Practice of Epidemiology

Toward a Road Map for Global -Oomics: A Primer on -Omic Technologies

Steven S. Coughlin*

* Correspondence to Dr. Steven S. Coughlin, Department of Epidemiology, Rollins School of Public Health, Emory University, 62 North Main Street, No. 510, Memphis, TN 38103 (e-mail: stevecatlanta@aol.com).

Initially submitted May 28, 2014; accepted for publication September 3, 2014.

As highlighted in a recent editorial in the *Journal (Am J Epidemiol. 2014;180(2):127–128)*, the research area of “-omics” includes genomics, proteomics, metabolomics, and nascent fields of scientific inquiry such as epigenomics and exposomics. These fields can be collectively referred to as “global -omics.” Increasing efforts have been made over the past 2 decades to identify and modify environmental risk factors among persons who are susceptible to disease because of their genotype and to integrate genetic information and other biological variables with information about individual-level risk factors and group-level or societal factors related to the broader residential, behavioral, or cultural context. In genome-wide association studies, only a small proportion of heritability is explained by genetic variants identified to date, which has prompted researchers in bioinformatics and biostatistics to take into account nonlinear relationships due to gene-environment or gene-gene interactions. The exposome, which is dynamic and variable, consists of all of the internal and external exposures an individual incurs over a lifetime. Both the epigenome and exposome change with age. The prenatal and perinatal periods are thought to be important times for epigenetic marking. Once the human epigenome has been fully mapped, identification of the effects of all deleterious environmental exposures according to duration of exposure and time period will be a complex undertaking, requiring collaborative epidemiologic studies.

bioinformatics; environmental exposures; epidemiologic methods; epigenomics; genetic epidemiology; genome; proteome; transcriptome

Abbreviations: GWAS, genome-wide association studies; SNP, single nucleotide polymorphism.

*Oomics* is an English-language neologism that refers to a field of study in biology ending in “-omics.” Examples include genomics, proteomics, and metabolomics (1). A related suffix (-ome) is often used to refer to the objects of study in these fields (e.g., the genome, proteome, or metabolome). The research area of -omics focuses on the collective characterization and quantification of large numbers of biological molecules that translate into the structure, function, and dynamics of an organism (2). In addition to relatively established fields of biological research, such as genomics, proteomics, transcriptomics, and metabolomics, there has been increasing interest in clarifying the totality of internal and external exposures. These emerging fields of scientific inquiry (e.g., microbiomics and exposomics) have several features in common, including the use of biomarkers to determine exposure, effects of exposure, and susceptibility factors; the use of technologies that produce large amounts of data; and the use of data-mining techniques to identify statistically significant associations (2). The importance of -omic technologies to the practice of epidemiology has recently been highlighted in the *Journal (3, 4)*.

In this article, key developments in the rapid growth of research on the genomic and molecular mechanisms of disease are considered in relation to transcriptomics, proteomics, metabolomics, epigenomics, and the emerging fields of exposomics and eco-exposomics. These fields can be collectively referred to as “global -omics.” An account of how these seemingly disparate areas of scientific inquiry intersect is provided, and important developments leading toward a road map for global -omics are discussed.
BACKGROUND

The Human Genome Project was launched in 1990, and the first individual genome sequences were reported in 2001 (5, 6). By the late 1990s, several articles had been published that explored the nascent field of public health genetics (7–10) and the intersection of genetics, public health, and preventive medicine (5, 7–9). The topics considered included genetic aspects of several illnesses, including coronary heart disease, hemochromatosis, breast cancer, colorectal cancer, and human immunodeficiency virus infection (11). Ethical, legal, and social issues that arise from genetic testing and screening of newborns and adults had also been described by that time (11–13).

