



Use of Metabolomics to Advance Research on Environmental Exposures and the Human Exposome: Workshop in Brief

DETAILS

12 pages | 8.5 x 11 | null
ISBN null | DOI 10.17226/23414

AUTHORS

Kellyn Betts and Keegan Sawyer, Rapporteurs; Board on Life Sciences; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine

BUY THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

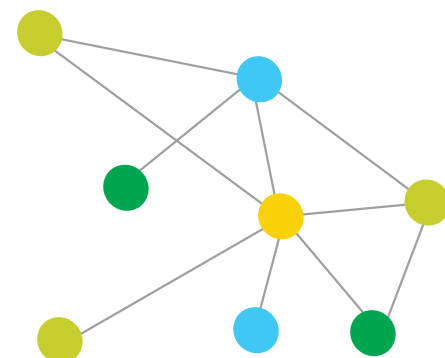
- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

Use of Metabolomics to Advance Research on Environmental Exposures and the Human Exposome

Metabolomics, the scientific study of small molecules produced from metabolism (metabolites) is a rapidly expanding area of research that enables scientists to better understand the physiological state of an organism and its response to different types of stimuli, including nutrients and pollutants. Metabolism is the array of chemical reactions that occur within a living organism to support its ability to grow, reproduce, and respond to environmental exposures, among other processes necessary to sustain life. Metabolites can be created in response to chemicals that originate endogenously (inside the body) or exogenously (outside of the body). Preliminary research suggests that metabolomics holds promise to advance understanding of the *exposome*. The *exposome* includes all of the environmental compounds an individual is exposed to from conception to death. This environmental correlate to the genome, first described in 2005 by Christopher Wild,¹ includes people's exposure to complex mixtures of chemicals, as well as the substances that can be produced in the body when chemicals are metabolized. For this reason, the Standing Committee on Emerging Science for Environmental Health Decisions of the National Academies of Sciences, Engineering, and Medicine held a workshop² to examine the potential for using metabolomics to characterize human environmental exposures and the exposome. Proofs-of-concept were discussed in two case studies on the cause of human Eosinophilic esophagitis and the effect of toxic pollutants on pregnancy in rats. Key workshop themes included technical capabilities and limitations to collect metabolomics data and implications of this new source of data for future environmental and public health research and public health policies.



**A Workshop in Brief for
the Standing Committee
on Emerging Science
for Environmental
Health Decisions**

May 28–29, 2015

¹ Wild, CP. 2005. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* 14:1849. doi: 10.1158/1055-9965.EPI-05-0456.

² Metabolomics as a Tool for Characterizing the Human Exposome: A Workshop, May 28–29, 2015. See <http://nas-sites.org/emergingscience/meetings/metabolomics-and-the-exposome> for archived video presentations and other workshop materials.

METABOLOMICS AND ENVIRONMENTAL EXPOSURES: A VISION

The metabolome's inclusion of all the small molecules involved in the chemical reactions that maintain the body's cells and organs essentially "caps off the pyramid of life," noted David Wishart of the University of Alberta (see Figure 1). Small changes made to the human genome (genetic material), epigenome (chemical compound that can tell the genome what to do), and proteome (the set of proteins expressed by the genome) are easily detected in the metabolome. Hence, metabolites, are "the canaries of the genome," Wishart stated. He emphasized that the human genome applies to the entire body, but every organ system is designed to have a different metabolome.

Wishart also emphasized that researchers may be able to use the metabolome to detect endogenous changes in response to environmental influences, such as environmental chemicals to which humans are exposed through occupation, diet, or other means. More than 90% of deaths and disease in the United States and other developed countries are due to some kind of environmental exposure, either an excess or a deficiency, noted Wishart. A major obstacle to determining how environmental exposures may contribute to disease is the relative dearth of publicly available information on human exposure to the some 80,000 chemicals registered for commercial use, added Roel Vermeulen of Utrecht University. Metabolomics can overcome this because of its ability to provide a broad, agnostic assessment

of the compounds that exist in a biosample, rather than being limited to a chemical or a class of chemicals selected in advance, explained Wishart. This allows the approach to identify exposures and potentially improve surveillance and elucidate emerging stressors.

One difficulty with research on human environmental health is that different disciplines use different measures for environmental exposures. Wishart pointed out that epidemiologists ask research participants to provide a self-report of their location, diet, behavior, and lifestyle. Molecular epidemiologists measure biomarkers inside people to identify internal doses. Exposure scientists measure what is found outside people, such as in the local air or water or on the skin. Metabolomics sidesteps the need for dietary and behavioral surveys and their attendant concerns about the reliability of the information that respondents provide, observed Oliver Fiehn of the University of California at Davis.

Instead, metabolomics measures both endogenous compounds created and assembled by our bodies and the exogenous compounds introduced by ingestion and our environmental exposures; the exogenous compounds are what is relevant to the exposome. Within the body, exogenous compounds may be transformed both by processes, including digestive enzymes, and the actions of microbes and other tissues, Wishart said.

Wishart's group and other researchers have cataloged nearly 42,000 metabolites associated with food, drugs, food additives, phytochemicals, and pollutants. Metabolomics studies have identified environmentally linked biomarkers related to pre-eclampsia, congenital heart defects, fetal growth restriction, chronic fatigue syndrome, and colonic polyps, Wishart said. He described three publicly available, online databases that he and his colleagues created to enable free access to data about the metabolome: DrugBank, Human Metabolome Database, and Toxic Exposome Database (see Box 1).

