Chapter 6: Unravelling the exposome: conclusions and thoughts for the future

Sonia Dagnino*

MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, UK

*Corresponding author: sonia.dagnino@imperial.ac.uk

Abstract

Since it was first defined by Christopher Wild, to today, the concept of the exposome has evolved into a valuable tool to evaluate human exposure and health. In the chapters of this book the definition of the exposome paradigm and concept is discussed. The most recent techniques based on OM-ICS, targeted and untargeted analysis for its characterization are described. The challenges arising from the amount of data that needs to be processed to understand the complex mechanism behind exposure and health outcome, as well as the latest statistical and data analysis methods have been discussed. Finally, multiple projects across the globe have, or are currently, tackling the application of the exposome concept to real studies, and these studies have been presented in this book. In this section, we will provide a chapter summary, and describe how the exposome paradigm has advanced since its definition. We will also explore questions such as: what have we learned so far? Does the exposome paradigm provide valuable information? And what needs to be improved? Finally, we will discuss the possibility of future applications of the exposome in disease causality and personalized medicine.

The evolution of the exposome through time

Since the completion of the Human Genome Project, the study of the genome was thought to be the key to understanding the causes of chronic diseases and mortality. Thus, the extensive studies carried on GWAS (genome-wide association study), has shown that this is not the case. It has now been widely accepted that only a small proportion of non-communicable diseases can be explained by the study of the Genome (Lichtenstein et al. 2000; Saracci and Vineis 2007; Rappaport 2016), and that the study of the exposome is a key element in the understanding of disease etiology. Since the first appearance of the terminology, the definition of the exposome has
Christopher Wild (2005) described the exposome as the ensemble of factors that “encompasses life-course environmental exposures (including lifestyle factors), from the prenatal period onwards” (Wild 2005). For the first time, the environmental exposures were considered as a whole that needs to be assessed simultaneously in order to understand its influence on the body. This novel idea was further expanded upon in 2010-2011 by Martin Smith and Stephen Rappaport (Rappaport 2011; Rappaport and Smith 2010). Therein, they highlighted that the current state of epidemiological knowledge pertaining to the impact of environmental exposures on human health was extremely poor. They further posited that epidemiologist should shift their views from narrow hypothesis considering one or two environmental risk factors to a more comprehensive view considering total exposures throughout individuals’ lifetime (Rappaport 2011).

In 2012, Wild stratified the exposome into three domains (Wild 2012):

i) The general external exposome: including the living environment, climate, social class, education

ii) A specific external exposome: including diet, smoking habits, physical activity, occupation etc.

iii) And finally, an internal exposome: which includes internal and biological factors, defined by the metabolome, microbiome, inflammation, oxidative stress and ageing.

This new description of the exposome stressed the need to carry transdisciplinary research, going from “molecular mechanisms, biotechnology, bioinformatics, biostatistics, epidemiology, social sciences and clinical research” (Wild 2012) in order to tackle the exposome in its whole.

More recently, the definition was expanded by Miller and Jones to include behavior. This definition goes beyond lifestyle as it also includes interactions with our surroundings (relationships, stress, physical emotions etc.) and considers endogenous processes that potentially modify exposures (Miller and Jones 2014). The revised and comprehensive complexity of the exposome is shown in Figure 1.
Although it is still in its infancy, the concept has been growing and several attempts have been made to define it and characterize it. The number of citations of the “exposome” in PubMed has risen from 17 in 2011 to 76 in 2017 (Figure 2). Indicating a growing interest in the subject.