Beginning in the 1990s, interest broadened beyond single-gene disorders evident in childhood (such as sickle-cell disease and phenylketonuria) to the genetic basis of common, adult-onset disorders that cause substantial morbidity and mortality. Examples of such genetic disorders include mutations in the human mutS protein homolog 2 (hMSH2) and human mutL protein homolog 1 (hMLH1) genes as causes of susceptibility to colorectal cancer, associations between factor V Leiden and thromboembolic disorders, an association between the ε4 allele of the apolipoprotein E gene (APOE) and late-onset Alzheimer’s disease, and identification of the high-iron Fe gene (HFE) for hemochromatosis (10, 14). Over the past 2 decades, there has been increasing interest in genetic screening for adult-onset disorders in high-risk families, as part of research protocols, or as a routine part of medical practice or genomic medicine. An emerging paradigm of disease prevention is the identification and modification of environmental risk factors among persons who are susceptible to disease because of their genotype (8, 9). For example, genetic testing for childhood asthma may improve the predictive value of environmental factors such as allergens and secondhand exposure to cigarette smoke. The identification of gene-environment interactions in the etiology of osteoporosis could result in preventive and therapeutic interventions for middle-aged persons at risk for later complications of the disease (15). For example, researchers have identified interactions between candidate genes for osteoporosis and nongenetic factors such as low calcium intake, vitamin D deficiency, and physical inactivity. Tests for other genetic conditions may allow for the identification of subgroups of patients who are more or less likely to benefit from preventive strategies such as the use of cholesterol-lowering drugs and replacement estrogens. According to this paradigm, genetic screening might allow for the identification of persons at high risk for an illness and targeted medical interventions.

Increasing efforts have been made in epidemiologic studies to integrate genetic information and other biological variables with information about individual-level risk factors and group-level or societal factors related to the broader residential, behavioral, or cultural context. Collaborative research teams and individual scientists have considered ways to integrate genetic information with social and contextual information in epidemiologic studies. Several authors have called for new research paradigms in epidemiology that more adequately take into account different levels of analysis at the molecular, individual, and societal or group levels (16–18). They have emphasized the importance of assessing both population risk factors for disease that exist at the group level (e.g., neighborhood effects) and more proximal risk factors that exist at the individual level (e.g., biological variables or genetic traits assessed using the methods of molecular genetics). This approach allows biological phenomena to be viewed within their social contexts and for individual-level explanations of disease causation to be integrated into broader models that incorporate interactions between individuals as well as group-level determinants and effect modifiers (16).

There is a special enthusiasm about the power of genomics to determine disease etiology (19). This has led to a tendency to view associations with genetic factors as causal and associations with factors related to the broader social, cultural, or behavioral context as noncausal (20). Although genetic and biological factors are commonly considered to be foundational in hierarchical models and theories of disease causation, there is no a priori reason to assume that causes found at the genetic or molecular level are any more significant than causes operating at another level, such as social factors (18, 21). Even in genome-wide association studies (GWAS) of coronary heart disease and other common complex diseases, only a small proportion of heritability is explained by the genetic variants identified to date (22). It is possible that familial clustering due to genetic factors has been overestimated and that important environmental or social influences (acting alone or in combination with genetic variants) have been overlooked. There may be a tendency to inadequately take into account the complexity of the human biological, physical, and social environment or the potential for confounding or effect modification by unmeasured genetic or environmental factors (19).

In 2005, cancer epidemiologist Dr. Christopher Wild proposed the exposome, which consists of all of the exposures an individual incurs over a lifetime and how those exposures are related to human health (23, 24). The exposome, which is dynamic and evolves throughout the lifetime of an individual, includes internal and external environmental exposures such as microorganisms, alcohol, and tobacco, as well as social, economic, and psychological exposures such as social capital, education, and mental stress (23, 24). Wild and other authors have further developed the exposome concept in more recent articles. For example, Miller and Jones (25) proposed a revised definition of the exposome that incorporates the body’s response to environmental influences and includes metabolic processes that can alter chemical exposures.

