Environmental and Molecular Mutagenesis 54:461–467 (2013)

Commentary

Advancing the Application of Omics-Based Biomarkers in Environmental Epidemiology

Paolo Vineis,[1,2]* Karin van Veldhoven,[1] Marc Chadeau-Hyam,[1] and Toby J. Athersuch[1,3]

[1]*MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom*

[2]*Molecular and genetic epidemiology Unit, HuGeF, Human Genetics Foundation, Torino, Italy*

[3]*Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, South Kensington, London, United Kingdom*

The use of omics represents a shift in approach for environmental epidemiology and exposure science. In this article, the aspects of the use of omics that will require further development in the near future are discussed, including (a) the underlying causal interpretation and models; (b) the "meet-in-the-middle" concept, with examples; (c) the role of "calibration" of measurements; and (d) the role of life-course epidemiology and the related development of adequate biostatistical models. Environ. Mol. Mutagen. 54:461-467, 2013. © 2013 Wiley Periodicals, Inc.

Key words: environmental epidemiology; exposome; systems biology; compartmental models; biomarkers

INTRODUCTION: OMICS AND SYSTEMS BIOLOGY

The identification of hazardous environmental pollutants is complex, particularly in relation to chronic, noncommunicable diseases. The main contributors to this complexity are the diversity of hazards that may exist, the typically low levels of environmental contaminants/pollutants, long latency periods, and largely unknown modes of action. The unravelling of environmental causes of disease is also limited by the technical difficulties in defining, and accurately measuring exposures, and by considerable spatial, temporal, and intraindividual variation. The complex and partially unknown interaction with underlying genetic and other factors that modulate susceptibility and response to environmental exposures further complicates the process of delineating and understanding environmental hazards. To address such difficulties, the concept of the "exposome" was proposed, initially by Wild [2005], with more recent detailed development in relation to its application to population-based studies [Wild, 2012]. The original concept was expanded by others, particularly Rappaport and Smith [2010] who functionalized the exposome in terms of chemicals detectable in biospecimens [Exposome]. The exposome concept refers to the totality of exposures from a variety of sources including, but not limited to, chemical agents, biological agents, radiation, and psychosocial components from conception onward, over a complete lifetime, and offers a conceptual leap in studying the role of the environment in human disease [Exposome; Rappaport and Smith, 2010; Wild, 2012].

The term "omics" refers to the quantitative measurement of global sets of molecules in biosamples using high-throughput techniques, in combination with advanced biostatistics and bioinformatics tools [Vineis et al., 2009]. In this article, we refer specifically to the application of omics in an epidemiological context and confine our use of the term "environment" to pollutants/environmental contaminants, as opposed to the totality of nongenetic factors encompassed by the exposome concept.

Grant sponsor: European Commission (ESCAPE, EnviroGenomarkers, Exposomics); Grant number: FP7–308610.

*Correspondence to: Paolo Vineis, MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, United Kingdom. E-mail: p.vineis@imperial.ac.uk

Received 6 November 2012; provisionally accepted 14 January 2013; and in final form 15 January 2013

DOI 10.1002/em.21764

Published online 21 March 2013 in Wiley Online Library (wileyonlinelibrary.com).

The exposome may be at least partially characterized by repeated measurements of global sets of biomarkers of exposure and early effects in biosamples collected at different critical life stages. These biomarkers include a wide range of molecules, including small molecule metabolites in blood (metabolomics), covalent complexes of chemical agents with DNA and proteins (adductomics), and downstream changes in gene expression levels and regulation (transcriptomics, epigenomics, and proteomics). Although encouraging progress has already been made in the use of omics technologies in environmental health research, considerable additional development is required, including the application of such methods to samples from existing biorepositories within large population studies.

Given the availability of data from high-throughput omics platforms, together with reliable measurements of external exposures, at least for some prioritized exposures, the use of omics may represent not only a leap forward in chronic disease epidemiology but also the investigation for platform-specific markers playing a role in the biological pathway linking exposure to disease risk. Additionally, the in-depth characterization of common or integrated patterns across platforms allows a full application of the concept of systems biology in environmental epidemiology.

