Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2012;41:24–32
doi:10.1093/ije/dyr236


REVIEW

# The exposome: from concept to utility

**Christopher Paul Wild**

International Agency for Research on Cancer, 150 cours Albert Thomas, 69008 Lyon, France. E-mail: director@iarc.fr

**Accepted** 22 December 2011

**Keywords** Exposome, exposure assessment, biomarkers, cancer epidemiology



## Introduction

The concept of the exposome was developed to draw attention to the critical need for more complete environmental exposure assessment in epidemiological studies;[1–4] environment is defined in this context in the broad sense of 'non-genetic'. The exposome, therefore, complements the genome by providing a comprehensive description of lifelong exposure history. Remaining focused on the element of application (to epidemiology) is a key to ensuring that the exposome is translated from concept to utility for better delineating the causes and prevention of human disease.

At the time of the original proposal, it was recognized that whereas exquisite tools had been developed to sequence the human genome and to interrogate individual susceptibility through genome-wide association studies (GWAS), there was a relative paucity of comparable tools, or indeed investment, in relation to exposure assessment. Given that cancer and other chronic diseases develop predominantly from a combination of environmental exposures played out on a particular genetic background, the inability to measure one part of the gene:environment combination with anything approaching the precision of the other will stymie progress. This becomes particularly acute as epidemiology aims to tease out relatively modest effect sizes associated with specific environmental exposures.

This commentary seeks to further define the exposome and to describe how its realisation may be achieved in epidemiological studies. The commentary focuses on cancer but many of the concepts are applicable to other chronic diseases.

## Defining the exposome

### Scope

The exposome is composed of every exposure to which an individual is subjected from conception to death. Therefore, it requires consideration of both the nature of those exposures and their changes over time. For ease of description, three broad categories of non-genetic exposures may be considered: internal, specific external and general external (Figure 1).

First, the exposome comprises processes internal to the body such as metabolism, endogenous circulating hormones, body morphology, physical activity, gut microflora, inflammation, lipid peroxidation, oxidative stress and ageing. These internal conditions will all impinge on the cellular environment and have been variously described as host or endogenous factors. Secondly, there is the extensive range of specific external exposures which include radiation, infectious agents, chemical contaminants and environmental pollutants, diet, lifestyle factors (e.g. tobacco, alcohol), occupation and medical interventions. In the past, these have been the main focus of epidemiological studies seeking to link environmental risk factors with cancer. Thirdly, the exposome includes the wider social, economic and psychological influences on the individual, for example: social capital, education, financial status, psychological and mental stress, urban–rural environment and climate. Here are encompassed the social determinants of health and the 'causes of the causes'.[5,6] It is also important to highlight the particular environment of the child in the earliest stages of life, namely the body of its mother, and in a small proportion of individuals, the additional early life exposure to *in vitro* cell culture. There is overlap in the three domains described above and sometimes difficulty in placing a particular exposure in one domain or another; for example, one can debate whether physical activity should be in the internal or specific external domains. Furthermore, the domains not only overlap but also may be considered as intertwined, in that the internal may at least partially be a response to the external. Measures in one domain or another may reflect to differing degrees one component of the exposome, e.g. the urban environment (general external), air pollution (specific external) and inflammation

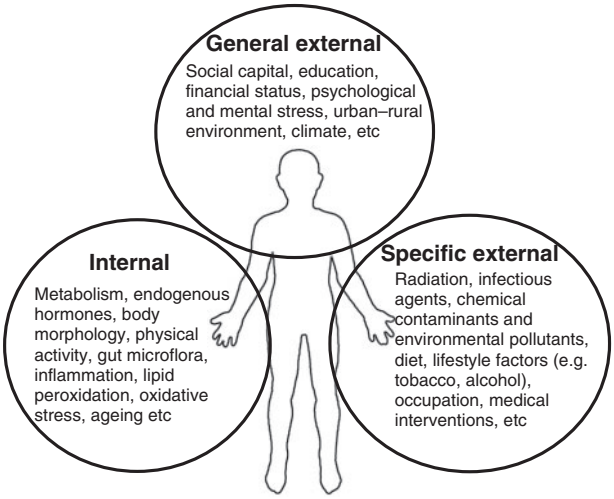


**Figure 1** Three different domains of the exposome are presented diagrammatically with non-exhaustive examples for each of these domains

(internal). Finally, the role of largely chance, stochastic events, distinct from the above domains, will need to be considered when interpreting measures of the exposome.[7] Despite these nuances and uncertainties, the current description serves the general purpose of bringing to mind the full breadth of the exposome.