The HERCULES Center (Health and Exposome Research Center: Understanding Lifetime Exposures) at Emory University’s Rollins School of Public Health (Atlanta, Georgia) is supported by the National Institute of Environmental Health Sciences, and its funding represents the first exposome-based center grant awarded in the United States (26). The European Commission awarded 2 large grants to pursue exposome-related research in 2012. The Exposomics Project at Imperial College London (London, United Kingdom) is using smartphones, global positioning systems, and environmental sensors to measure exposures. The Human Early-Life Exposome (HELIx) Project at the Barcelona-based Centre for Research in Environmental Epidemiology (Barcelona, Spain) is focusing

Am J Epidemiol. 2014;180(12):1188–1195
on early-life exposures (27). It builds upon 6 existing birth cohorts across Europe and assesses the exposome during the prenatal period and early childhood through the use of geographic information systems, personal sensors, -omics platforms, and biomarkers. The Environmental Health Risks in European Birth Cohorts (ENRICO) Project, part of the European Commission’s 7th Framework Programme, has inventoried 37 European birth cohort studies that collected data on environmental exposures (28). The Health and Environment-wide Associations based on Large Scale Population Surveys (HEALS) Project, also part of the 7th Framework Programme, is examining complex, dynamic interactions between DNA sequence, gene expression, epigenetic modifications of DNA, and environmental factors that influence disease. In the United States, the National Children’s Study (https://www.nationalchildrensstudy.gov/Pages/default.aspx) will provide data and biological samples on the first 21 years of life in a large cohort followed prospectively.

These ongoing initiatives and collaborative projects are addressing the current lack of information about mechanisms of disease at the molecular, cellular, individual, and societal levels. Important outstanding questions remain about biological and physiological processes that account for replicable associations with genetic markers. Genetic markers such as those examined in GWAS and in molecular epidemiology studies may be far removed from complex physiological processes that are more important risk factors for disease (29). In addition, individual- and societal-level risk factors, commonly examined in epidemiologic studies, may represent only crude markers of disease risk and do not provide detailed information about causal pathways. These scientific challenges require systematic approaches to acquire and utilize information about the timing, intensity, and duration of exposures and associated biological responses and health outcomes (30).

AREAS OF -OMICS RESEARCH

Genomics

The completion of the Human Genome Project in 2001 and the 1000 Genomes Project in 2010 (5, 31–33) was made possible by advancements in high-throughput genomic technology, including sequencing machines that can read over 250 billion bases per week (34). The traditional candidate-gene approach has permitted the identification of approximately 2,850 genes underlying Mendelian diseases and over 1,100 genetic loci that play a role in common polygenic diseases such as coronary heart disease, prostate cancer, and osteoporosis (35, 36).

The rationale and methods for GWAS have been explained elsewhere (22, 35, 36). These studies utilize automated, high-throughput genotyping technologies to analyze biological specimens collected from cases and controls for up to 1 million single nucleotide polymorphisms (SNPs), at a relatively low cost (37). The genome-wide association approach allows interrogation of the entire human genome in large numbers of unrelated individuals (36). In contrast to studies of candidate genes, there may be no a priori hypotheses about genetic associations with disease in GWAS. Nearby SNP alleles tend to be inherited together more often than expected by chance due to linkage disequilibrium (35). Because of the strong associations among the SNPs in most chromosomal regions, only a few SNPs need to be typed to predict the likely variants at the rest of the SNPs in a particular region. As a result, it is not necessary to type all 10 million common SNPs in persons with and without disease in order to identify sites that differ in frequency between the groups. Studies involving the typing of a few hundred thousand or up to 1 million of these SNPs can test the hypothesis that 1 or more common variants explain part of the genetic risk for a disease or trait (37). If a SNP increases the risk of a common disease, then there will be a statistical association in the population between disease and that SNP and several nearby SNPs due to linkage disequilibrium (22). Quality control procedures and the replication of results obtained from GWAS are important considerations, since genotyping errors can cause spurious false-positive or false-negative results that obscure true associations (35). Extensive data cleaning and quality control are required in GWAS to detect problems that can result in spurious findings (22).