This article addresses some aspects that are key to the understanding, development, and application of omics in environmental epidemiology: (a) the underlying causal interpretation and models; (b) the "meet-in-the-middle" (MITM) concept, with examples; (c) the role of "calibration" of measurements; and (d) the role of life-course epidemiology and the related development of adequate biostatistical models.

CAUSAL INTERPRETATIONS

It has been proposed that the causal interpretation typical of reductionism is bottom-up, whereas theories of top-down causation (or of whole-part causation) are more relevant to systems biology. Philosophers who have contributed to the latter have used concepts such as integration, pattern and function [Sellars, 1970], or organization [Van Gulick, 2007]. However, we acknowledge that the two approaches are complementary and that probably much (if not most) successful science has been bottom-up rather than top-down. Therefore, we are not advocating a substitution of bottom-up with top-down, but complementarity of the two. Moreover, we are aware that the terms "bottom-up" and "top-down" have been used in the exposome literature [particularly in Rappaport and Smith, 2010] with slightly different meanings: bottom-up is the omic-based investigation of specific exposures or classes of exposures, whereas top-down refers to complete agnostic investigation of omic signals irrespective of hypotheses on relevant exposures. The discussion on causal models fits into this conceptual context, as we will try to show below.

In brief, the bottom-up, causal reductionist view is that complex, higher-level systems are nothing but the sum of their parts and that the behavior of the system is determined by the intrinsic properties of its constituent parts. The behavior of these parts is independent of their relationship to other parts or the system as a whole. In contrast, the top-down view is that the intrinsic properties of the constituents of a system are not the sole determinants of its overall behavior; the relationship between the constituent parts, that is, the organization of the system, gives rise to higher-order laws that govern this behavior. Below, we consider more detailed aspects of the bottom-up and top-down approaches.

Bottom-Up

The traditional neo-Darwinian theory based on random mutations is an example of typical bottom-up interpretation of causality. The assumption is made that by identifying individual key genes one can discover the underlying causes of (higher order) biological phenomena. For example, the discovery of the gene encoding the tumor suppressor, p53, and of its role as a "guardian of the genome," led to the hypothesis that its mutations were crucial for carcinogenesis. However, it is now clear that mutations in the gene encoding p53 are neither necessary nor sufficient to cause cancer. By construction, the bottom-up approach is neither able to adequately consider the role of selection operated by conditions external to the cell nor does it address important aspects of biology that have relevance to carcinogenesis, such as the tension between stability, variation, and plasticity. Organisms (and cells) require stability to transfer their properties from one generation to the following, variability to respond to major environmental challenges, and plasticity to respond to minor ones. All these are forces operating at both the low and higher levels of causation that drive the link between causes and effects but cannot be accounted for based on bottom-up approaches.

Top-Down

The causal framework underlying top-down approaches relies on multiple, ordered, and correlated levels contributing to the observed phenomenon. Information driving the observed process cannot be entirely derived by lower levels; lower levels are typically necessary, but are rarely sufficient. The fact that they are not sufficient is expressed by the concept of emergence of new properties in a complex system (as in biological systems). It has been proposed [Deacon, 2007], in fact, to distinguish three levels of emergence in complex systems.

First-order emergence can be illustrated by the viscosity of liquids, turbulence in large bodies of water, and

feedback systems (such as thermostatically controlled heating systems). The physical processes linking elements of the first-order emergence impose constraints on the system and ultimately define its behavior.

Second-order emergence occurs when there is the amplification of a fluctuation beyond the constraints from physical processes. Systems in which this occurs are called nonlinear. The simpler sort is self-organizing, in which higher-order patterns selectively drive the incorporation of lower-order constituents into the system or select among possible states of the lower level entities. More complex second-order emergent systems are also "autopoietic," that is, they change the lower-order constituents themselves and their interactions. All of life involves second-order emergence of this more complex sort.