If the broad domains described above are combined, then a picture of a comprehensive exposome begins to emerge, integrating the internal environment of the body, the specific external agents to which one is exposed and the social, cultural and ecological contexts in which the person lives their life. Notwithstanding the descriptions which are formulated, it is important to strive to distinguish between the inherent composition of the exposome, its biological consequences (i.e. the effects of the exposures of interest) and the more practical question of the methodology applied to measure the exposome. I will return to these two latter points below.

### Dynamic

The dynamic nature of the exposome presents one of the most challenging features of its characterization. As a consequence, its myriad components need to be considered in relation to their temporal variation. In effect, at any given point in time, an individual will have a particular profile of exposures. Therefore, to fully characterize an individual's exposome would require either sequential measures that spanned a lifetime, or a smaller number of measures that captured exposure over a series of extended periods. Whether this is necessary or desirable will be considered below, but for now let us accept this is what, in principle, the exposome comprises.

One can imagine, therefore, innumerable cross-sectional measures of the exposure profile building to a continuous real-time monitoring, which cumulatively would represent the exposome of the individual (Figure 2). Within this continuum, there will be periods of only gradual change in the exposure profile; periods with dramatic changes in specific components, e.g. a change in occupation, or a course of medical treatment; and periods when there is a radical change in global exposure profile in a short period, for example, when one is born or migrates. Capturing the completeness of this inherent variability will be demanding, but an approximation to the complete exposome can now be envisaged. To consider how we must turn to the practical question of measurement.

## Measuring the exposome

Measuring the exposome must be disentangled from the discussion of its theoretical composition to avoid obscuring the underlying objective by the methodology applied. The potential for this conflation has at least partly arisen because advances in laboratory sciences, permitting the simultaneous analysis of thousands of individual entities such as metabolites, proteins, etc., have developed in parallel with the proposal for measuring the exposome. This laboratory technology, captured under the generic label of 'omics', will surely yield important elements of the exposome, but not its entirety.

In fact, a more nuanced approach is required, drawing on a range of diverse tools in order to capture the full range of the three domains of the exposome described above. A 'one-agent-at-a-time' approach, using different methods to measure each individual exposure, is clearly unrealistic. A common feature therefore of the approaches to define the exposome is the integration of a wide range of individual exposures in a single measurement. This does not imply that single highly specific measures of exposure to individual external agents have no value. On the contrary, such tools can transform a research area, as has been demonstrated with chronic infections, such as hepatitis B virus, or chemical carcinogens, such as aflatoxins.[8] However, these targeted approaches will not go far in characterizing the whole exposome, partly a reflection of the complex range of exposures already identified but also the many exposures occurring which remain uncharacterized. The assimilation of the known and unknown is a primary goal of the new generation of methods to assess the exposome. I will describe below some of the tools that may help translate the exposome from concept to utility, beginning with biomarkers before going on briefly to highlight other promising avenues (Table 1).

### General 'omics'

It is now possible to sequence the entire three billion DNA bases of the germline genome of a specific individual or alternatively, of the cancer genome to