Genome-wide association scanning has so far identified variants or chromosomal regions associated with more than 40 complex, non-Mendelian diseases or traits, including type 1 and type 2 diabetes; breast, prostate, and colorectal cancer; coronary heart disease and other cardiovascular diseases; neuropsychiatric conditions; and autoimmune and infectious diseases, although the estimated odds ratios are mostly small (35). In order to have sufficient statistical power to detect genotypic odds ratios on the order of 1.1–1.4, it is necessary to include large numbers of cases and controls. Studies of disease subtypes or cases with a younger age at onset or a more rapidly progressive course of disease are also likely to be informative (35). The study of rare variants requires large-scale resequencing analyses (35).

Most deleterious genetic variants are found in protein-coding regions (exons) or in sequences that control gene expression (promoters). However, gene expression can be altered by SNPs found in noncoding gene segments (introns). Hindorff et al. (38) examined characteristics of reported trait- and disease-associated SNPs and found that 45% of reported trait/disease-associated SNPs were intronic and 43% were intergenic. Moreover, the trait-/disease-associated SNPs were significantly depleted in intergenic regions, which supports the assertion that intergenic regions have the smallest ratio of functional DNA to total DNA, even though they may contain important regulatory sequences (38).

GWAS raise special challenges due to the very large number of tests of disease-SNP associations and the potential for false-positive findings. Biological specimens may be examined to look for patterns of gene expression. The area around a trait-/disease-associated SNP may be sequenced to identify rarer variants with more apparent functional significance (22). Ioannidis and Khoury (39) have outlined opportunities for improving validation practices in genomics research.

Quantitative methods for analyzing data from GWAS allow for the identification of statistical associations which may then be replicated in independent samples or validated through studies of gene expression and proteomics. Statistical tests are performed to identify associations between SNPs passing quality thresholds and the disease or trait of interest (40).
Stringent levels of statistical significance are required because of the vast number of tests performed (22). Biostatistical analyses of data from GWAS have often been “agnostic” in the sense that they ignore prior knowledge about disease pathobiology (41). In addition, analyses have often examined associations with one SNP at a time and ignored their genomic context and gene-gene and gene-environment interactions. More recent approaches in bioinformatics have strived to take into account nonlinear relationships due to gene-gene and gene-environment interactions (41). Genotypic effects may depend upon environmental exposures. Parametric statistical models, which play an important role in contemporary genetic epidemiology, have limited power for modeling high-order, nonlinear interactions (42). Efforts are being made to apply data-mining techniques, machine learning, and advanced computational methods—for example, random forests, neural networks, and multifactor dimensionality reduction—to the immense amounts of data obtained from GWAS (42, 43). Such approaches make fewer assumptions about the functional form of the model.

Transcriptomics, metabolomics, and proteomics

Although the initial successes of GWAS have focused attention on genomics, proteomic, metabolomic, and transcriptomic information is also important. Tzoulaki et al. (3) provided a useful account of the design and analysis of metabolomics studies in epidemiologic research. Because there are approximately 500 different histological cell types in the human body and each one has its own gene expression, proteome, and metabolome, humans have over 500 dynamic cellular metabolomes (3). Metabolomic profiling, which must take into account the temporal relationships between the onset of disease and treatment and progression of disease, can provide insights into the molecular pathways that underlie biological processes and measure response to therapy. Studies of microRNAs have also been described (4). Gene expression data obtained from cell lines and other human tissues (e.g., blood, brain, liver) are often key to understanding physiological or regulatory pathways (42, 44). Studies have shown that quantitative analyses of gene expression data and transcriptome information can help to clarify results obtained from GWAS (43). Using data from expression quantitative trait locus studies, Nicolae et al. (45) showed that SNPs associated with complex traits and diseases are more likely to be expression quantitative trait loci than other SNPs chosen from GWAS (and matched on minor allele frequency). Their results indicate that annotating SNPs with a score reflecting the strength of the evidence that the SNP is an expression quantitative trait locus can enhance the ability to discover true associations in GWAS and clarify the biological mechanism accounting for the associations (45). These findings suggest ways to more accurately characterize SNP signals in GWAS with respect to target genes and biological function. In order to detect differences in gene expression across tissue types, disease, and time, computer algorithms and computational methods have been developed for analyzing data obtained using high-throughput techniques (46, 47). To better understand physiological and pathological processes at the molecular level, pathway analysis and differential network analysis have been applied to detect and quantify signaling cascades and regulatory networks (47).