Third-order emergence involves further interaction between levels and appears only in biological systems. Here, a variety of second-order forms emerge and are selected by the environment, but in such a way that a representation of their form is introduced into the next generation. An example is the evolutionary process where the microlevel (the genome) in interaction with the higher-level organism's environment directs the construction of the organism (the mid-level), whose reproductive fate is determined top-down by the environment. The preservation of information regarding the organism's success in the environment is the means by which a population of successful organisms can be produced and maintained, within which future viable fluctuations appear. In this way, a two-stage process of emergence occurs that results in downward causation not just from top to mid-level but from top to bottom—from environment to genome.

To summarize, components in complex systems such as the human body are arranged in a hierarchy, with each level consisting of modules that obey local rules and interact with those in the adjacent "higher" and 'lower' levels of organization. Ellis [2012] has also reviewed the top-down causal mechanisms and discussed five distinct types (algorithmic top-down causation, top-down causation via nonadaptive information control, top-down causation via adaptive selection, top-down causation via adaptive information control, and intelligent top-down causation). Of these, top-down causation via adaptive selection relates to natural selection and most biological phenomena.

THE MEET-IN-THE-MIDDLE PRINCIPLE

In relation to environmental epidemiology, top-down causality can be investigated in a variety of ways. The simplest approach is to start by determining the associations between exposures, intermediate markers, and disease. Building on this, we have recently proposed a new approach that is known as the MITM [Chadeau-Hyam et al., 2011] and includes by construction a multilayer causal framework. Its implementation aims at addressing the challenge of identifying causal relationships that link exposures and disease outcome. This approach is usually based on a combination, within a population study, of a prospective search for intermediate biomarkers, which are elevated in subjects who eventually develop disease and a retrospective search for links of such biomarkers to past environmental exposures (Fig. 1). The reasoning is in three steps. The first step of this approach consists in the investigation of the association between exposure and disease. The next step consists in the study of the relationship between (biomarkers of) exposure and intermediate omics biomarkers of early effects; and third, the relation between the disease outcome and intermediate omics biomarkers is assessed. The MITM stipulates that the causal nature of an association is reinforced if it is found in all three steps.

To illustrate its flexibility, we describe below the practical use of the MITM paradigm in the context of a literature review process using real data.

Endocrine-Disrupting Chemicals and Omics Technologies: A Literature-Based Example

Endocrine-disrupting chemicals (EDCs) are compounds, of both synthetic and natural origin, in the environment that interfere with (i.e., mimic and/or antagonize) the actions of endogenous hormones by altering hormone synthesis, secretion, transport, binding, action, or elimination and thereby disrupt the functions of the endocrine system. Many studies have focused on the association between exposure to EDCs and breast cancer risk. There are numerous pathways to EDC exposure, and although a few of these compounds are administered for medicinal purposes, exposure to the vast majority of these EDCs is unintended. The epidemiological evidence to support or refute a relationship between EDCs and breast cancer is still limited, and therefore, it is a potentially fruitful field of application of the MITM concept. Here, we show how the MITM concept can be applied in a systematic review of the literature.

We searched PubMed using MESH terms for literature published up to May 2011. The search was performed in four stages: for (a) gene expression profiling, (b) proteomics, (c) epigenomics, and (d) metabolomics. More details are given in a separate paper (van Veldhoven et al., in preparation). Three candidate genes (i) have been reported in more than one study as differentially expressed after exposure to EDC and (ii) have also been found to be differentially expressed in breast cancer tissue and patients (Table I).