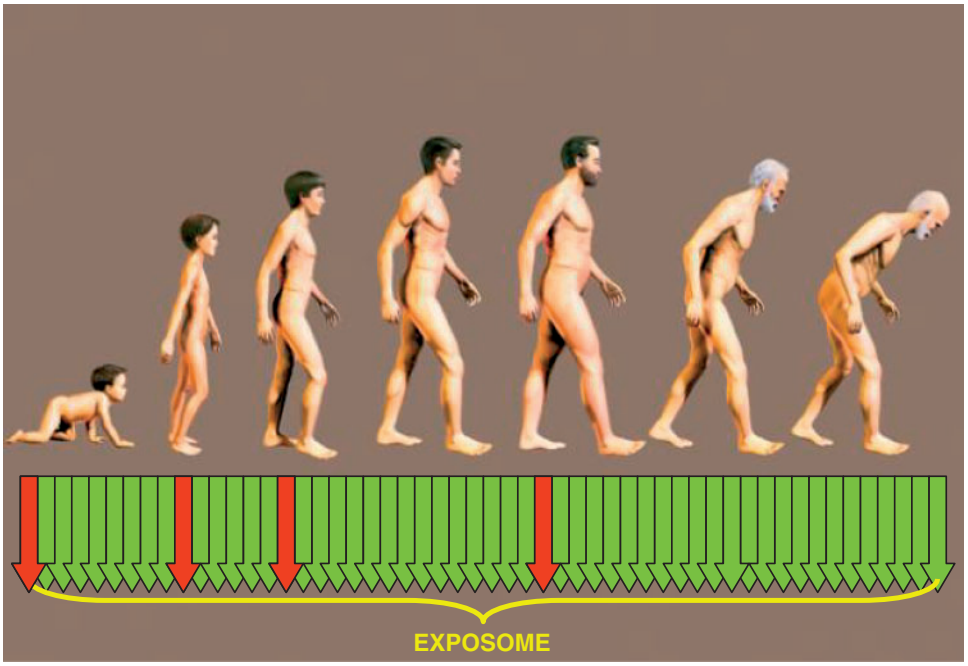


**Figure 2** The exposome would require measurement of exposures over time across the lifecourse of an individual (*in utero* exposures are included but not represented on this schema). The darker arrows indicate possible time-points where representative cross-sectional exposure assessments could be made in order to capture different key periods: *in utero*/infancy; childhood; adolescence; and adulthood

**Table 1** Some examples of approaches and tools to measure the exposome

| Approach | Tools |
|---|---|
| Biomarkers (omics) | |
| General | Genomics, transcriptomics, proteomics, metabolomics, epigenomics |
| Targeted | Adductomics, lipidomics, immunomics |
| Sensor technologies (including mobile phones) | Environmental pollutants, physical activity, stress, circadian rhythms, location [global positioning systems (GPS)] |
| Imaging (including mobile phones, video cameras) | Diet, environment, social interactions |
| Portable computerized devices (including palmtop computers) | Behaviour and experiences (ecological momentary assessment), stress, diet, physical activity |
| Improved conventional measurements (combined with environmental measures) | Job-exposure matrices; dietary recall (e.g. EPIC-Soft) |

capture the somatic mutations occurring during the pathogenic process. The first complete germline genome sequence of a human being spawned a book and cost millions of dollars;[9,10] >5 years later, it is unremarkable and increasingly affordable.[11] The latter characterization provides one essential platform for the study of the genetic and environmental aetiology of the disease.

Human genome sequencing has been closely followed by analogous tools necessary to characterize downstream biological events in the form of profiles of RNA expression (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in cells, tissues or body fluids. Epigenomics is providing complementary information, for example, by characterizing the methylome and microRNA profiles.[12] The scope and resolution of each of these omics approaches is being refined at a rapid pace and is permitting fascinating and valuable progress in understanding the mechanisms underlying cancer development.

Each of the omics methodologies provides information on thousands of individual component parts, which in principle can be correlated with disease endpoints, in analogous fashion to GWAS. The characterization of a pathological condition by its transcriptome, proteome, metabolome or epigenome promises a foundation for improved diagnosis and treatment of disease, including more accurate patient stratification. Huge financial investment is being made in utilizing this combination of knowledge and technology in the clinic and that will take its course. The questions we must ask here are: does this knowledge of underlying disease mechanisms

and technological advance offer anything to epidemiology and public health? More specifically, what do the omics technologies offer to characterization of the exposome?

I think the answer is: much, but not everything. The crucial point, and one where data are currently limited, is whether qualitative and quantitative relationships exist between exposures originating from the three domains described above and the profiles revealed by the omics technologies. The hypothesis is that different components of the exposome will leave their mark or fingerprint, so that one may travel not only forward from the molecular characteristics to the clinic but also back to the exposures, epidemiology and public health.[3]

Pilot studies of the relationship between exposure and molecular profiles are promising, with a number of different exposures, at least in relation to the 'specific external' domain of the exposome, having been related to differences in gene, protein and metabolite patterns in peripheral blood cells (see Wild[2] for a summary). For example, occupational exposure to relatively low levels of benzene (<1 ppm) in workers in China has been demonstrated to result in changes detected by transcriptomics, proteomics and epigenomics.[13] Smoking status was related to the pattern of microRNA expression in normal bronchial epithelial cells.[14] Similarly, arsenic exposure results in a defined set of epigenetic changes.[15] In addition, it is fascinating to see the whole-genome somatic mutation spectrum in human cancers in relation to specific environmental exposures including tobacco smoke and UV light.[16,17] These early results suggest that it may be possible to detect exposure through an indirect mechanism by measuring the impact of an agent on, for example, a panel of a few tens of different microRNAs or metabolites. Of course, it must be borne in mind that these relationships will be modulated by the genetic background of the individual concerned (see also 'What does the exposome offer other than improved exposure assessment?' section below).