The metabolome and the proteome have a role in oxidation and reduction reactions that occur at the cellular level. Biological reactions such as the formation of free radicals and other oxidation and reduction reactions allow the body to utilize essential nutrients (e.g., glucose) and adapt to environmental challenges such as those posed by microorganisms. Reduction and oxidation (redox) biological reactions play an important role in homeostasis and occur continuously as part of signaling of transcriptional responses and DNA repair; they may also play a role in aging and in adverse health outcomes such as atherosclerotic cardiovascular disease and certain types of cancer (30).

Epigenomics

The epigenome consists of nonnucleotide modifications of DNA that lead to changes in gene expression. The resulting epigenetic modifications (e.g., DNA methylation) may account for deleterious effects of environmental agents on gene expression (34). DNA methylation has been shown to lead to changes in chromatin structure and gene expression. High-throughput sequencing techniques with analysis of the pattern of DNA methylation and chromatin modifications are used for methylation analysis. A map of epigenomic variations has been created that shows how interindividual differences in DNA sequences can be amplified to result in phenotypic variations (34, 48). The International Human Epigenome Consortium (http://www.epigenome.org/) has the goal of mapping hundreds of reference epigenomes over the next 10 years.

The epigenome changes with age. The prenatal and perinatal periods are thought to be important time periods for epigenetic marking (34). Epigenetic changes that occur during the prenatal and perinatal periods may have important effects on adult health. Transgenerational effects are included in epigenomics. Some environmental exposures have been shown to affect pregnant mothers, the fetus, and the germ cells of the fetus, leading to third-generation effects (49). One challenge in epigenomic marker studies is that changes may occur as a result of oxidative stress or unrelated early-life exposures or transgenerational effects (50).

Gaining a clear understanding of how variations in DNA sequences and variations in epigenetic markings contribute to the development of complex diseases is challenging (34). Once the human epigenome has been fully mapped, identification of the effects of all deleterious environmental exposures according to duration of exposure and time period (e.g., the prenatal period or early childhood) will be a complex undertaking and will likely require collaborative epidemiologic studies (34). Studies of the effects of only a few environmental factors (e.g., nutrition, severe stress, and environmental pollution) will yield a huge amount of data (51).

Microbiomics

In a study of 124 Europeans, Qin et al. (52) found over 3.3 million microbial genes. The microbiome, consisting of $10^{14}$ bacteria that reside in and on the human body, plays
an important role in the development and maintenance of the immune system, and it has also been associated with obesity and atherosclerosis (53–55). In addition, the diversity of microbes in the human gut plays a key role in digestion (52, 56). Although it is clear that the microbiome has a profound effect on nutrition and metabolism, substantial work remains to be accomplished to identify associations with human diseases and health outcomes.