One of the main challenges when considering (potential) intermediate biomarkers is to understand whether they belong to the causal pathway linking exposure and disease risk, whether they are simply a side effect of

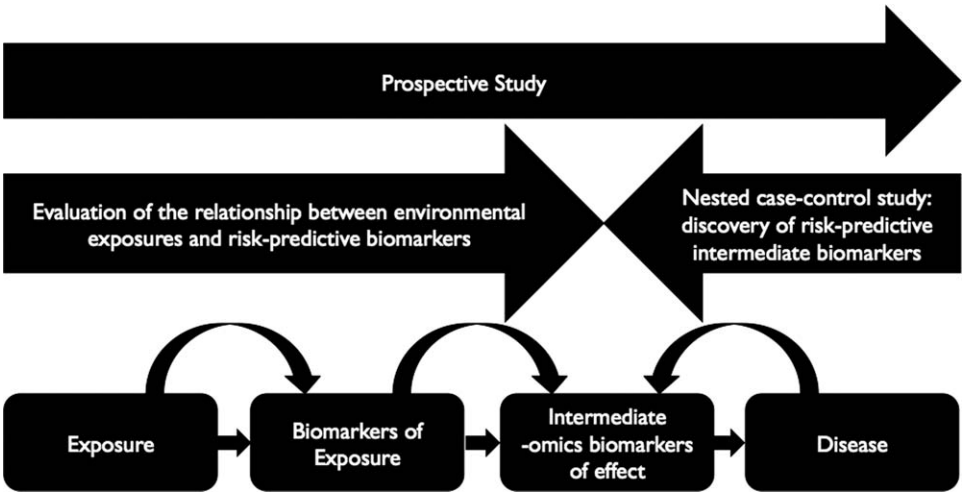

Fig. 1. The "meet-in-the-middle" approach. In prospectively collected cohorts, biological samples are characterized using omics platforms (e.g., transcriptomics, proteomics, and metabonomics) to identify molecules that represent intermediate markers of early effect. These are used to link exposure metrics/biomarkers of exposure with disease endpoints.

exposure or disease, or whether their measurement is confounded by some other exposure. A way to show that levels of EDCs in the blood contribute to breast cancer risk independently of confounding by other risk factors would be to show that the levels of EDCs are associated with different genotypes or polymorphisms, which, in turn, predict disease (concept of Mendelian randomization [MR]) [Ebrahim and Davey Smith, 2008]. For example, three studies found higher breast cancer risk associated with higher exposures to PCBs (well-established ECDs) among postmenopausal white women with a polymorphism in the *CYP1A1* gene [Moysich et al., 1999; Laden et al., 2002; Zhang et al., 2004]. Being assorted randomly from one generation to the next, variants of the *CYP1A1* gene are independent of confounding factors. Hence, the finding that a polymorphism in the *CYP1A1* gene was associated with breast cancer would provide indirect proof of a genuine involvement of PCBs in the etiology of the disease. However, because genome-wide association studies (GWAS) and familial studies of breast cancer point to a rather small genetic effect (5–10%), the lack of the two levels of association should not be taken to mean the absence of a potentially important causal exposure.

In addition to similarities, the differences between MITM and MR should also be stressed. In particular, in MR, (a) genetic susceptibility precedes exposure and is expected to modulate it; (b) gene variants are fixed from birth; (c) and gene variants are usually not confounded.

TABLE I. Genes Found to be Differentially Expressed After EDC Exposure and in Relation to Breast Cancer

Study		Number of significant associations (*N*)	CD44	CEBPD[a]	CYP1B1[a]	GALR2	IGFBP3	LAMB3	PGR	SYK	THBS1[a]	TIMP3	TP53
Exposure versus omics													
Boehme		86		X		X		X	X		X		
De	Exposure												
	PCB-153	51			X		X			X			
	PCB-77	41	X				X						X
Buterin	Cell line												
	MCF7	133		X	X							X	
	T47D	76			X						X		
Epigenomics versus BC													
Jovanovic		43						X	X	X			
Huang		186	X	X	X	X	X	X	X	X	X	X	X

The top part of the table displays genes found to be differentially expressed (in gene expression studies) after endocrine disrupting compounds (polychlorinated biphenyl, dichlorodiphenyltrichloroethane, and Dichlorodiphenyldichloroethylene) exposure. The lower part shows genes that have been reported to be associated with breast cancer in epigenetic studies. The established overview of this table is a comparison of genes mentioned in both Steps 2 and 3 of the MITM approach (see main text).