David Balshaw of the National Institute of Environmental Health Sciences (NIEHS) noted that the 2012 National Research Council report *Exposure Science in the 21st Century: A Vision and a Strategy* observed that the collection of better exposure data can provide more precise information

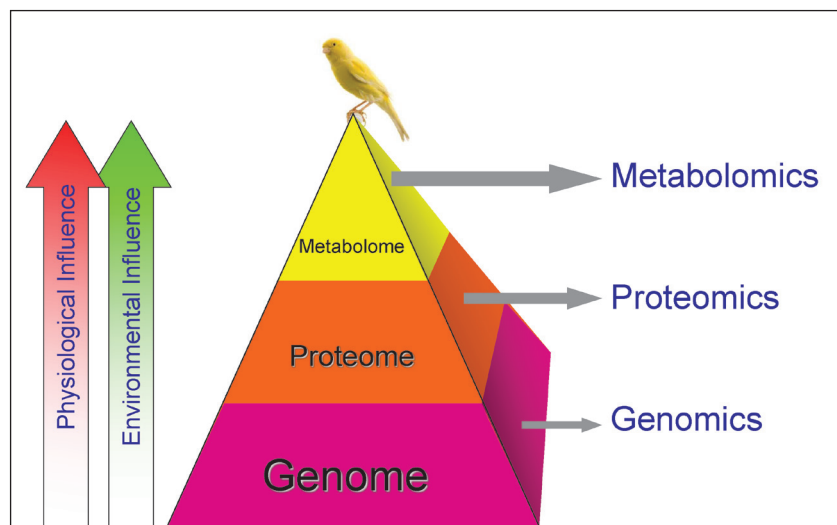


FIGURE 1: The 'Omics Pyramid of Life. Changes in the genome and proteome can be detected in the metabolome. The metabolome also can reflect changes in response to environmental influences. SOURCE: Wishart workshop presentation, slide 5.

regarding risk estimates and lead to improved public health and ecosystem protection.³ The report described in length how metabolomics is being applied to biomonitoring of chemicals in humans and in wildlife. NIEHS is using metabolomics technology to help fulfill the organization's strategic plan to transform exposure science, said Balshaw.

METABOLOMICS FOR EXPOSOMICS

The discovery that genes do not play as large a role as initially expected in disease helped drive interest in the insights possible through metabolomics, said Steve Rappaport of the University of California at Berkeley. Chirag Patel of Harvard University Medical underscored the idea that metabolomics technologies represent an affordable option for comprehensively accessing exposure data. In particular, he emphasized that science is at “a critical juncture in thinking about how to use these new genomics technologies to ascertain the human exposome.” Since 2005, research on the exposome has risen rapidly, noted Rappaport (see Figure 2). A key concept underlying the use of metabolomics to study the exposome is that the observed characteristics and traits of humans and other organisms, their phenotypes, are a function of genes *and* the environment. Metabolomics has the potential to provide insights on the cause of major diseases, including cardiovascular disease and many cancers, said Rappaport.

Rappaport and others are optimistic about metabolomics' potential for providing insight into the human exposome. The large number of genome-wide association studies (GWASs) conducted to date have made clear that genetics plays a relatively small role, less than 20% in some of the most important health conditions affecting humans, Rappaport explained. A recent meta-analysis of more than 2,700 publications in *Nature Genetics*

³ See <http://www.nap.edu/catalog/13507>.

Box 1. Online Metabolomics Databases

DrugBank: contains information about 1,240 metabolites and 1,550 drugs, including drug–drug interactions.

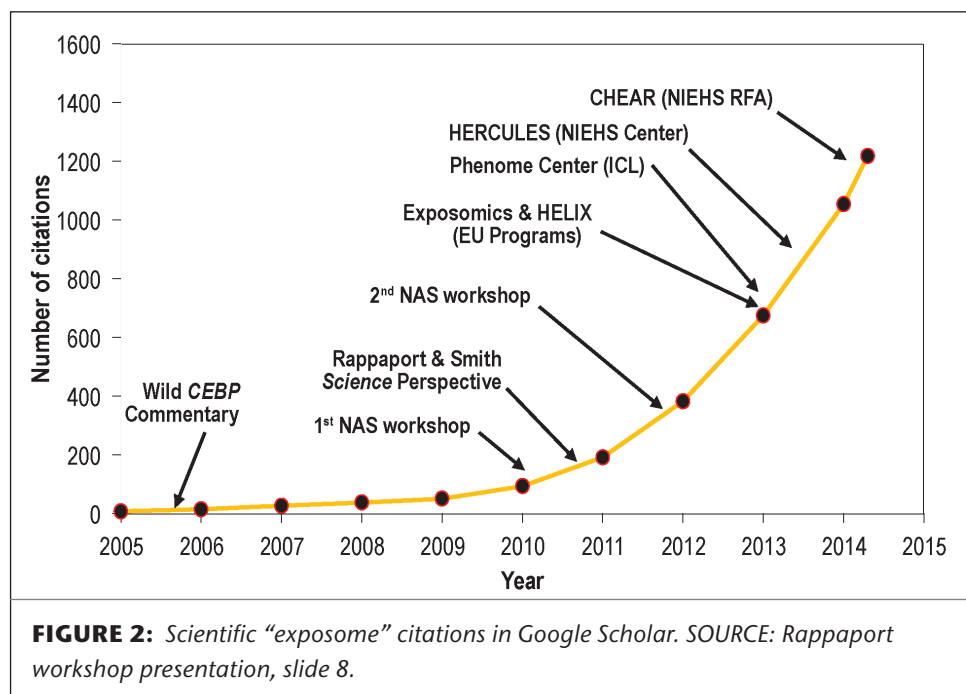
Human Metabolome Database: details information on more than 40,000 quantified, detected, and expected metabolites in humans and their known roles in health and disease.