Exposomics

Success in mapping the human genome has led to the complementary concept of the exposome (2, 23, 24). As noted above, the exposome consists of all of the exposures an individual incurs over a lifetime (embryonic, fetal, newborn, early childhood, preadolescence, adolescence, young adulthood, middle age, older adulthood, and old age). The exposome concept is partly indebted to long-standing interest in a life-course approach to epidemiology (57). Longitudinal epidemiologic studies have frequently examined environmental exposures in a life-course fashion, including vulnerable periods such as fetal development and early childhood. The exposome concept takes into account the full spectrum of possible exposures (diet, dietary supplements, food additives, pesticide residues, microbial organisms and infections, physical exposures, environmental pollution, smoking, alcohol consumption, exercise, infections, vaccinations, occupational exposures, consumer products, therapeutic drugs, severe stress, etc.), the timing of exposures, and acute and chronic responses to exposures (30). In contrast to the genome, the exposome is highly variable and dynamic and evolves throughout the lifetime of an individual (23). The exposures that are taken into account by the exposome concept go far beyond those that are quantitatively evaluated in environmental epidemiology today. Heretofore, environmental epidemiology has rarely extended beyond the Environmental Protection Agency’s 129 priority pollutants in external media (e.g., ambient air) (58) or the roughly 300 environmental chemicals measured in human biological fluids (52).

Exposomics (the study of the exposome) relies on the application of both internal and external exposure assessment methods and techniques. External exposure assessments include laboratory analyses, direct reading instruments, and large-scale surveys (2). External exposure assessment is needed because the use of measures of biological samples alone does not capture the full impact of complex exposures such as secondhand tobacco smoke inhalation or particulate matter air pollution (50). In addition, several potentially harmful environmental exposures (e.g., heat, high altitude, electromagnetic radiation) do not have direct correlates as protein adducts or metabolites (50). Thus, the application of the exposome concept to environmental health problems requires linkages with the external environment.

Lioy and Rappaport (59) noted that both “bottom-up” and “top-down” approaches have been used to characterize the exposome of individuals. The bottom-up approach focuses on each category of external exposure (air pollution, radiation, etc.) to quantify contaminant levels and to sum those levels over all categories to construct individual exposomes. However, this approach, which employs external monitoring of exposures, would require tremendous effort to evaluate the vast number of largely unknown analytes, and it would also overlook important endogenous exposures. The top-down approach uses untargeted -omic methods to measure features of exposures in biological specimens and is therefore appealing to scientists who have used biomonitoring to assess exposure levels. This approach is relatively efficient because both exogenous and endogenous exposures may be reflected in measures obtained from blood specimens (59). The idea is to use -omic profiles to generate hypotheses to identify exposures, to develop specific biomarkers for high-throughput screens, and to determine sources of external and internal exposure. Metabolomics studies have applied this approach to identify previously unknown exposures associated with cardiovascular disease (60, 61). Lioy and Rappaport recommended use of both the bottom-up and top-down approaches to characterize individual exposomes and suggested that researchers see these 2 approaches as complementary (59). Many experts in the field recommend both the bottom-up and top-down approaches on the basis of current scientific considerations.

A third approach to characterizing the exposome, which incorporates public health surveillance data and other existing data sets on exposures, behavior, and disease morbidity and mortality, has also been proposed (62, 63). For example, van Tongeren and Cherrie (62) called for the assembly of mobile phone data to track the movements of individuals over time; consumer information about use of electricity and natural gas, which is routinely collected by utility companies; and data about food purchases tracked by supermarkets. The idea is to use aggregated data on exposures and personal behavior, contextual data, and information from environmental sensors to characterize the exposome of individuals for use in epidemiologic studies. A dynamic, multidimensional data information system, which also incorporates public health surveillance data and other existing data sets, has been developed by researchers at Meharry Medical College (Nashville, Tennessee) (63).

Gaining a thorough understanding of the exposome will require both improved methods for collecting exposure information via questionnaires and continued progress in identifying biological markers suitable for large-scale epidemiologic studies. Self-reported information that is collected via questionnaires is commonly validated using information or samples collected from small subsets of the overall study population. Validation studies may employ more detailed questionnaire data or laboratory analyses of biological markers found in blood, urine, exhaled air, or hair (64).