The application of omics technologies to biological samples in epidemiological studies has been described as a 'top–down' approach to measuring the exposome.[4] This agnostic approach is consistent with GWAS studies where the identity and functional significance of the emerging associations between biomarker and disease are investigated post hoc. However, the 'top-down' approach lacks a crucial link between exposure to a given agent and the molecular profiles observed. To identify risk factors and implement prevention strategies such a link is a requirement.

In the case of GWAS, the DNA sequence itself is the exposure of interest, i.e. it is the variant at a specific genetic locus that is hypothesized to confer an altered risk of disease. Loci significantly correlated with the disease can be subsequently compared with existing databases to see if they fall within genes or other sequences of known function; if so something can be inferred about potential mechanism of action. In the case of other omics methodologies, some specific proteins or metabolites may be identified through available databases or through additional structural or chemical characterization, but the annotation of the individual components is far less comprehensive than for the genome. In addition, once the nature of the biomarker has been established, it is necessary to establish a link to the exposure, which led to an alteration in the level of that biomarker, recognizing the potential modulation of that relationship by both genetic and other environmental factors. Therefore, by entering through the portal of agnosticism, the link between a biomarker and an exposure will frequently require two additional steps: chemical or structural identification and establishment of a relationship with an exposure.

There will be exceptions to this assertion; for example, in the case of metabolomics, some of the endogenous or exogenous metabolites themselves are known and directly represent the internal or specific external exposures of interest. Examples include glucuronide metabolites of tobacco-specific nitrosamines or aflatoxin metabolites. But, in general, an additional effort is needed to link exposure to the omics data. This may be addressed in studies designed specifically for the purpose, for example, where short-term interventions are performed to modulate the exposure and observe resultant changes in the omics pattern.[18] Alternatively, if exposure has already been well-characterized in a group of subjects, for example, by a detailed dietary assessment or an occupational history, then the 'omics' pattern can be examined in subjects stratified for the exposure of interest.[13] Subsets of biomarkers validated by this type of studies could subsequently be used to measure specified exposures in epidemiological research.

## Targeted omics

The omics methods described above are predominantly of the global type, capturing thousands of individual component entities. There is however another type of assay that still includes multiple endpoints but is targeted at specific subsets of related molecules. Each of these targeted approaches can add further valuable components to the overall characterization of the exposome. For example, Rappaport and colleagues[19] recently summarized approaches towards the simultaneous analysis of a broad range of different chemical DNA and protein adducts, termed 'adductomics'. A number of research groups have used mass spectrometry techniques to examine the chemical binding of electrophiles to DNA,[20] serum albumin[21] and glutathione[22] among other targets. Other approaches have involved a focus on molecules grouped not by the exposure but by biological process; examples include lipid metabolism (lipidomics)[23] and

the immune system (immunomics).[24] In a first study of its kind, 266 different known environmental agents were measured in relation to type 2 diabetes in what was termed an environmental-wide association study.[25] In this study, novel associations were reported between environmental exposures and disease, with effect sizes similar to the strongest seen in GWAS. However, the greater challenges of excluding confounding and bias in environmental-wide association studies compared with GWAS is a critical consideration in interpretation of such approaches.[26]

## Other methods applicable to the exposome

Measuring the exposome will require a multi-faceted approach including further refinement of conventional tools and methods, innovative technologies as well as the biomarker approach discussed above. Each of these can make an important contribution in relation to the different domains of the exposome under consideration and each has its limitations. For example, for a number of environmental and occupational exposures, significant advances have been made by refining conventional measurements through a more detailed consideration of the exposure scenario. Examples include improved job-exposure matrices in occupational settings incorporating workplace measurements;[27] dosimetry for non-ionizing radiation[28] and geographic information systems (GISs)[29] to estimate domestic radon exposure.

In relation to new technology, as summarized in Table 1, the biomarkers based on either general or targeted omics technologies are being developed in parallel with other novel measurement tools. A significant investment in new technologies has come through the National Institutes of Health, USA: Gene, Environment and Health Initiative (GEI) (http://www.gei.nih.gov/exposurebiology/index.asp). A number of the initiatives and related publications are described on the GEI website. Approaches under development include personal monitors that capture a wide range of environmental pollutants; monitors for physical activity or physiological measures such as heart rate and blood pressure; devices that track location through GPS technology. Similarly, indicators of psychosocial stress may be captured in real-time through electronic diaries, whereas the use of mobile phones or video technology may allow measures of social interaction. Other initiatives include increasingly sophisticated dietary assessment through portable devices (palmtop computers, mobile phones), biosensors or web-based technology.[30,31]

These technologies are not without challenges. Ensuring the devices have sufficient robustness and acceptable dimensions for ease of use is important. The cost of equipment and therefore its application to population-based studies may be high, at least initially, and there will be requirements for training subjects in the use of the devices. When appropriate technology is available, there are study design concerns including, for example, the risk of selection bias due to computer literacy or internet access.