[a]Genes reported by at least two studies as differently expressed after PCB/DDT/DDE exposure and also reported to be associated with breast cancer in epigenetic studies.

In the MITM approach, exposure precedes changes in biomarkers, the latter are variable in time and also subject to confounding; therefore, the interpretation of MITM requires more caution. In both cases, lack of evidence (for MD or MITM) is not evidence of lack of causality.

MITM: An Example From a Single Study on Epigenomics and Breast Cancer

The second example is based on preliminary data from a small study on epigenomics (unpublished). Ninety-six breast cancer cases and their 96 matched controls nested within the European Prospective Investigation into Cancer and Nutrition cohort underwent genome-wide methylation profiling (Illumina 450K chip). We identified the 10 CpG sites whose methylation fraction exhibited the strongest association with breast cancer onset. We looked at the association between these candidates and well-established breast cancer risk factors. Of the 10 candidate signals, four signals (cg01546046, cg08890338, cg13298827, and cg13665852) were associated exclusively with reproductive factors, and in particular with ever having breast fed and age at menopause, of hundreds of different exposure items. The application of the MITM approach in the current context is based on the investigation of risk factor versus disease; omics versus disease; and omics versus risk factor. The observation lends credibility in this case not only to the original association (risk factor and disease, which is well known) but in particular to the association between omics and disease, which is novel. This example illustrates one key advantage of the MITM to perform causal inferences based on high-dimensional data. Although typical multiple-testing correction strategies (e.g., Bonferroni correction) may provide stringent criteria to guard against false-positive candidates, they may also exclude weaker yet relevant markers. This is especially true when small sample sizes are considered. The MITM approach can therefore complement the list of putative influential markers by considering their biological relevance, irrespective of their formal statistical significance.

PRACTICAL AND TECHNICAL REQUIREMENTS

Calibration

Existing population-based studies are often limited by poor exposure assessment and uncertain biological plausibility of associations that are detected, leading to false-positive and false-negative findings. This is of particular importance in studies on environmental exposures, such as those focused on air pollution and water contamination. Many studies conducted with in-depth exposure assessment tend to be too small for causal inference on disease, and there is a conflict between "depth" and "breadth," that is, accuracy of individual information and study size. Well-conceived, fully controlled short-term intervention studies, such as the Oxford Street Randomized Trial [McCreanor et al., 2007], have shown that acute changes can occur in lung and heart function at low or very low levels of exposure to air pollutants. However, the identification of long-term effects has been problematic due to the lack of the same degree of accuracy in exposure assessment as obtained in short-term experimental studies. A challenge is therefore in bringing together both types of investigations and in linking personal exposure monitoring (high precision, used mostly in short-term studies) with up-to-date land use regression modeling and satellite-based exposure assessment (less precise methods applicable in large populations). A recently EU-funded project called EXPOsOMICS will enable the collection, in several European populations, of modeled exposures based on refined models and cutting edge technologies such as omics, as well as personal exposure measurements (PEM). Based on samples in which both types of data are available, we will define calibration models to optimize the prediction of true exposure (using PEM studies) from (possibly complex combinations of) variables in the modeled exposure matrices. Statistical methods involved include classical measurement error models and their Bayesian alternatives (filtering models), providing calibrated exposure estimates for air and water pollutants. The resulting estimates will be used within EXPOsOMICS for (a) the characterization of the internal response to the external component of the exposome and (b) to study the association between the calibrated exposures and disease endpoints. The latter includes the calibration component.

A Conceptual Model of Life-Course Disease Risk

Population studies of chronic diseases have traditionally recruited middle-aged subjects, whereas there is strong evidence that (a) the risk of disease is influenced by early exposures, including in *utero*, and (b) life stages include critical periods (during which changes in exposure have long-term effects on disease risks or related, intermediate markers) and sensitive periods (during which an exposure has stronger effect on development and, hence, disease risk than at other times) [Ben-Shlomo and Kuh, 2002].