The reason for its slow development relies in the difficulty of the measurements. Measuring the totality of all exposures at a given time point already presents itself with huge challenges. Additionally, characterization of the exposome requires longitudinal measurements – all exposures through lifetime – creating a dense, 3-dimensional array of data and meta-data. This rich dataset must be analyzed with sophisticated bioinformatics tools to identify associations between exposures and disease.
The first attempt to define measuring the exposome was put forward by Rappaport, who proposed two complementary methods to tackle the exposome (Rappaport 2011). The first method was called a bottom-up approach. This consists of measuring all external exposures by analyzing environmental samples (air, water, food etc.) and linking them to health status. The second approach, is top-down. This approach consists of measuring exposures and internal processes by sampling individuals’ blood, urine or other body fluids and linking them to health outcomes. Both methods have limitations, the bottom-up approach does not consider internal processes, such as metabolism, oxidative-stress, and inflammation but it does provide a link to the exposure source. In the top-down approach, the analyses consider all exposures as well as metabolites, by-products, and internal effects, but it does not identify the source of the exposure. In 2010, the national academy of science organized a symposium “The Exposome: A Powerful Approach for Evaluating Environmental Exposures and Their Influences on Human Disease”, the discussion concluded that the top-down approach would make more sense, as it allows to “focus on specific toxicants that have a known effect on human health”(National Academy of Science 2010). The discussion also suggested the use of OMICs technologies, which had already proven efficient for biomarker discovery, and should be applied to the study of the exposome. These thoughts lead to the birth of several large international projects for the study of the exposome.

The study of the exposome: international projects

Over the past decade, multiple exposome-related projects and centers have been created. The main projects and centers are summarized in Table 1. Most of these projects have been described in detail in Section 5 of this book.

In Europe, the REA (Research Executive Agency) has devoted a significant attention to the early understanding of the exposome. Financed by the Seventh Framework Program for Research and Technological Development (FP-7), three major projects have been created: EXPOsOMICs, HELIX and HEALS. EXPOsOMICs has been the first large project created for the understanding of the exposome (Vineis et al. 2017). The project has focused on a top-down approach, focusing on air pollution and water contamination. Although this project has now officially ended, a great amount of data has been produced and analysis is still ongoing. Outputs of EXPOsOMICs in-
clude the production of a European database combining OMICs and environmental exposure measurements on over 3000 individuals. The outcome of EXPOsOMICs will further guide future projects and research in the exposome.

The HELIX project (Vrijheid et al. 2014), is devoted to understanding how the exposome influences health at an early stage. The in utero as well as early-life exposure are critical as they can influence health outcomes further in life. Exposure science had, until now, focused primarily on evaluating single exposure-health effect relationships. HELIX provides the framework to assess the effect of multiple exposures in these critical stages. The HELIX project is still ongoing but will end in 2018.

The HEALS project has a focus on creating new analytical and computational methods to integrate Exposome-wide association studies (EWAS)(HEALS 2017). The aim of HEALS is to introduce new methodologies for data analysis and modeling tools to tackle large scale exposome studies. Similarly, to HELIX, HEALS will end in 2018.

In the US, significant funding has been provided by the National Institute of Environmental Health Sciences (NIEHS) to create the HERCULES exposome Research Center (HERCULES 2017). The aim and project of the Center have been described in detail in chapter 5.1 of this book. In brief, HERCULES provides key infrastructure and funding for exposome-related projects. The creation of the center has already lead to the birth of the Children's Health Exposure Analysis Resource (CHEAR) project (NIEHS 2017), which, similarly to HELIX, focusses on the exposome of critical early-stages of life.

In Japan, a substantial effort has been put in the creation of the Japan Environment and Children’s study (Kawamoto et al. 2014), involving 100,000 parent-child pairs followed from birth to 13 years of age. The study will periodically measure exposure to chemicals by direct analysis and questionnaires as well as lifestyle and health outcomes.

Other efforts across the globe include the creation of the International Exposome Center, the I^3 Care Centre, which is a global collaboration between the Chinese University of Hong Kong, the University of Utrecht, and the University of Toronto. The Centre aims at “expanding the exposome concept through the development of new investigative tools to discover and quantify modifiable risk factors”(Center 2017) and provides a framework for the development of new international projects on the understanding of the exposome.

