Exposome, 2021, **1(1)**, osab002
DOI: 10.1093/exposome/osab002
Advance Access Publication Date: 30 November 2021
Editorial

Integrating the exposome into a multi-omic research framework

Gary W. Miller *

Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA

*To whom correspondence should be addressed: E-mail: Gary.miller@columbia.edu

In June 2021, I was asked to speak to the National Advisory Environmental Health Sciences Council, the advisory body of the National Institute of Environmental Health Sciences (NIEHS). The topic was how to integrate the exposome into the multi-omic framework that exists within the larger National Institutes of Health (NIH). Since the audience of that presentation was limited, I decided to pseudo-transcribe my thoughts for the pages of *Exposome*.[1]

The omic revolution

The ability to measure components of a biological system on an omic scale has transformed science from genomics, transcriptomics, proteomics, metabolomics, to many others.

To those studying the environmental factors that contribute to disease, it can appear as if genomics and its close relatives are gobbling up all of the funding resources, resulting in what appears to be a gene-centric view of disease and health. This leaves the non-genetic contributors understudied. In the *Tyranny of Science* by Paul Feyerabend, the philosopher outlines many ways that the dogmatic rules of science suppress other paths toward knowledge, whether they are from within science itself or from other domains such as philosophy and religion.[2] I would argue that science is not inherently tyrannical, but rather scientists are using the rather dogmatic scientific method(s) that they were trained to use. In a similar vein, it is difficult to fault geneticists for using genetic tools to solve problems especially given how successful the field has been over the past few decades. In fact, the advances in genomics are arguably some of the greatest accomplishments in the study of human biology and disease. That said, we want to avoid a tyranny of genetics. Rather we want to build upon the extraordinary success of genetics by partnering with the field to explore the non-genetic drivers of disease and to better understand how our environment interacts with genes to improve our understanding of disease causation. Ideally, this combined knowledge will lead to strategies to prevent diseases and improve health.

NIH, whose mission is "to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability" must use all of the tools at its disposal to achieve this mission. Given the truism that our phenotype is a result of our genes and environment, then a more robust analysis of our environment is warranted. The NIH institute, NIEHS, is dedicated to understanding the impact of the environment on human health and most of the other institutes already incorporate some environmental factors into their study of disease to a degree. However, the importance of the environment in human disease is still under-appreciated and under-studied. The environmental contributors of disease must be better integrated into the study of biology and they need to be studied in a more comprehensive fashion.

Our methodological reductionism, which strives to study a biological system at the lowest possible level, has served us well for centuries. Advances in cryo-EM and single-cell analysis reveal the power of the reductionistic approach. However, humans operate as a series of integrated systems and it is necessary to take a systems-level approach to understand human health. Geneticists have studied individual genes using the reductionistic approach and this has yielded great results, but they didn't stop there. The field aggressively pushed for and adopted a more systematic and comprehensive approach to uncovering the genetic basis of disease. Study sections that review grant applications started to consider projects that were not hypothesis-driven, a rather uncommon practice. Rather than proposing that a particular gene or genetic pathway was responsible for a specific disease, scientists started proposing genome-wide association studies (GWAS)-based approaches to interrogate the entire genome for that specific disease and the results have been glorious. GWAS is the epitome of the anti-reductionist approach with polygenic risk scores attempting to merge the converging genetic factors of disease. Genetics does an excellent job of balancing a range of reductionistic and systems-level approaches in uncovering the genetic basis of disease.

Thus, even though we should continue our reductionistic efforts to study how environmental chemicals alter protein folding, DNA integrity, and receptor activity as is done in molecular toxicology and through targeted studies, we must also approach the environment on a systems level, which is what the exposome was designed to do. We must strive toward an exposome-wide association study (ExWAS or EWAS) approach that incorporates a comprehensive and systematic analysis of the exposome. GWAS studies are extremely powerful, but they are complemented by the hypothesis-based studies that further examine how a particular gene alters biology and how that gene is regulated. In a similar manner, the exposome-based approach can be hypothesis

Received: August 02, 2021. **Accepted:** August 03, 2021

generating in a systematic manner that can be coupled with hypothesis-driven toxicological and environmental health studies to identify mechanistic insights.

Purveyors of the exposome have been pursuing strategies to study a wide array of environmental insults for many diseases, and in doing so attempt to measure the corresponding biological alterations that result from these exposures. Frequently, exposome scientists are asked to determine the environmental component of samples that have already been analyzed on a multi-omic level. Thus, much of the information needed to identify the corresponding biological responses has already been captured. Such an approach leverages the data generated from genetic, epigenetics, transcriptomic, proteomic, lipidomic, and metabolomic analyses. In addition, these studies typically have rich phenotypic data. Therefore, if the exposome field wants to slide an "exposomic layer" into the multi-omic framework it must do so under the aegis of those running the larger parent studies. This requires a partnership mentality rather than an "us versus them" or "gene versus environment" mindset.