The idea of a sequence of critical and sensitive periods leads to the concept of "chain of risk," that is, the interplay of early exposures and late exposures. The practical use of this concept implies having access to multiple life stages in exposure assessment and epidemiological studies and repeated measurements of biomarkers at different time windows. This approach requires an intergenerational epidemiological study design. The intrinsic dynamic nature of the exposome, which encompasses all exposures from conception to death, makes statistical

analysis a challenging task requiring the inclusion of a temporal component in causal inferences [Vermeulen and Chadeau-Hyam, 2012]. Unfortunately, there is currently no single cohort that covers all periods of life from conception to death, and therefore, we need to look at different age cross sections using information from separate cohorts. In addition, the vast majority of cohorts just have one spot biological sample, which makes the life-course assessment of (causal) relationships even more difficult.

Although epidemiology usually treats time in a rather simplified manner, for example, by stratifying risk measures by duration of exposure or latency since first exposure, recent developments have proposed refined ways to account for a dynamic exposure function (exposure history) and its effect on health outcomes (exposure increments) [Lubin et al., 2009, 2012; Richardson et al., 2012]. Alternative approaches include models for the reconstruction on the natural history of disease progression such as compartmental models. In these approaches, the population is subdivided into states corresponding to observed (or assumed) steps in the disease process and transition from one state to another reflects an evolution of the health status of the studied population (or of each individual included). Such models were initially developed to study spread of infectious diseases but were extended to noncommunicable diseases. In chronic disease epidemiology, they can be used to estimate the evolution of health status among known steps of disease progression. These stages can either be observed or hidden (e.g., if the prevalence of the underlying asymptomatic affection cannot be measured) [Chadeau-Hyam et al., 2010; Vineis and Chadeau-Hyam, 2011]. Their application not only quantifies the impact of risk factors/exposures (and their history) on the disease risk but also gives an insight into their role in the dynamics of disease progression and provide refined estimates of their predictive performances: their ability to help in the reconstruction of the individual (or population based) trajectories across health states.

The extensive investment in the field of GWAS has given rise to numerous high-dimensional datasets and opened doors to a now well-established "good practice" consisting in the replication of findings through validation in multiple independent epidemiological studies. This same requirement for causal assessment of any putative marker is expected in the field of omic-based biomarkers, through multiple nested validations. This is even more important in this case given the additional uncertainties related to environmental exposures and their high variability [Ioannidis and Khoury, 2011].

As discussed in the Introduction section, the exposome may prove to be one of the best ways of addressing "systems biology" because of the powerful analytical omic instruments now available and should contribute to a cross-omics validation of the findings: if signals found in different omics platforms do belong to the same (and biologically relevant) pathway, then their causal link with the health outcome should be strengthened. However, the identification of these patterns remains computationally challenging as we seek for common signals within a large set of highly correlated data. The recent developments in network theory will help in that respect and should enable the investigation of relationships between external environment and internal changes at different levels (DNA, transcripts, miRNA, proteins, metabolites, etc.). The application of network models as developed by Albert-Laszlo Barabasi investigate "emergence" of new properties from the interaction of different underlying molecular phenomena. This has provided an exciting new approach [Barabasi, 2007] and contributed to the identification of complex protein–protein interactions, in which some proteins exert a more prominent role (hubs) and others less so [Vidal et al., 2011]. Central to this approach is that molecules are not seen in isolation but within their biochemical contexts (networks). Components are "nodes" and interactions are "edges," using the tools of graph theory [Joffe et al., 2012]; for example, in the field of metabolomics, small molecule metabolites are represented by nodes, and metabolic reactions (metabolite interconversions) are represented by edges. It appears that a series of "basic and reproducible organizing principles" regulate different processes like cell functioning and technological and social systems. Some important discoveries have been made by applying the network theory, for example, it was discovered that hub proteins tend to correspond to more ancient genes and have evolved more slowly and that the deletion of the corresponding genes has a larger diversity of phenotypic outcomes [Vidal et al., 2011]. Another important concept is that multiple diseases may not be independent, that is, a "phenome" can be reconstructed based on the underlying molecular networks that do not correspond to the existing classification of diseases (e.g., International Classification of Diseases).