Toxic Exposome Database: describes 3,670 metabolites, including toxic drugs, pesticides, herbicides, endocrine disruptors, carcinogens, solvents, Polychlorinated biphenyls, and furans.

SOURCE: Wishart workshop presentation, slides 19–21.

looked at more than 17,000 expressed traits studied in more than 14 million human twin pairs and calculated that the average heritability of all traits is 0.49.⁴ This suggests that the exposome can help elucidate the other 50% of traits—those that are not strictly heritable, Patel said.

Although it is improbable that scientists will be able to measure any individual's complete exposome, it is feasible to capture useful data on the exposome via metabolomics. To do so requires what Toby Athersuch of Imperial College referred to as “a meet in the middle approach,” that is, prospective cohort epidemiological studies that begin following people before they exhibit any symptoms of disease,



⁴ Polderman, TJC, et al. 2015. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47:702–709.

sometimes beginning as early as in utero. Exposure information may come from questionnaires and environmental monitoring data. Metabolomics can be used on samples collected through these studies to determine the relationship between exposures and risk-predictive intermediate biomarkers. Such studies would involve selecting some of the people who subsequently become sick and matching them with control individuals to use metabolomics to identify risk-predictive intermediate biomarkers, Athersuch explained. Rappaport's term for such studies is "exposome-wide association studies" (EWASs). The approach has already been used to identify new causes of disease, and Rappaport predicts that EWASs will advance research on the etiologies of disease.

Both NIEHS and the European Union are piloting exposome initiatives. NIEHS has been developing methods and technologies for measuring the exposome for a decade, Balshaw said. Exposome research features prominently into two of the institute's current strategic plan goals on environmental exposures, and metabolomics is part of

NIEHS's efforts to expand and transform its research into environmental exposures. The research also encompasses understanding the role of combined environmental exposures, including the microbiome and nonchemical stressors, such as psychosocial stress. The agency is currently implementing suggestions from a January 2015 workshop on the exposome related to exposome technologies and methods and the need to develop an exposome clearinghouse to facilitate research and data sharing. Balshaw said that the research efforts are benefiting from \$48 million redirected from the National Children's Study, which is now being used to investigate the influence of the environment on children's health via studying the exposome.

The European Union's initiative, which began in 2012, is its largest ever in the environment and health research area. Vermeulen is active in the initiative's exposomics project to link collected exposure data on air pollution and water contaminants with biochemical and molecular changes in the human body. The project's metabolomics work is being led by Augustin Scalbert of the International Agency

Case Study 1

Metabolomics and the Environmental Etiology of Disease

To exemplify the potential of metabolomics, Wishart described his research on Eosinophilic esophagitis (EoE), a chronic immune condition that affects the stomach. Symptoms of EoE include difficulty swallowing, heartburn, and sometimes food impaction. EoE occurs in children and adults, explained Wishart. The incidence of the disease is "exploding almost at the same rate as autism," in many countries, particularly among middle-aged men, Wishart stated. People with EoE often have other allergy-related immune conditions, including asthma, eczema, and celiac disease. EoE is difficult to diagnose with limited treatment options and is often confused with gastroesophageal reflux disease (GERD).

EoE's rapid rise in incidence throughout the world suggests an environmental link, as do the discrepancies between the numbers of cases in differing areas, Wishart said. Smoking appears to be protective, while early life exposure to antibiotics and wheat substantially increase risk. Wishart and colleagues conducted an untargeted metabolomics study based on urine samples from approximately 30 age and gender-matched EoE patients and 30 controls. They found

unexpected and striking differences in heavy metals, including lead, manganese, cadmium, and barium. Additional research correlated the findings with genetic studies linking the disease to reduced levels of metallothionein, a metal binding protein that can grab heavy and toxic metal to keep them out of circulation. This can conceivably result in heavy metals being released in urine, Wishart said. Unfortunately, the researchers have not been able to identify an obvious source of exposure to the metals.

In a subsequent study involved with EoE, GERD, and healthy patients the researchers used several spectrometry, chromatograph, and bioinformatics tools in their analysis to look for metabolomics distinctions among the three patient groups. Wishart and his team succeeded in identifying urinary biomarkers that could be used in diagnosing the condition. The research not only revealed identifying distinctions, but also that many of the metabolites that distinguish the EoE patients were of bacterial origin. Their finding supports the possibility that EoE symptoms are linked, and potentially caused by, changes in the gut microbiome.

for Research on Cancer, and the group is building a database it calls Exposome-Explorer with information about biomarkers of environmental exposure potentially relevant to diseases. The effort includes collecting data on expected ranges of exposure in different populations to illuminate correlations and patterns between exposure and disease, Vermeulen said. It is essentially “a reference exposome database,” he said.