Going to the bank

Scientists interested in the exposome and environmental health would be wise to listen to the apocryphal words of the famed bank robber Willie Sutton. When asked why he robs banks his reported response was "Because that's where the money is." Whether or not he actually said it is of no consequence, as it is self-evident. The biobanks from specific disease studies or larger population-level projects are storing critical information that is literally there for the taking. I have long advocated taking advantage of biobanked samples from these ongoing multi-omic studies and extracting exposomic or environmental value from these stored samples, as well as using biological samples from current and ongoing studies.[3] To take advantage of these precious samples, the exposome field must demonstrate that our methods are reliable, repeatable, rigorous, and reproducible.

Moreover, many of these studies have geocoded information on participants' home and work addresses. With our extensive knowledge of exposures via satellites, ground-based sensors, and chemical usage maps it is possible to construct fairly detailed exposure matrixes by having this geocoded information. The level of resolution from these geospatial approaches may not be sufficient to uncover causative links in individuals, but may identify trends at the community, neighborhood, or regional level. It also allows examination of historical data to examine changes in land use, pollution, or atmospheric conditions. The power comes when one couples that geospatial data with what can be gleaned from that biobanked plasma or serum and the extensive phenotypic and multi-omic data. The individual-level data capture genetic differences in enzyme expression, transport dynamics, and provide specific information on internal dose of hundreds to thousands of chemicals. The merging of geospatial data with multi-omic data is at very early stages, but we know that from an epidemiological perspective, geospatial data can reveal reliable information, e.g. cancer clusters. Statistical methods can reveal if an apparent cluster of disease is truly significant and then the individual level exposome analysis can identify those at highest risk. Although we must continue to be wary of the ethical issues surrounding where people live and health, as it has the potential to impact housing costs, property values, and health disparities.

A multi-omic exposome organon?

What is an organon you ask? An organon is a system of principles for use in scientific or philosophical inquiry. It is analogous to the constellation of beliefs that Thomas Kuhn described in *The Structure of Scientific Revolutions* that helps define a field.[4] If we are going to try to integrate the exposome in NIH-supported multi-omic studies of disease, we must set some standards on how to do this. This starts with a clear operational definition and a specific list of goals.

From an NIH perspective, the exposome still needs to sharpen up its definitions, goals, and approaches. However, many NIH institutes have embraced the concept and are starting to integrate the exposome into their research programs. The exposome definition of "the measure of cumulative environmental exposures and corresponding biological responses" is a good starting point. We need to obtain a comprehensive measure of the environmental exposures and then be able to compare those results to the biological responses. Leveraging existing and ongoing cohorts will let us focus on the data generation, which is the more pragmatic approach. We ultimately will need cohorts that are designed to study the exposome, but we can take giant leaps by collaborating with investigators who have been studying specific diseases and accumulating phenotypic and molecular data. For the sake of discussion, I propose the following working definition for the multi-omic framework:

> The exposome: the cumulative measure of environmental exposures and corresponding biological responses that can be derived from biological samples (blood, urine, tissue).

Although this definition excludes a geospatial component it does not prevent its inclusion at a later time. I would encourage the exposome field to first focus on getting the exposome built into the multi-omic framework of NIH and then build out the geospatial component in parallel. At this point, geospatial data do not have a clear path into the classical mulit-omic framework. This is a challenge that we in the exposome field will need to help solve. If we can develop the methods to merge geospatial environmental data with our exposomics data from human biofluids it will help us interpret the findings and allow identification of sources and potential interventional strategies.

In order for an exposome component to be incorporated into multi-omic studies, the measures must be truly omic. By this, I mean a systematic and comprehensive analysis. We must go beyond targeted approaches. We must incorporate untargeted or non-targeted high-resolution mass spectrometry even if we have thousands of semi-annotated or unannotated peaks. The type of high-density data that comes off of a mass spectrometer, even in an annotated form, blends well with genomic and proteomic data. The approaches developed for statistical genomics are also amendable to the mass spectrometry-generated data. Cheminformatics and bioinformatics are progressing rapidly and the known unknowns and unknown unknowns will be shrinking in number in short order. The environmental epidemiologist and regulatory scientist rely on rigorous measures of targeted environmental exposures. Geneticists and disease-focused scientists, however, are much more comfortable with the uncertainty that comes with

high-resolution mass spectrometry-based upon their use of transcriptomics, epigenomics, proteomics, and NextGen sequencing. We should take advantage of that comfort level to integrate the "exposome layer" into the NIH multi-omic framework.

The NIH is ready for the exposome. Now, it is time for the exposome to be ready for the NIH. The exposome field must develop the organon that clearly defines how the exposome will deliver environmental data to the multi-omic framework. We need agreement on an operational definition for multi-omic research, we need to optimize, validate, and harmonize our methods, and we must utilize the ontologies that allow the seamless integration of our environmental data into the multi-omic system. This will require considerable planning but the potential dividends for human health are extraordinary.

References

1. Miller GW. The Exposome: a new field. New J Exposome. 2021; 1(1):1–2.
2. Feyerabend P, Oberheim E. Tyranny of Science. Polity Press; 2011.
3. Miller GW.f The Exposome: A New Paradigm for the Environment and Health. Cambridge Academic Press; 2020.
4. Kuhn TS. The Structure of Scientific Revolutions. University of Chicago Press; 1962.