There are additional challenges aside from the measurement devices themselves. Not least is the fact that in many epidemiological cohorts, there is an under-representation of individuals from lower socio-economic strata; it would be important to ensure that the inclusion of new technologies helped reverse rather than exacerbate these problems of recruitment. Indeed, the need for close collaboration between social scientists, epidemiologists and laboratory scientists is essential if technological and methodological advances are to help in unravelling the complex relationships between social interactions, biological effects and disease risks.

Each of the above approaches assesses a component of the exposome, but in greater detail or with greater precision than previous methodologies. At the same time, it must be noted that the majority of these initiatives are in the early stages of development, face a number of challenges as alluded to briefly above and will require extensive validation prior to application in epidemiological studies.

## Timing of measurement

The above-mentioned methods address the diversity of the exposome but, as noted earlier, the exposome is in constant flux. Obtaining the totality of the exposome would therefore require either methods that integrate those fluctuations over time or a series of snapshot measures at specific times in the lifespan of an individual, building to an approximation of the full exposome (illustrated in Figure 2).

In relation to the former, it is too early to say how far the biomarker and other methods may provide this sought-after integration, but one might predict that capturing long-term past exposures in a comprehensive fashion is unlikely. In relation to the snapshot approach, Rappaport and Smith[4] proposed a number of key stages of life where cross-sectional measures of the exposome could be made, including gestation, early childhood, puberty and the reproductive years. These represent rational selections likely to exhibit significant differences in exposure patterns for a given individual. This strategy of partial characterization may allow significant advances in aetiological studies, even while falling short of a full description of the exposome. Consideration of the demands on the subject from a practical and ethical standpoint will be a key factor if repeat collections of bio-specimens and data are to be made.

The question of timing of exposure measurement also raises a limitation in that the large, prospective cohort studies of chronic diseases (e.g. UK Biobank) typically comprise middle-aged adults. Thus exposure assessment will be focused on a relatively narrow time window, which may not be the only or even the most important in relation to disease risk. Notably, there is increasing evidence that early life

and *in utero* exposures or those during adolescence, including factors as diverse as birthweight, diet and maternal psychosocial stress, are important to chronic disease risk later in life.[32–34] Therefore, in developing the exposome, the implications of life-course exposure must be a central consideration and the value of existing birth cohorts not overlooked.[32,35] Pragmatically, the issue of timing of exposure assessment may have to be tailored to specific hypotheses by examining exposure in cohorts of different ages.

## Is there a need for an individual to know their exposome?

The view has been expressed that an individual's genetic make-up, as determinable by whole-germline genome sequencing, will provide one significant entry point to personalized medicine.[36] In principle, the exposome could be thought of as an accompaniment to the genome to inform personal decisions about lifestyle or medical interventions. However, in terms of public health impact, I do not believe this should be the goal of exposome-orientated research. This is particularly so in the international context, where in many low- and middle-income countries, access to an individualized genome or exposome and the possibility to effect subsequent behavioural or medical interventions will remain constrained for the foreseeable future.

The primary purpose of the exposome should be to identify risk factors in epidemiological studies. The aim is to generalize observations from a group of individuals to the population as a basis for public health decisions. In this context, the full exposome of an individual may not need to be established. A more likely scenario would be a partial exposome established by one or more cross-sectional exposure measurements at different time-points, typically in a prospective cohort study. From this information, an association between an exposure and a risk may be identified. Even this partial exposome, where neither the totality of exposures nor the dynamic coverage is comprehensive, can still yield important aetiologic findings.

It may be that as the exposome is more completely measured, it yields information presented as useful at the personal level (one can all too easily envisage the development of direct-to-consumer products with their attendant risks). Some newly identified biomarkers, however, may have personal utility, as for example does high blood cholesterol, first identified as a biomarker of cardiovascular disease risk through epidemiological studies. Nevertheless, it is likely that a more appropriate analogy for the vast majority of the exposome information will be with low penetrance genes, where the information is useful at the population but not the individual level.[37]

As work on characterizing the exposome evolves, it is important to retain the notion that even a partial description can lead to major public health benefits. Certainly, in the first instance, a drive to deliver a personal exposome for individual health choices should not be allowed to deflect from this more public health-orientated goal.