ADVANCES IN MEASUREMENT TOOLS

A range of tools can be used to analyze metabolic compounds in biological samples. Wishart and Athersuch discussed some of the different analytical chemistry instruments being used in metabolomic studies, including nuclear magnetic resonance spectroscopy, direct infusion mass spectrometry, time of flight mass spectrometry, and mass spectrometry coupled with gas chromatography or liquid chromatography (LC), and inductively coupled plasma mass spectrometry. Additionally, Wishart pointed out that the glucose meter is one of the most widely used metabolomics devices.

How these tools are used for metabolomics is rapidly evolving and improving, Wishart said. Emerging trends in metabolomics include efforts to miniaturize and automate the process. Hand-held devices are becoming available that measure anywhere from 10 to 30 metabolites at a time.

Scientists have recently devised ways to use matrix-assisted laser desorption/ionization, a soft ionization technique used in mass spectrometry, to image and measure metabolites at very low masses.

The main approach to metabolomic analyses discussed by workshop participants is *untargeted*, where researchers simply try to identify as many metabolites as possible in a given sample. This is also known as the chemometric, agnostic, non-targeted, and data-driven approach. “Untargeted analyses get at the heart of the epidemiological question, ‘what are the causes of disease,’” Rappaport commented. Alternatively, researchers can use a *targeted*, knowledge-driven approach to focus on particular metabolites.

Many of the methods used for untargeted analysis were developed by the Human Metabolome Project (HMP), which had the goal of characterizing all of the metabolites in the human body in a range of fluids and tissues, including saliva and cerebrospinal fluid, as well as blood and urine. Another way to measure exposure is through exhaled breath.

Because many of the compounds discovered through metabolomics studies have not been cataloged, researchers often use multiple platforms and multiple passes to identify samples by generating data that can be compared and contrasted. For example, Fiehn pointed out that an effective combination is gas chromatography coupled with time

Case Study 2

Metabolomics and Cross-Generational Effects of Environmental Exposures

Susan Sumner of the Research Triangle Institute described surprising results brought to light by her work with metabolomics in cross-generational rodent studies over the past decade. During 2006, when metabolomics was relatively new, Sumner was working on a study involving pregnant rats dosed with butyl benzyl phthalate that involved collecting urine and tissues from the offspring at 26 days after delivery. Long after the butyl benzyl phthalate was gone, her group observed differences in the urinary and tissue metabolomes of the offspring of exposed mothers. The results suggest that the “memory” of exposure and what it does to endogenous pathways may be a fruitful area of future research.

Similarly, Sumner said that her work with human cord blood to investigate the “in utero exposome” revealed detectable differences in the metabolites

of fetuses whose mothers were exposed to arsenic. Similar studies have been conducted to tease out the impacts of polycyclic aromatic hydrocarbons on the in utero exposome with Federica Perera at Columbia University, Sumner said. They are currently working on how to model the data. Other studies looking at neonatal exposures suggest that some metabolomic differences may be protective. For example, hydroxybutyrate injured rat kidneys but showed a protective effect against the herbicide paraquat. Researchers at the University of California at Berkeley documented reduced breast cancer incidence and mortality in communities with high arsenic exposure. “We have to be really careful about how we are interpreting this data when we see something change—is it a good change or is it a bad change,” Sumner pointed out. Sumner’s findings emphasize the need to understand the relevance of the data being generated.

of flight instruments and mass spectrometry and LC coupled with electrospray ionization, as well as hydrophilic interaction LC with mass spectrometry.

Blood and urine samples are currently the main biosamples used in metabolomics studies. More than 3,500 compounds have been documented in the human serum metabolome, and more than 2,600 have been identified in the human urine metabolome. An effort led by Rappaport to catalog all of the molecules measured via metabolomics in significant numbers in human beings shows that drugs, foods, and endogenous compounds tend to be found at very similar levels.⁵ Environmental chemicals exist at much lower levels. “The dynamic range [of these molecules] is enormous, spanning 11 orders of magnitude,” Rappaport said.

Databases created through the HMP include information about more than 42,000 compounds, including ones that are expected to exist but have not yet been formally detected. When possible, the data include associated information, such as normal and abnormal concentrations and links to diseases.

COLLECTING METABOLOMICS DATA: TECHNICAL CHALLENGES AND OPPORTUNITIES

Metabolomics relies on analytical chemistry for identifying metabolites and analytical chemists face a number of unique challenges in identifying these small molecules. Because metabolites can occur at a wide range of concentrations, from milli-Molar to femto-Molar, identifying them all requires a group of instruments, Wishart said. “No one technology is all-encompassing,” said Anthony Macherone of Agilent Technologies.

Measurements covering thousands of small molecules are desired to understand how they change over time and within large cohorts, which requires both a high sensitivity and a large dynamic range, Erin Baker of Pacific Northwest National Laboratory pointed out. Biological changes are best understood when xenometabolites and endogenous metabolites are measured simultaneously, she said. This involves collecting lots of information about many disparate types of chemicals as rapidly as possible. These aspects of the task make it very different than what is involved in many other kinds of analytical chemistry, Macherone stressed. Accurately detecting individual metabolites requires that chemists “differentiate real

signals from the din of noise,” Macherone said. This involves distinguishing what is real and significant from normal variation and background noise, he explained. “We need to be able to do this from a qualitative and quantitative perspective, and, most importantly, with reliability,” he said.