CONCLUSIONS

The introduction of omics in environmental epidemiology represents a new paradigm for studying the influence of environmental agents on health over a lifetime. The ability to generate efficiently parallel omics datasets and recent advances in personal environmental monitoring provide a strong technological backdrop for exploiting cohort biospecimens to assess personal-level exposures and responses. Several aspects of these studies require methodological development to make best use of such opportunities. What is clear is that the scope of the exposome concept will require the input and collaboration of researchers across many disciplines for exposome studies to succeed and will rely as much on the design and

philosophy that underpin them as on our ability to generate complex datasets about their participants.

ACKNOWLEDGMENTS

The authors are grateful to Steven Rappaport and Pietro Ferrari for thoughtful comments.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing of the manuscript. P. Vineis and K. van Veldhoven conducted the literature analysis on endocrine-disrupting chemicals.

REFERENCES

Barabasi AL. 2007. Network medicine—From obesity to the "diseasome". N Engl J Med 357:404–407.

Ben-Shlomo Y, Kuh D. 2002. A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. Int J Epidemiol 31:285–293.

Boehme K, Simon S, Mueller SO. 2009 Gene expression profiling in Ishikawa cells: A fingerprint for estrogen active compounds. Toxicol Appl Pharmacol 236:85–96.

Buterin T, Koch C, Naegeli H. 2006 Convergent transcriptional profiles induced by endogenous estrogen and distinct xenoestrogens in breast cancer cells. Carcinogenesis 27:1567–1578.

Chadeau-Hyam M, Clarke PS, Guihenneuc-Jouyaux C, Cousens SN, Will RG, Ghani AC. 2010. An application of hidden Markov models to the French variant Creutzfeldt–Jakob disease epidemic. J R Stat Soc Ser C: Appl Stat 59:839–853.

Chadeau-Hyam M, Athersuch TJ, Keun HC, De Iorio M, Ebbels TM, Jenab M, Sacerdote C, Bruce SJ, Holmes E, Vineis P. 2011. Meeting-in-the-middle using metabolic profiling—A strategy for the identification of intermediate biomarkers in cohort studies. Biomarkers 16:83–88.

De S, Ghosh S, Chatterjee R et al. 2010 PCB congener specific oxidative stress response by microarray analysis using human liver cell line. Environ Int 36:907–917.

Deacon T. 2007. Three levels of emergent phenomena. In: Murphy N, and Stoeger WRSJ,, editors. Evolution and Emergence: Systems, Organisms. Oxford: Oxford University Press. pp 88–110.

Ebrahim S, Davey Smith G. 2008. Mendelian randomization: Can genetic epidemiology help redress the failures of observational epidemiology? Hum Genet 123:15–33.

Ellis G. 2012. Top-down causation and emergence: Some comments on mechanisms. Interface Focus 126–140.

Exposome. Exposome Alliance Project. Available at: http://exposomealliance.org/. Accessed on February 2013.

Huang TH, Esteller M. 2010. Chromatin remodeling in mammary gland differentiation and breast tumorigenesis. Cold Spring Harb Perspect Biol 2:a004515.

Ioannidis JP, Khoury MJ. 2011. Improving validation practices in "omics" research. Science 334:1230–1232.

Joffe M, Gambhir M, Chadeau-Hyam M, Vineis P. 2012. Causal diagrams in systems epidemiology. Emerg Themes Epidemiol 9:1.

Jovanovic J, Ronneberg JA, Tost J, Kristensen V. 2010. The epigenetics of breast cancer. Mol Oncol 4:242–254.

Laden F, Ishibe N, Hankinson SE, Wolff MS, Gertig DM, Hunter DJ, Kelsey KT. 2002. Polychlorinated biphenyls, cytochrome P450 1A1, and breast cancer risk in the Nurses' Health Study. Cancer Epidemiol Biomarkers Prev 11:1560–1565.