## What does the exposome offer other than improved exposure assessment?

In the first phase of molecular cancer epidemiology, it was recognized that some categories of biomarkers, e.g. DNA adducts, are an integration of exposure and various biological processes that vary at the individual level e.g. carcinogen metabolism, DNA repair and cell turnover, and that these latter processes are at least partially a result of inter-individual genetic variation. This type of biomarker was therefore not going to strictly correlate with exposure, but might be associated with cancer risk because it reflected the consequences of that exposure on a pathway relevant to carcinogenesis. The application of omics technologies to the relationship between human exposure and disease provides the same challenges and opportunities, albeit on a far greater scale, in that the endpoints measured will represent a composite mix of direct responses to exposure and downstream biological effects.

The challenge therefore is to recognize, understand and interpret the interplay between the exposure and the biological responses to that exposure. The opportunity is that the information gathered can indicate not only the link between an exposure and a disease but also provide insights into the mechanisms by which an exposure might be exerting its effects. Such insights may contribute to the weight of evidence in assigning causality to an exposure–disease association and open up avenues to prevention through modulation of specific identified biological pathways. This link between exposure, mechanisms and disease has been described succinctly by Vineis and Perera[38] as a 'meet-in-the middle' approach to biomarker discovery. A recent application of the principle has been described, whereby metabolomics was used to identify biomarker associations both with cancer (colon and breast) and with exposure (dietary information).[39]

The opportunity to elucidate the link between exposure, mechanisms and disease is one of the most exciting for cancer research in the coming decade. The ability to interrogate specific biological pathways may provide completely novel insights into how seemingly diverse risk factors, such as obesity, physical inactivity, immune suppression, psychosocial stress and other behavioural traits play out on common pathways to yield a common phenotype.[40,41] In this

regard, suggestions that measures of psychosocial stress, even *in utero*, are correlated with oxidative stress, inflammation and telomere length[34,42] provide intriguing insights into how seemingly disparate risk factors may act through common biological pathways, revealed through an integrative systems biology approach. As mentioned above, however, it is particularly in these novel areas of investigation that a true inter-disciplinary collaboration between social scientists, epidemiologists and laboratory scientists is crucial.

## The exposome: from concept to utility

The current commentary attempts to provide further definition to the concept of the exposome and an indication of how the concept may be translated to utility. Technological advances in biomarkers, personal monitors, imaging etc., offer ways to construct the exposome with increasing completeness. Large-scale prospective cohorts, intervention studies and other designs provide opportunities for application to epidemiological investigations. In this context, it is important to highlight that the biomarker methodologies, by definition, require the availability of biological specimens. The quality of these samples (collection, processing and storage) is as vital as the methods applied to them.[43] In developing and validating biomarker approaches, feasible application to the nature and amount of biological samples in biobanks has to be a defining criterion.

The increasing number of prospective cohorts and their associated biobanks do indeed offer a rich resource. However, their value in unravelling the genetic and environmental contributions to chronic diseases is predicated on the availability of adequate measures of environmental exposure.[1] Hence, there is a marriage, not of convenience but of necessity, between the cohorts and the new tools to assess the exposome.

The need for synergism goes further in that addressing the exposome implies the need for inter-disciplinary research. The juxtaposition of molecular mechanisms, biotechnology, bioinformatics, biostatistics, epidemiology, social sciences and clinical research requires collaboration across disciplines that currently use different paradigms, tools and even language. Training researchers to have the linguistic skills to order more than a beer outside their discipline of origin is an important goal. A vital component of the above inter-disciplinary effort is the challenge of processing and analysing the large volumes of data generated by the methodologies for characterizing the exposome. The application of omics technologies, potentially in a longitudinal fashion, poses enormous challenges to both bioinformatics and biostatistics. This point is not developed further here, but has to be a part of any future assessment of the human exposome.

Hopes of identifying the causes of cancer and other chronic diseases in cohort studies depend not only on scientific interest, application and innovation; but they also rely on funding and cooperation. Three remarkable features of the human genome project are notable: technological advances (increased speed and reduced cost of DNA sequencing), huge financial investment and a commitment to international cooperation, whereby the responsibility for sequencing specific human chromosomes was assigned to different teams across the world. The exposome would progress most rapidly by an analogous approach. For the exposome, the technology is emerging, the investment is showing small signs of life (e.g. the above-mentioned NIH USA: GEI and the recent FP7 European Union call for funding of the European exposome initiative), but the international cooperation is largely absent. At first sight the 'chromosomes' of the exposome are not so easily identified and shared out for a coordinated effort. Neither is there the uniformity and factory-like standardization of a single DNA sequencing methodology.