Compounding the challenge is the reality that many small molecules have the same masses but a different chemical makeup. For example, Baker mentioned a National Institute of Standards and Technology (NIST) database that includes 18 biological molecules with the same mass as testosterone, to 7 significant digits. This renders differentiating between them using approaches based on mass spectrometry difficult, she said.

The capabilities of the analytical equipment offered by different vendors are somewhat similar, Macherone observed. LC coupled with electrospray ionization, as well as hydrophilic interaction LC (HILIC) and reversed phase chromatography, “covers a great deal of the exposome space,” he said. He added that these technologies are particularly useful for identifying small polar molecules. Gas chromatography coupled with time of flight instruments (GC-TOF) with or without mass spectrometry can identify many primary small metabolites, Fiehn said. GC-TOF is also good at identifying volatile compounds, he said. Gas chromatography with electron ionization mass spectrometry is helpful for non-polar compounds, as well as those that are volatile or semi-volatile, Macherone said.

HILIC also is a good technology for secondary metabolites, according to Fiehn. Because nuclear magnetic resonance spectroscopy can be used to identify a wide range of chemicals, it can be helpful for screening, but there can be sensitivity issues, agreed Macherone and Sumner. Inductively coupled plasma mass spectrometry is helpful for identifying metals, Athersuch said. “There are lots of ways to approach the problem” posed by attempting metabolomics analyses, Sumner commented.

A newer technique for conducting non-targeted analyses aimed at identifying xenobiotics is ion mobility spectrometry (IMS), a fast separation technique coupled with mass spectrometry (IMS-MS). Baker said that its use can increase confidence in xenometabolite discovery because of the ability to distinguish between chemical isomers that contain the same numbers of atoms in different configurations. She is creating an IMS-MS small molecule database to buttress the technique’s utility. Baker has also investigated how Field Asymmetric Waveforms and RapidFire cartridge approaches can be coupled

⁵ Rappaport, SM, et al. 2014. The blood exposomes and its role in discovering causes of disease. *Environmental Health Perspectives* 122(8):769–774. doi: 10.1289/ehp.1308015.

with IMS to enhance resolution, sensitivity, and speed. The very high throughput of RapidFire-IMS-MS removes some key obstacles to assessing the exposome, she said. It can enable higher-confidence identification of xenometabolites and be used to rapidly generate libraries. Both IMS-MS and RapidFire-IMS-MS are promising for use as screening platforms, Baker said. The RapidFire-based approach may also help with method development, Macherone noted.

Macherone described an alternative to untargeted and targeted metabolomic analyses that he termed semi-targeted analyses. He explained that untargeted screening analyses may not detect 70% of the persistent organic pollutants linked to the exposome and typically do not use validated methods. Targeted methods tend to be validated, more rigorously developed, and the most sensitive. Semi-targeted methods combine aspects of both targeted and screening methods to screen for unknowns but also target other compounds that can be quantitated. An example would be a situation involving 60 compounds that a researcher wants to quantitate, in addition to interrogating that data for everything else, he explained. Macherone also described what he called *blended* methods that rely on bioinformatics in addition to semi-targeted approaches. The bioinformatics tools can aid in combining information produced by analyzing the samples using different equipment. Using the bioinformatics tools to combine and align data allows “much broader coverage” on the order of tens of thousands of compounds, he said.

ANALYZING METABOLOMICS DATA

Metabolomics data can be collected through a range of chemistry tools and methods, but analyzing the data requires additional approaches that go beyond those conventionally used by analytical chemists. Both Fiehn and Pieter Dorrestein of the University of California at San Diego have created tools to help researchers analyze metabolomics data. “Whenever you deal with big data, there are two things you have to think about,” organizing it and visualizing it, Dorrestein said. “Without these abilities, it is nearly impossible to interpret the data or to create a hypothesis,” he stated.

The tools that Dorrestein and Fiehn developed help surmount the fact that many of the compounds revealed by metabolomics have never been catalogued. An untargeted analysis from a 50-microliter blood sample may identify more than

1,000 compounds, but in many cases 50% or more of these compounds are unknown, Fiehn pointed out. Dorrestein stated that for every molecule with a chemical identifier, there are 50 that cannot be identified with mass spectrometry.

Other factors contributing to the large number of unknowns include the reality that the action of enzymes can produce novel—and sometimes very odd—compounds, Fiehn stated. Chemical damage can also occur during sample preparation and analysis, he said. Key to determining what is in a metabolomics sample is how its MS peaks are defined and identified, Fiehn said. Because there are many ions per compound, including isotopes, adducts, and fragments, mass information alone may not be sufficient to definitively identify them. The Fiehn group’s free MS-DIAL software program works with data from tandem mass spectrometry (MS-MS) analyses that fragment compounds to capture information about their structure, which dictates how they will fall apart. The software capitalizes on an algorithm developed more than a decade ago to evaluate ion mobility to identify where an ion belongs in a series of peaks. The software also incorporates the automated BinBase database analysis tool for annotating mass spectra to evaluate whether a spectra has been logged before and whether it represents a “good” peak. Fiehn claims that when using the software, “false positives are very, very rare.”

The Dorrestein group’s tool incorporates what he called “a similarity scoring function” to identify how similar molecules and molecular fragments are based on MS-MS data. Although the group’s Global Natural Products Social Molecular Networking software platform is not published, it has more than 4,000 users in more than 80 countries. The software’s knowledge base includes what chemists know about how molecules fall apart to find related molecules and consider how metabolism will act on the compounds. It also allows for additional attribute data, such as biological information, to be associated with spectra data and enables users to share and compare their spectra with all of the other spectra in the database. Dorrestein’s software alerts a researcher if portions of his or her dataset match someone else’s dataset. Recently, he said, the software matched a component of a spectrum linked to coral reef bleaching with one identified by a researcher studying cystic fibrosis. “It turns out they have the same inflammatory response ... that manifests itself in the same type of lipids,” he said.