SUMMARY AND CONCLUSIONS

In summary, only a small proportion of heritability is explained by genetic variants identified to date. Results from twin studies, for example, suggest that the genetic contributions to cancer and degenerative diseases are about 10% (65). The exposome, which consists of all of the internal and external exposures an individual incurs over a lifetime, is dynamic, variable, and changes with age. The prenatal and perinatal periods are likely to be important times for epigenetic marking. Once the human epigenome has been fully
mapped, identification of the effects of all deleterious environmental exposures according to duration of exposure and time period will be a complex undertaking, requiring collaborative epidemiologic studies. Although there have been calls for a Human Exposome Project, no such project exists at this time.

There is great interest in several scientific disciplines in measuring and describing the vast number of biomarkers that characterize human health and disease (65). In the not-too-distant future, it may be possible to use this information to prevent or treat disease (66). As Buck Louis and Sundaram have noted (65), moving forward requires greater awareness of the exposome paradigm and a willingness to explore and think in novel ways. The exposome concept offers an expansive view of environmental exposures over the life course and is likely to clarify disease etiology by pointing to new associations with disease. Efforts should be made to combine information from different levels (e.g., the levels of the exposome, genome, epigenome, and metabolome) in order to identify possible causal pathways and opportunities for intervention. Longitudinal studies are needed that include sizeable numbers of persons of various ages who are healthy or have preclinical or clinically diagnosed illness. This will help to rule out reverse causation and facilitate the identification of metabolomic and proteomic markers of preclinical disease and disease progression. Recommendations for the reporting of prognostic marker studies have been provided (67). Additional research is needed to clarify biological mechanisms and the causality of exposure-disease associations so that effective public health interventions can be devised (53). Of particular interest is the identification of mediators of illness and injury that are potentially modifiable. VanderWeele (68, 69) has provided an account of how direct, indirect, and interactive effects can be decomposed in epidemiologic studies that examine mediation and interaction.

A road map for global -omics should logically take into account important ongoing initiatives such as efforts to build upon information obtained from the Human Genome Project and to overcome the limitations of GWAS by sequencing large numbers of entire human genomes (i.e., whole-genome sequencing). Ongoing collaborative efforts to unencrypt the human epigenome are also likely to be key. Large-scale collaborative epidemiologic studies, including projects that have already been undertaken in the United States and Europe, will play a key role in combining genomic, epigenomic, and metabolomic data with information about environmental exposures that individuals incur over a lifetime, while taking into account the timing of exposures (e.g., prenatal, early childhood) and acute and chronic responses to exposures. The potential scientific contributions of ongoing studies and future initiatives will be maximized only if further progress is made in clarifying the human exposome and in refining advanced computational methods that are suitable for studying gene-environment interactions and epigenomic effects using complex multidimensional, dynamic data systems. Progress is being made in bioinformatics and biostatistics to take into account nonlinear relationships due to gene-environment or gene-gene interactions.

A road map for global -omics must also address the need for laboratory resources to build upon existing technologies such as high-throughput assays, microfluid technologies, gene and protein chips, and expanded resources for biobanking (65, 70). The US Department of Veterans Affairs has invested in large-scale biospecimen repositories. The National Institute for Occupational Safety and Health is exploring new technologies and tools for measuring internal and external exposures and developing and validating biomonitoring techniques for monitoring responses to environmental exposures (2). In the future, it may be possible to create environmental chips that measure exposures across the life span.

Finally, efforts to develop, combine, and analyze large, complex data sets should be guided not only by scientific considerations but also by ethical considerations, such as individual privacy and confidentiality of health information, maximizing benefits in an equitable fashion, and policy considerations related to how best to utilize finite resources for health research. An example of the latter is the ongoing policy debate over how many public resources should be spent on genomics versus traditional epidemiologic approaches for preventing disease and protecting health. Efforts to incorporate metabolomics and proteomics into epidemiologic studies will require large amounts of public funding, even though automated laboratory and processing methods have reduced costs. Despite the expense and complexity, however, -omics research is likely to become an increasingly important part of epidemiologic studies in the foreseeable future.

ACKNOWLEDGMENTS

Author affiliation: Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia (Steven S. Coughlin).

Conflict of interest: none declared.

REFERENCES