Consequently, the exposome does not translate as easily to big science, to a 'moon-shot' solution, as did sequencing the human genome. However, there are recognizable divisions of emphasis, for example, on metabolism, inflammation, diet, obesity, physical inactivity, environmental pollutants, social determinants, etc., which could be assigned to specific international teams with defined goals and shared expertise. There could be coordinated international investment with national funders able to target their contribution to exposures of regional or national priority. There could be further efforts to coordinate the major national prospective cohort studies, such as has been achieved at a regional level, thus facilitating pooled statistical analyses (e.g. in Asia, see Rolland *et al.*[44]). Data generated on the exposome could be shared in a common, publicly accessible database. Each of these cooperative initiatives would provide key elements to an international effort to elucidate the human exposome.

The time is ripe to translate the exposome from concept to utility. An international coordinated effort from governments, international organizations, scientists, funders and the public would have much to offer, not only in the aspects mentioned above, but also in ensuring that exposures most relevant to low- and middle-income countries are not neglected; for it is in these regions that the burden of cancer and other non-communicable diseases is set to rise most dramatically.[45] A truly international response to improving environmental exposure assessment will be one that applies the most sophisticated science to encompass the challenges of rich and poor alike and presents a global response to the causality and prevention of disease.

## Acknowledgements

The author is grateful to his colleagues Silvia Franceschi, Rodolfo Saracci, Joachim Schüz and Paolo Vineis for their discussions and comments on a draft of this article. This work was supported by European Union grant No. 260791 entitled ''Eurocan-platform: A European Platform for Translational Cancer Research''.

**Conflicts of interest:** None declared

## References

[1] Wild CP. Complementing the genome with an 'exposome': the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;**14:**1847–50.
[2] Wild CP. Environmental exposure measurement in cancer epidemiology. *Mutagenesis* 2009;**24:**117–25.
[3] Wild CP. Future research perspectives on environment and health: the requirement for a more expansive concept of translational cancer research. *Environ Health* 2011; **10(Suppl 1):**S15.
[4] Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. *Science* 2010;**330:**460–61.
[5] Hiatt RA, Breen N. The social determinants of cancer: a challenge for transdisciplinary science. *Am J Prev Med* 2008;**35:**S141–50.
[6] Friel S, Marmot MG. Action on the social determinants of health and health inequities goes global. *Annu Rev Public Health* 2011;**32:**225–36.
[7] Davey Smith G. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int J Epidemiol* 2011;**40:** 537–62.
[8] Groopman JD, Kensler TW, Wild CP. Protective interventions to prevent aflatoxin-induced carcinogenesis in developing countries. *Annu Rev Public Health* 2008;**29:** 187–203.
[9] Levy S, Sutton G, Ng PC *et al*. The diploid genome sequence of an individual human. *PLoS Biol* 2007;**5:**e254.
[10] Venter JC. *A Life Decoded: My Genome: My Life*. New York: Viking Adult, 2007.
[11] Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;**458:**719–24.
[12] Kulis M, Esteller M. DNA methylation and cancer. *Adv Genet* 2010;**70:**27–56.
[13] Zhang L, McHale CM, Rothman N *et al*. Systems biology of human benzene exposure. *Chem Biol Interact* 2010;**184:** 86–93.
[14] Schembri F, Sridhar S, Perdomo C *et al*. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci U S A* 2009;**106:**2319–24.
[15] Smeester L, Rager JE, Bailey KA *et al*. Epigenetic changes in individuals with arsenicosis. *Chem Res Toxicol* 2011;**24:** 165–67.
[16] Lee W, Jiang Z, Liu J *et al*. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;**465:**473–77.
[17] Pleasance ED, Cheetham RK, Stephens PJ *et al*. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010;**463:**191–96.
[18] Walsh MC, Brennan L, Pujos-Guillot E *et al*. Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr* 2007;**86:**1687–93.
[19] Rappaport SM, Li H, Grigoryan H, Funk WE, Williams ER. Adductomics: characterizing exposures to reactive electrophiles. *Toxicol Lett* 2011; doi:10.1016/j.toxlet.2011.04.002 [Epub 8 April 2011].
[20] Kanaly RA, Matsui S, Hanaoka T, Matsuda T. Application of the adductome approach to assess intertissue DNA damage variations in human lung and esophagus. *Mutat Res* 2007;**625:**83–93.
[21] Rubino FM, Pitton M, Di FD, Colombi A. Toward an 'omic' physiopathology of reactive chemicals: thirty years of mass spectrometric study of the protein adducts with endogenous and xenobiotic compounds. *Mass Spectrom Rev* 2009;**28:**725–84.
[22] Wagner S, Scholz K, Sieber M, Kellert M, Voelkel W. Tools in metabonomics: an integrated validation approach for LC-MS metabolic profiling of mercapturic acids in human urine. *Anal Chem* 2007;**79:**2918–26.
[23] Bou Khalil M, Hou W, Zhou H *et al*. Lipidomics era: accomplishments and challenges. *Mass Spectrom Rev* 2010;**29:** 877–929.
[24] Yan Q. Immunoinformatics and systems biology methods for personalized medicine. *Methods Mol Biol* 2010;**662:** 203–20.
[25] Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 2010;**5:**e10746.
[26] Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007;**4:**e352.
[27] Teschke K, Olshan AF, Daniels JL *et al*. Occupational exposure assessment in case-control studies: opportunities for improvement. *Occup Environ Med* 2002;**59:**575–93.
[28] Schmiedel S, Bruggemeyer H, Philipp J, Wendler J, Merzenich H, Schuz J. An evaluation of exposure metrics in an epidemiologic study on radio and television broadcast transmitters and the risk of childhood leukemia. *Bioelectromagnetics* 2009;**30:**81–91.
[29] Andersen CE, Raaschou-Nielsen O, Andersen HP *et al*. Prediction of 222Rn in Danish dwellings using geology and house construction information from central databases. *Radiat Prot Dosimetry* 2007;**123:**83–94.
[30] Schatzkin A, Subar AF, Moore S *et al*. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev* 2009;**18:**1026–32.
[31] Illner AK, Harttig U, Tognon G *et al*. Feasibility of innovative dietary assessment in epidemiological studies using the approach of combining different assessment instruments. *Public Health Nutr* 2011;**14:**1055–63.
[32] Wild CP. How much of a contribution do exposures experienced between conception and adolescence make to the burden of cancer in adults? *Cancer Epidemiol Biomarkers Prev* 2011;**20:**580–81.
[33] Burdge GC, Lillycrop KA, Jackson AA. Nutrition in early life, and risk of cancer and metabolic disease: alternative endings in an epigenetic tale? *Br J Nutr* 2009;**101:** 619–30.
[34] Entringer S, Epel ES, Kumsta R *et al*. Stress exposure in intrauterine life is associated with shorter telomere length in young adulthood. *Proc Natl Acad Sci U S A* 2011;**108:** E513–18.