The task of analyzing metabolomics data is hampered by the lack of a universal library for MS data, Fiehn, Dorrestein, and Patel agreed. MassBank of North America, NIST's Mass Spectrometry Data Center, and Scripp's METLIN are publicly available metabolomics databases that, when considered together, contain spectral data for 40,000 compounds, but more than 40 million compounds are listed in PubChem and ChemSpider.

Other tools for helping identify compounds uncovered in untargeted analyses that Fiehn mentioned included a group designed to make key predictions associated with metabolomics data. LipidBlast predicts MS-MS spectra. CarniBlast predicts LC retention times. AcylCoABlast predicts LC-MS-MS spectra. archaeBlast predicts MS-MS spectra linked to the microbiome. Ab Initio predicts MS-MS spectra from first principles.

WHERE TO GO FROM HERE?: IMPLICATIONS FOR RESEARCH DIRECTIONS AND PUBLIC POLICY

Workshop attendees discussed a variety of ways to help metabolomics move forward. Dean Jones of Emory University commented on the value of identifying which compounds everyone tends to share. His group's research suggests that somewhere between the "high-hundreds and 8,000 compounds may be part of our shared metabolome. Having greater communication between different databases would be valuable to the research community," Dorrestein observed. A tool for helping identify combinations of markers that may indicate exposure to a single entity could prove valuable, Wishart said. Along the same lines, Sumner pointed out the potential value of looking for "imprints" from drug exposures. Jones also mentioned his desire to have a way to identify and differentiate between adaptive versus potentially adverse responses. Sumner and Andrew Patterson of Pennsylvania State University agreed on the value of taking repeat samples from research subjects over time to see what changes and what remains constant.

The metabolome holds promise for helping to quantify endogenous chemicals that also have exogenous exposures so we understand relative "backgrounds" from physiology versus nutritional or other environmental sources, Daston said. He and fellow committee member William Farland, professor of Environmental and Radiological Health Sciences at Colorado State University, agreed that this could prove to be some of the "low-hanging fruit" for

metabolomics because scientists already know the identity of many compounds produced endogenously, with both standards and clinical ranges for essential parts of the diet and precursors that scientists know are or should be there.

The harder part from a risk standpoint is the lack of understanding regarding how much additivity may exist or if the endogenous load is sufficient to produce disease states, Daston cautioned. For example, "we know that there are associations between estrogens and certain cancer types, but we really don't have a strong grasp on how much," he said. Another more challenging issue, he added, is determining what components in a complex mixture may have the same biological target and therefore contribute to the same disease—or health-promoting states, in the case of vitamins—as the endogenous materials. "We probably have the technology, in terms of high-throughput and cellular assays, to start to address that, but we need to have the exposure information to know how much there is and in what combinations," he said.

Metabolomics has the potential to help identify the causes of environmentally mediated disease, observed Elaine Cohen Hubal of the U.S. Environmental Protection Agency's Office of Research and Development. Although risk assessment and chemical evaluation are separate exercises, they must be carried out in the context of what causes disease. Metabolomics can help provide answers in both areas, she said.

Benjamin Blount of the Centers for Disease Control and Prevention's (CDC's) biomonitoring laboratory, pointed out that the amount of biological sample available plays a role in determining what is possible. He pointed out that the amount of blood collected in CDC's U.S. National Health and Nutrition Examination Survey (NHANES) limits how many of the 200-plus targeted compounds can be analyzed in NHANES samples. For determining causation, the ability to determine the dose to the target tissue is a very important concept, agreed Blount and Hubal.

Standardizing and harmonizing methods so the results can be compared is very important, Blount said. Committee Member Ivan Rusyn of Texas A&M's Department of Veterinary Medicine and Agricultural Sciences pointed out the potential value of having researchers focus on a set number of chemicals and trying to work out methods that would be portable from laboratory to laboratory. This could allow many more chemists to reliably measure at least parts of the metabolome and help move the science forward, he said.

Rappaport and other panelists and some audience members detailed reasons why exposome data collected through metabolomics are likely to become more important in the coming years. The field already has amassed more weight of evidence than existed at the beginning of the human genome project, Wishart pointed out. “The proof of principle is there ... arguably we should be getting \$2–3 billion that the [National Institutes of Health] put toward the Human Genome Project,” he said. Ironically, he said his work is not fundable by conventional National Institutes of Health grants.

“If we had tried to sequence the human genome by scattered efforts featuring individual genes, we’d still be doing it now,” Wishart continued. He credits the concerted top-down effort for advancing the field to where it is now, where researchers can print their own microarrays. Without such a top-down effort to standardize methodology, findings on the human genome would not be replicable. Athersuch agreed on the “great need” for national and international collaboration to try to harmonize protocols for international sharing of the data generated from precious sample resources.