Table 1. Project and research centers devoted to the study and understanding of the EXPOSOME across the world.
<table>
<thead>
<tr>
<th>Project</th>
<th>Objectives of the project</th>
<th>Funder</th>
<th>Start date</th>
<th>End date</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPOsOM-ICs</td>
<td>Develop new approaches to assess environmental exposures and to link them to biochemical and molecular changes in the body, with focus on Air pollution and water contaminants.</td>
<td>EU FP-7</td>
<td>2012</td>
<td>2017</td>
</tr>
<tr>
<td>HELIX</td>
<td>Combining all environmental hazards that mothers and children are exposed to and link it to health, growth and development of children</td>
<td>EU FP-7</td>
<td>2013</td>
<td>2018</td>
</tr>
<tr>
<td>HEALS</td>
<td>Functional integration of omics derived data and biochemical biomonitoring to create the internal exposome at the individual level</td>
<td>EU FP-7</td>
<td>2013</td>
<td>2018</td>
</tr>
<tr>
<td>Hercules</td>
<td>Provide key infrastructure and expertise to develop and refine new tools and technologies for the study of the exposome.</td>
<td>NIEHS</td>
<td>2013</td>
<td>2022</td>
</tr>
<tr>
<td>CHEAR</td>
<td>Advance understanding about how the environment impacts children’s health and development</td>
<td>NIEHS</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>JECS</td>
<td>Identify harmful factors in the environment affecting children's growth and health, and to investigate the relationship between such factors and children's health condition.</td>
<td>Japan</td>
<td>2011</td>
<td>2032</td>
</tr>
<tr>
<td>I3 Care</td>
<td>International Exposome Center</td>
<td>Netherlands-Hong-Kong-Canada</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Although these projects offered great advances in the description and understanding of the exposome, we are still far from understanding the full mechanisms and implications of exposome science. Further projects and investments will be needed to pursue advances in this domain.

**Are Omics tools suitable for the evaluation of the internal exposome?**

Characterization and measurement of the exposome is multi-disciplinary and relies on multiple analytical tools and technologies. The implementation of omics (big-data) technologies has yielded significant advances for the study of the internal exposome. As discussed in this book, omics provide useful tools to assess effects of exposure at a certain time-points. Metabonomics, Epigenomics and Transcriptomics have been extensively described in Section 3 of this book. To these, we must add other very important omics, such as Genomics, Proteomics, Adductomics (Rappaport et al. 2012), and the Microbiome – which should also be considered when assessing the exposome. These technologies allow the acquisition of many measurements, and they are often based on untargeted approaches, allowing an agnostic analysis of markers. When associated to exposures and outcomes, omics data can help identify biomarkers of exposure.

Overall, omics technologies have several advantages that make them very suitable to assess the exposome. Omics measurements can be applied on freshly obtained human samples, but also on samples from existing biorepositories within large cohort studies. Furthermore, as mentioned by Vineis *et al.*, it is possible to “investigate platform-specific markers playing a role in the biological pathway linking exposure to disease risk” (Vineis *et al.* 2013), each marker can give valuable information on mechanism and pathways involved in disease etiology and exposure. Finally, measurements on different omics platforms can be applied on the same samples, allowing the potential of cross-omics studies. If a signal belonging to a same biologically relevant pathway is found in multiple omics, the causal link from exposure to health outcome is reinforced.
Although very useful for the characterization of the exposome, omics measurements also present several limitations. The first is that omics markers can vary through time. Therefore, when they are measured on only one sample, bias can be introduced. Perrier et al. reported that using multiple samples pooled together from one specimen can reduce the attenuation bias without increasing analytical costs (Perrier et al. 2016). Undeniably, cost is another limitation for omics measurements, although over time these are being reduced, it has been estimated that the cost of a full suite of omics measurement can go from €50,000 to €100,000 for a study with 500 subjects (Siroux et al. 2016). The development of new cost-effective technologies in the future will be necessary to perform larger studies.

Other limitations include the amount of data produced by omics measurements. The management of these large data sets is a great challenge. Firstly, it requires a significant computational investment both for data processing and data storage. Secondly, as illustrated in section 3 of this book, the amount of data produced requires new statistical methodologies to interpret the significant signals. Finally, another important limitation of omics technologies relies in biological interpretation. Thousands of measured signals cannot always be interpreted and annotated due to technical limitations, making it sometimes impossible to form conclusions based on the findings. This limitation can be reduced by the creation of databases of signals and the sharing of data among researchers which can contribute to the identification of unknowns.

