* 2015;12(suppl 1):S41–S54.
27. Schulte PA, Perera F, eds. *Molecular Epidemiology: Principles and Practices.* San Diego, CA: Academic Press, Inc.; 1993.
28. Rothman N, Hainaut P, Schulte P, et al. *Molecular Epidemiology: Principles and Practices.* (IARC Scientific Publication no. 163). Lyon, France: International Agency for Research on Cancer; 2011.
29. Rappaport SM, Li H, Grigoryan H, et al. Adductomics: characterizing exposures to reactive electrophiles. *Toxicol Lett.* 2012;213(1):83–90.
30. Ahlbom A, Navier IL, Norell S, et al. Nonoccupational risk indicators for astrocytomas in adults. *Am J Epidemiol.* 1986; 124(2):334–337.
31. Howard JH, DeBord DG, Hoover MD. The future of industrial hygiene. *Synergist.* 2014;25(8):7.
32. Potera C. The HELIX Project: tracking the exposome in real time. *Environ Health Perspect.* 2014;122(6):A169.
33. Longley PA, Goodchild MF, Maguire DJ, et al. *Geographic Information Systems and Science.* 2nd ed. New York, NY: John Wiley & Sons, Inc.; 2005.
34. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect.* 2004;112(9):1007–1015.
35. Martin Sanchez F, Gray K, Bellazzi R, et al. Exposome informatics: considerations for the design of future biomedical research information systems. *J Am Med Inform Assoc.* 2014; 21(3):386–390.
36. Eagle N, Pentland A. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.* 2006;10(4):255–268.
37. Madan A, Cebrian M, Lazer D, et al. Social Sensing for Epidemiological Behavior Change. In: *UbiComp '10: Proceedings of the 2010 ACM Conference on Ubiquitous Computing, September 26–29, 2010, Copenhagen, Denmark.* New York, NY: Association for Computing Machinery; 2010:291–300.
38. Harrison C, Jorder M, Stern H, et al. Using online reviews by restaurant patrons to identify unreported cases of foodborne illness—New York City. *MMWR Morb Mortal Wkly Rep.* 2014; 63(20):441–445.
39. Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One.* 2010;5(5):e10746.
40. Patel CJ, Cullen MR, Ioannidis JP, et al. Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels. *Int J Epidemiol.* 2012;41(3):828–843.
41. Patel CJ, Chen R, Kodama K, et al. Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. *Hum Genet.* 2013; 132(5):495–508.
42. Rappaport SM. Biomarkers intersect with the exposome. *Biomarkers.* 2012;17(6):483–489.
43. Patel CJ, Manrai AK. Development of exposome correlation globes to map out environment-wide associations. *Pac Symp Biocomput.* 2015;20:231–242.
44. Rappaport SM, Smith MT. Environment and disease risks. *Science.* 2010;330(6003):460–461.
45. Rappaport S, Barupal D, Wishart D, et al. The blood exposome and its role in discovering causes and disease. *Environ Health Perspect.* 2014;122(8):769–774.

46. Patel CJ, Ioannidis JP. Studying the elusive environment in large scale. *JAMA*. 2014;311(21):2173–2174.
47. Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J Epidemiol Community Health*. 2014;68(11):1096–1100.
48. Patel CJ, Rehkopf DH, Leppert JT, et al. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. *Int J Epidemiol*. 2013;42(6): 1795–1810.
49. Kelsey JL, Whittemore AS, Evans AS, et al. *Methods in Observational Epidemiology*. New York, NY: Oxford University Press; 1996.
50. Chadeau-Hyam M, Campanella G, Jombart T, et al. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ Mol Mutagen*. 2013;54(7):542–557.
51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995;57(1):289–300.
52. Smith GD, Lawlor DA, Harbord R, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*. 2007;4(12):e352.
53. Mias GI, Snyder M. Personal genomes, quantitative dynamic omics and personalized medicine. *Quant Biol*. 2013;1(1):71–90.
54. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. *Environ Mol Mutagen*. 2013;54(7):480–499.
55. Patel CJ, Ioannidis JP, Cullen MR, et al. Systematic assessment of the correlations of household income with infectious, biochemical, physiological, and environmental factors in the United States, 1999–2006. *Am J Epidemiol*. 2015;181(3): 171–179.
56. Von Korff M, Koepsell T, Curry S, et al. Multi-level analysis in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol*. 1992;135(10):1077–1082.
57. Diez Rouz AV, Aiello AE. Multilevel analysis of infectious diseases. *J Infect Dis*. 2005;191(suppl 1):S25–S33.
58. Bingenheimer JB, Raudenbush SW. Statistical and substantive inferences in public health: issues in the application of multilevel models. *Annu Rev Public Health*. 2004;25:53–77.
59. Witte JS, Greenland S, Kim L-L, et al. Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology*. 2000;11(6): 684–688.
60. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78–85.
61. Thomas RS, Clewell HJ III, Allen BC, et al. Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol Sci*. 2011;120(1): 194–205.
62. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
63. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640–648.
64. Buck Louis GM, Sundaram R. Exposome: time for transformative research. *Stat Med*. 2012;31(22): 2569–2575.
65. Herceg Z, Vaissière T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics*. 2011;6(7):804–819.
66. Aiken RD. Quality-of-life issues in patients with malignant gliomas. *Semin Oncol*. 1994;21(2):273–275.
67. Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet Gynecol*. 2011;13(3):189–195.
68. Ferlini A, Scotton C, Novelli G. Biomarkers in rare diseases. *Public Health Genomics*. 2013;16(6):313–321.
69. Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013;41(database issue):D801–D807.

(Appendix follows)

Appendix Table 1. Exposomic terms used in this article and their definitions

Term	Definition
Environment- or exposome-wide assessment studies (EWAS)	Studies that collect multiple kinds of exposure data from multiple sources which are then related to health effects
Exposome	The totality of exposures experienced by an individual during his or her life and the health impact of those exposures (1)
Exposome informatics	Data management framework for dealing with large multiscale data sources in exposomic studies (32)
Genome	The genetic material of an organism, encoded in DNA and including both the genes and the noncoding sequences of the DNA (68)
Genome-wide association studies (GWAS)	Studies that evaluate markers across the genome to elucidate associations with diseases
Geographic information systems (GIS)	Systems that manage geographic data (33)
Metabolome	Sum of all low-molecular-weight metabolites present in a biological sample (69)
-Omic technologies	The collective characterization of components and measurement of molecules from a biological field of study, which involves a large-scale data acquisition system that can be used to measure biological states or responses; examples include the genome (DNA), transcriptome (RNA), and proteome (proteins)
Phenome	All of the phenotypes of a cell, tissue, or organism
Proteome	The full set of proteins encoded by a genome (68)
Reality mining	Analysis of behavioral and self-reported data extracted from social network software and other portable-device applications (36)
Sensors	Devices that measure an event or change; there are many diverse types available
Total exposure	All exposures experienced by an individual
Transcriptome	Complete set of RNA transcripts produced by the genome at any one time (67)