Russell Thomas, director of EPA’s National Center for Computational Toxicology, said that his agency, together with the Tox21 Consortium, has curated a set of more than 8,300 compounds, including industrial, pharmaceutical, and environmental chemicals, which can be used for metabolic process standards and identifying metabolites. EPA also created mixture formats as well as metabolic clearance assays that can be used to derive metabolic process standards. EPA is already beginning to make formats available, Thomas said. Having a collection of chemicals to use as “pure standards” to determine the metabolic process from them would be a good basis for moving forward, Wishart observed. Fiehn contended that it is still a bit too early to determine what standards and platforms to use because discovery work is still needed. But he and Blount agreed on the need for the concerted effort by agencies, together with vendors and maybe committees from research labs, to provide awards to standardize methods and codify best practices.

Fiehn also raised concerns that NIH policy may render data produced by some labs inaccessible. In response, Balshaw said that this issue is a topic NIEHS has discussed a great deal. Balshaw said that NIEHS’s goal is to gather as much data as possible, but added that “there are limitations on what we can make publicly available.” CDC developed a model for making de-identified versions of its NHANES data

available. “Some places will allow de-identified data to be shared,” he said. “We have been pushing internally to try to change the policy.” Some investigators may need incentives to make their data and datasets publicly available, said Chirag Patel, an assistant professor at Harvard Medical School’s Center for Biomedical Informatics.

Robert Briggs pointed out the need for what he called “anchor endpoints” for biomarker analytical chemistry. He suggested that instead of trying to come up with a standard for every metabolite out there, it may make sense to find out what is stable over a long time and let everyone measure that.

Farland challenged workshop attendees to identify studies that would demonstrate the proof-of-principle to allow the community to show that the science deserves a concerted effort and funding. Wishart pointed out that when the Human Genome Project was announced, only a few genetic diseases had been identified. He argued that research such as the work led by Stanley Hazen of the Cleveland Clinic using an untargeted metabolomics approach to produce small-molecule metabolic profiles in human plasma to predict risk for cardiovascular disease, which has been further developed in more recent projects, provides a strong justification.⁶

And other examples are coming out now focusing on diabetes and cancer that strengthen the case, he added.

Jones pointed out that a proof-of-principle already exists for some animal models: a recent study in mice dosed with atrazine used metabolomics to provide a window into the metabolic network structure. Teaming that research with an investigation into atrazine metabolites in humans would provide insight into the internal dose–response. “I think the principles are there,” he said. “It’s a matter of getting capabilities that are widespread so that metabolomics is part of the fabric of all of our research institutions.”

Moving forward will require addressing challenges in study design, exposure signal assessment, and data analysis, Vermeulen said. “We’re slowly changing from the idea of single disease entities to overlapping clinical categories,” which means thinking more about the phenome and diseaseome, he said. This also helps us move toward understanding adverse outcome pathways (AOPs) based on changes to the metabolome associated with

⁶ Wang, Z, et al. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57–63. PMID: PMC3086762.

chronic low-level exposures and biomarkers of response, Farland said.

Daston agreed on the need to find a way to make connections between this use of metabolomics to identify the exposome and providing insights in AOPs. Hubal said that her agency's water ecological research program is working on effects-based methods for monitoring effluent using toxicogenomics and other *in vitro* assays. The researchers overlay what the tests and assays show are being perturbed with the AOPs of interest based on the kinds of endpoints that might be regulated. This allowed them "to start to piece together how the sources are related," she said.

Daston stated his belief that metabolomics research into the exposome will help scientists fill in the gap in the middle ground between what we know about disease states and AOPs. He pointed out that researchers know a lot about disease states and the initial chemical interactions with biology, and there also is an increasing awareness of receptor interactions, adduct formation, etc. Metabolomics data can help flesh out the vast middle ground between these two ends of the disease spectrum and provide information needed to characterize AOPs. "Having greater metabolomics data [may help identify] the next level of logical organization that leads from those initial interactions to disease state on the organ or organismal level," he said. The tools may also help pinpoint chemicals that might have different molecular targets but all converge to produce the same sort of toxicity to an organ such as the liver or kidney, he added.

Anna Navas-Ancien of Johns Hopkins University observed that it will be critical to have AOP curves to know which outcomes may be adversely affected. The potential for unmeasured confounding and confounding in general "is a major challenge in this framework," she said. This includes the potential for measurement error and selection bias, she observed.

The metabolome can help us identify individuals or populations that are more sensitive or susceptible by flagging pre-existing disease or concurrent exposures that impact the response to an exogenous exposure, said Standing Committee member Joyce Tsuji of Exponent Engineering and Scientific Consulting. She pointed out that chemical metabolites may be the source of adverse responses. "If we can measure differences among individuals in their metabolism and connect that all together with the exposome, we might have a better understanding of who is more sensitive and likely to have those

adverse outcome pathways, versus other people." This line of research might also bring to light other factors that trigger susceptibility, such as other exposures or a person's nutritional state, but she stressed the importance of separating causes from effects.

"We are getting to a really interesting point in time where we have ever-increasing resources that we can use to mine data," Vermeulen observed. In the European Union, more than 280 organizations from more than 30 countries are working together to collect biosamples from millions of people, including both adults and children in different populations, he pointed out. For example, Vermeulen is involved in the European Prospective Investigation into Cancer and Nutrition study,⁷ a large cohort study with more than half a million participants from 10 European countries. This includes large contrasts in "exposure experiences" from environmental, food, and other sources, he noted.