Apart from these limitations, omics are still considered the best option for the characterization of the internal exposome. Moreover, the continuous development of new technologies increases the capabilities of evaluating internal signals.
What advances have been made for the external exposome?

As reviewed in chapter 1.2 and section 3 of this book, the assessment of the external exposome can rely on many different tools and technologies. Environmental exposure and health data can now be obtained by using models based on geographical information systems. Data obtained by geo-localization technologies can be used to build models for exposure assessment. These models can be applied to large populations, and further be refined with direct measurements obtained by personal sensing on smaller sub-sets of individuals. A great example of this technique has been applied as part of EXPOsOMICs. GIS (geographical information systems) based models have been successfully applied to air pollution estimates in multiple cities in Europe (Gulliver et al. 2017) and models have been refined with data collected by personal monitoring devices. Other methods such as remote sensing, smartphone based sensors and apps, questionnaires, photos as well as personal dosimeters and sensors are currently being used in multiple projects, as reviewed in chapter 3.2 of this book. Additionally, new easy-to-use and low costs sensors are being developed, such as silicone wristbands. These have been applied as passive sampling devices for the evaluation of exposure of 30 individuals where 49 compounds were assessed with a simple and easy to use method (O’Connell et al. 2014). Smartphone and sensor-based technologies based on wireless devices present a great advantage, in fact, the information stored by these devices can be made available through internet portals through collaborations with app developers. The collected data can provide information on many subjects without the need of setting up a specific study. Additionally, continuously collected data can significantly improve the estimation of long-term and chronic exposures by reducing the gaps in data collection. These technologies are well suited to evaluate exposures related to location such as air pollution, noise, green-spaces etc., or what is more widely defined as the “urban exposome”.

One of the greatest limitation of these technologies, similarly to omics, is the amount of data produced. Great computing power is required to store, manage and to perform statistical analysis on these large datasets. Other issues include privacy and ethical issues, particularly related to data collected through apps and stored on the internet. Researchers must ensure that consent of participants is obtained and that proper and secure data management and storage insuring privacy is maintained.

The analysis of the external exposome also relies on direct measurements. As explained in section 3 of this book, high throughput technologies based on mass-spectrometry allow the analysis of thousands of compounds in food, air, water, dust, etc. to evaluate exposure to pollutants. As mentioned in chapter 3.1, limitations of these technologies rely today in the fact that
they are often semi-quantitative and can only be applied to a small number of samples. Further development in technologies for analysis and sample preparation are required to apply these technologies to a large number of samples with greater precision.

**Unravelling the exposome: where do we go from here?**

Over the past decade, research studies and government funded projects have offered guidance for the future needs and goals of the exposome. These precursor projects provide great advances in the field and will serve as examples and guides for future research. In this section we will highlight a few points that need to be considered in future exposome research.

**Collaborations and data sharing**

In 2012, Wild offered suggestions to transfer the exposome from concept to utility (Wild 2012). He pointed out that one of the main difficulties for the exposome would be the attempt from single groups of researchers to tackle the concept in its whole, he recommended that greater success would be achieved through collaboration, focusing on “defined goals and shared expertise” and sharing the information in the public domain. This approach has been proven successful by multiple collaborative projects such as EXPOsOMICs and HELIX (Vineis et al.; Vrijheid et al. 2014).

The concept of the exposome is broad, and therefore requires the collaboration across multiple disciplines: epidemiology, biology, statistics, computer science, exposure science etc. In 2014, the National Academy of Sciences, Engineering and Medicine held a workshop to establish which were the obstacles of data sharing in environmental research (Principles and Obstacles for Sharing Data from Environmental Health Research: Workshop Summary 2016). As part of this discussion, it was noted that the quality of the data is key to the sharing. Quality controls and common scientific language must be applied to make sharing and collaborations successful. Epidemiologist and statistician need to collaborate with experts in mass spectrometry, GIS and more, to ensure that proper study designs are applied, and that data is produced with proper methodologies and analyzed correctly. Similar conclusions were discussed in 2015 at the NIEHS Exposome workshop, during which several recommendations were made for future research in exposome. One of the recommendation was based on the need to develop databases and coordinating centers for “collaborative data storage, access and analysis”. Recommendations also included the establishment of language and methodological standards for sample collection, and use of technologies (Dennis et al. 2016). These principles have already been implemented by a recent NIEHS initiative as part of the Children’s
Health Exposure Analysis Resource (CHEAR) (NIEHS 2017). The CHEAR project will serve as a structure to facilitate the evaluation of early life-exposome. Methods and data collection will be harmonized among platforms to ensure cohesion among selected cohorts.