Lubin JH, Purdue M, Kelsey K, Zhang ZF, Winn D, Wei Q, Talamini R, Szeszenia-Dabrowska N, Sturgis EM, Smith E, Shangina O, Schwartz SM, Rudnai P, Neto JE, Muscat J, Morgenstern H, Menezes A, Matos E, Mates IN, Lissowska J, Levi F, Lazarus P, La Vecchia C, Koifman S, Herrero R, Franceschi S, Wunsch-Filho V, Fernandez L, Fabianova E, Daudt AW, Maso LD, Curado MP, Chen C, Castellsague X, Brennan P, Boffetta P, Hashibe M, Hayes RB. 2009. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: A pooled analysis of case–control studies. Am J Epidemiol 170:937–947.

Lubin JH, Cook MB, Pandeya N, Vaughan TL, Abnet CC, Giffen C, Webb PM, Murray LJ, Casson AG, Risch HA, Ye W, Kamangar F, Bernstein L, Sharp L, Nyren O, Gammon MD, Corley DA, Wu AH, Brown LM, Chow WH, Ward MH, Freedman ND, Whiteman DC. 2012. The importance of exposure rate on odds ratios by cigarette smoking and alcohol consumption for esophageal adenocarcinoma and squamous cell carcinoma in the Barrett's Esophagus and Esophageal Adenocarcinoma Consortium. Cancer Epidemiol 36:306–316.

McCreanor J, Cullinan P, Nieuwenhuijsen MJ, Stewart-Evans J, Malliarou E, Jarup L, Harrington R, Svartengren M, Han IK, Ohman-Strickland P, Chung KF, Zhang J. 2007. Respiratory effects of exposure to diesel traffic in persons with asthma. N Engl J Med 357:2348–2358.

Moysich KB, Shields PG, Freudenheim JL, Schisterman EF, Vena JE, Kostyniak P, Greizerstein H, Marshall JR, Graham S, Ambrosone CB. 1999. Polychlorinated biphenyls, cytochrome P4501A1 polymorphism, and postmenopausal breast cancer risk. Cancer Epidemiol Biomarkers Prev 8:41–44.

Rappaport SM, Smith MT. 2010. Epidemiology. Environment and disease risks. Science 330:460–461.

Richardson DB, Cole SR, Langholz B. 2012. Regression models for the effects of exposure rate and cumulative exposure. Epidemiology 23:892–899.

Sellars R. 1970. Principles of Emergent Realism: The Philosophical Essays of Roy Wood Sellars. In: Warren WP, editor. St. Louis, MO: Warren H. Green.

Van Gulick R. 2007. Who's in charge here? And who's doing all the work? In: Heil J. and Mele A., editors. Evolution and Emergence: Systems, Organisms, Persons. Oxford: Oxford University Press. pp 233–256.

Vermeulen R, Chadeau-Hyam M. 2012. Dynamic aspects of exposure history-do they matter? Epidemiology 23:900–901.

Vidal M, Cusick ME, Barabasi AL. 2011. Interactome networks and human disease. Cell 144:986–998.

Vineis P, Chadeau-Hyam M. 2011. Integrating biomarkers into molecular epidemiological studies. Curr Opin Oncol 23:100–105.

Vineis P, Khan AE, Vlaanderen J, Vermeulen R. 2009. The impact of new research technologies on our understanding of environmental causes of disease: The concept of clinical vulnerability. Environ Health 8:54.

Wild CP. 2005. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 14:1847–1850.

Wild CP. 2012. The exposome: From concept to utility. Int J Epidemiol 41:24–32.

Zhang Y, Wise JP, Holford TR, Xie H, Boyle P, Zahm SH, Rusiecki J, Zou K, Zhang B, Zhu Y, Owens PH, Zheng T. 2004. Serum polychlorinated biphenyls, cytochrome P-450 1A1 polymorphisms, and risk of breast cancer in Connecticut women. Am J Epidemiol 160:1177–1183.