[35] Merlo DF, Wild CP, Kogevinas M, Kyrtopoulos S, Kleinjans J. NewGeneris: a European study on maternal diet during pregnancy and child health. *Cancer Epidemiol Biomarkers Prev* 2009;**18:**5–10.
[36] Offit K. Personalized medicine: new genomics, old lessons. *Hum Genet* 2011;**130:**3–14.
[37] Vineis P, Schulte P, McMichael AJ. Misconceptions about the use of genetic tests in populations. *Lancet* 2001;**357:** 709–12.
[38] Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomarkers Prev* 2007;**16:** 1954–65.
[39] Chadeau-Hyam M, Athersuch TJ, Keun HC *et al*. Meeting-in-the-middle using metabolic profiling – a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers* 2011;**16:**83–88.
[40] Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM. DNA methylation in white blood cells: Association with risk factors in epidemiologic studies. *Epigenetics* 2011;**6:**828–37.
[41] Thayer ZM, Kuzawa CW. Biological memories of past environments: Epigenetic pathways to health disparities. *Epigenetics* 2011;**6:**798–803.
[42] Wolkowitz OM, Mellon SH, Epel ES *et al*. Leukocyte telomere length in major depression: correlations with chronicity, inflammation and oxidative stress – preliminary findings. *PLoS One* 2011;**6:**e17837.
[43] Hewitt R, Hainaut P. Biobanking in a fast moving world: an international perspective. *J Natl Cancer Inst Monogr* 2011;**2011:**50–51.
[44] Rolland B, Smith BR, Potter JD. Coordinating Centers in Cancer Epidemiology Research: the Asia Cohort Consortium Coordinating Center. *Cancer Epidemiol Biomarkers Prev* 2011;**20:**2115–19.
[45] World Health Organization. Global status report on noncommunicable diseases 2010. Geneva, Switzerland: WHO Press, 2011.