One issue that may be solved in the longer term is the reality that while the main biosamples are blood and urine, many of the metabolites of interest have a relatively short residence time in those biosamples, Vermeulen said. For next-generation biobanks we may want to consider increasing the range of biological samples collected to include some with longer residence times to provide more information about accumulated exposures, including hair, teeth, and adipose tissue, he said.

Amassing banks that contain biosamples collected before any signs of disease emerge will be very important for identifying metabolomics signatures that are predictive of disease and could in time both help pinpoint causes and aid in early diagnosis, Rappaport said. "The U.S. has not developed cohorts, especially starting in very early life, when many diseases get started, and bio-repositories of fluid in epidemiology studies," he commented. In comparison, the Europeans have set up a number of cohorts from which good bio-specimens are being collected, he said.

Rappaport's group is using metabolomics and adductomics to study small amounts of blood from the "blood spots" collected from California babies born since 1986 who have developed childhood leukemia. "We're seeing a lot of features, not necessarily molecules, that can be tested between cases and controls," he said. The ability to collect good exposure data is key to efforts to use samples collected historically, Vermeulen and Rappaport agree. "What is needed is a convergence

⁷ See <http://epic.iarc.fr>.

of information,” Vermeulen said. Ways to do this is by building up databases showing what kind of interpretation to give to perturbations, and running controlled experiments related to exposures that appear to be likely candidates.

Vermeulen brought up the need to try to gauge the long-term impact of very early life impacts during windows of susceptibility. As an example, he pointed to data showing that younger girls (ages 2–9) who experienced more severe hunger during the “hunger winter” of 1944-1945 had higher rates of breast cancer later in life than older girls (>age 10) who experienced severe hunger. A potential source of data now and in the future is the very early birth cohorts, who are now in their 20s. He also noted that multi-generational cohorts could be valuable study subjects. Carolyn Mattingly of North Carolina State University responded that this was a good example, but she pointed out that there might not be an exposomic signature decades after an event. She reminded the attendees about the value of keeping in mind that context is important. “Using metabolomics interchangeably with exposomics may get us into trouble,” she cautioned. Not all metabolites are the result of chemical reactions with environmental exposures. Rusyn agreed. He argued that while much of the focus in metabolomics is on understanding how changes in small molecules and hormones can help us understand the causes, progression, and outcomes of diseases, much of the actionable intelligence is on the exposure side. Metabolomics data could help bolster efforts to assess what people have been exposed to and

whether metabolomics responses may be correlated to experimental models, either in vivo or in vitro studies. Finding ways to ensure that animal and cell experiments use concentrations of doses that represent realistic exposures is also important. “We know that metabolomics shows us some indicators of pre-existing disease and lifestyle that can combine with environmental exposures to produce adverse outcomes,” Farland added. If metabolomics indicates that a person has problems with kidney function or liver damage, this may suggest that the person will be more susceptible to some environmental pollutants. “The ability to couple the findings of those exposures with individuals whose metabolome has that type of signature is very powerful,” he said.

In the future, Vermeulen expressed hope that data will be aggregated into multi-generational cohorts to address life-course epidemiology. From an exposure standpoint, follow-up data collection would be very helpful. It will also be important to find ways to integrate across age, omics, and health outcomes, he said. How to deal in a sophisticated way with the data about simultaneous exposure to a wide range of mixtures is another challenge, he said.

“There are a variety of ways that metabolomics and exposomic approaches can help us to understand the harm caused by different exposures,” stated Blount. Where we go from here with the metabolome and the exposome, moving toward risk assessment, is going to be a multi-step process, he concluded.

DISCLAIMER: This Workshop in Brief was prepared by Kellyn Betts and Keegan Sawyer, PhD, as a factual summary of what occurred at the meeting. The planning committee's role was limited to planning the workshop. The statements made are those of the authors or individual meeting participants and do not necessarily represent the views of all meeting participants, the planning committee, the Standing Committee on Emerging Science for Environmental Health Decisions, or the National Academies of Sciences, Engineering, and Medicine. The summary was reviewed in draft form by Toby Athersuch, Imperial College; Erin Baker, Pacific Northwest National Laboratory; Elaine Cohen-Hubal, U.S. Environmental Protection Agency; Neha Garg, University of California at San Diego; and Martyn Smith, University of California at Berkeley, to ensure that it meets institutional standards for quality and objectivity. *The review comments and draft manuscript remain confidential to protect the integrity of the process.*

Planning Committee for the Workshop on Metabolomics as a Tool for Characterizing the Exposome: **William H. Farland** (*Chair*), Colorado State University; **David Balshaw**, National Institute of Environmental Health Sciences; **Ana Navas-Acien**, Johns Hopkins Bloomberg School of Public Health; **Chirag Patel**, Harvard University; **Ivan Rusyn**, Texas A&M University; **David Wishart**, University of Alberta; **Lauren Zeise**, California Environmental Protection Agency.

Sponsor: This workshop was supported by the National Institute for Environmental Health Sciences.

**About the Standing Committee on Emerging Science
for Environmental Health Decisions**

The Standing Committee on Emerging Science for Environmental Health Decisions is sponsored by the National Institute of Environmental Health Sciences to examine, explore, and consider issues on the use of emerging science for environmental health decisions. The Standing Committee's workshops provide a public venue for communication among government, industry, environmental groups, and the academic community about scientific advances in methods and approaches that can be used in the identification, quantification and control of environmental impacts on human health. Presentations and proceedings such as this one are made broadly available, including at <http://nas-sites.org/emergingscience>.