**Development of new tools and methods to measure biological responses**

Another difficulty to tackle in future exposome research is how to consider all aspects of a life course. Up to now, most projects have focused on snap-shots – measuring the exposome at specific time-points throughout the life-course and associating these with disease. Complete exposome studies need to account for longitudinal, multiple exposures and responses. In this context, the elucidation of important pathways is not simple and will require the development of new methodologies and models. In their review, Stingone et al. suggested a two-step approach to facilitate the interpretation of the exposome’s complexity (Stingone et al. 2017). The first step is the identification of critical windows of exposure, which will be studied in detail. The second step relies on the creation of solid and repeatable study designs which can facilitate the comparisons across different exposome measurements over time. Stingone et al. recommend the study based on merging different cohorts and databases that have focused on short windows. For this purpose, new methods that integrate individual studies at different specific life-stages, the relationship between multiple exposures, different omics measurements, and health outcomes must be developed. This approach is currently being tested in the EXPOsOMICs project. As illustrated in Chapters 4.1 and 4.2 of this book, new approaches to address the high-dimensional data produced in exposome studies that can link data to pathway analysis and network construction can contribute to converge data from different platforms and guide biological interpretation.

**Other applications, the Exposome, and precision medicine**

In recent decades, epidemiology has shifted its attention to chronic diseases in developed countries. This shift, as well as the failure of genomics to explain the causative factors of many chronic diseases, has led to a reconsideration of the health impact of environmental exposures. The study of the exposome provides growing evidence that exposures to toxicants (air, water, soil, food, household products, etc.) contributes to many chronic diseases. In fact, many diseases have been linked to environmental exposure such as diabetes (Chevalier and Fenichel 2015), Alzheimer’s disease (Genuis and Kelln 2015), reduced fertility (Buck Louis et al. 2013), cardiovascular diseases (Xu et al. 2009), cancer (Cao et al. 2011; Kim et al. 2013; Teitelbaum et al. 2015) and more (Bijlsma and Cohen 2016).
It was estimated that over 4.9 million deaths could be attributed to exposure to environmental chemicals in 2011 (WHO 2015). This figure stresses the importance of considering the effect of exposures on health outcomes. However, little has been done to bring knowledge of the possibilities of including environmental health in prevention and diagnosis to clinicians. Advances in exposome technologies, such as the development of OMICs (including genomics), and personal sensors can provide tools for clinicians to evaluate the role of environmental exposures in patient’s health. Classical views of disease onset and progression need to be replaced and considered holistically as caused by a complex interplay of genetic and environmental factors (GxE as referred in Chapter 1.1). In current medicine, diagnosis is based on clinical practice guidelines which consider the patients not as an individual but as a group (Ziegelstein 2017). Precision medicine is defined by considering unique biological characteristics of each individual and tailor the diagnosis and therapeutic specifically to the individual’s phenotype. In this context, exposome tools could be used to integrate and interpret data about disease causation and prevention. Genomics, metabolomics, proteomics, epigenomics and other omics can deliver information that can lead to more precise diagnosis (Collins and Varmus 2015). Epigenetic for instance, can provide evidence of inherited effects and alterations due to environmental exposures which can be implicated in disease causation. The inclusion of regular comprehensive monitoring of omic levels as well as the use of sensor-based technologies could contribute, in the future, to assess differential risk in patients. These recent developments could soon offer a bridge between environmental science and medical research and indicate a new pathway for personalized medicine engaging environmental scientists, epidemiologists, clinicians and individuals that would help the understanding of the links between the environment and human health.
References


