

Commentary

Understanding environmental causes of disease: what can we expect from new concepts and technologies?

Paolo Vineis¹, Aneire E. Khan¹, Jelle Vlaanderen², Roel Vermeulen²

1. MRC Centre for Environment and Health, Imperial College, London, UK

2. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands

Emails of authors:

p.vineis@imperial.ac.uk

aneire.khan@imperial.ac.uk

J.J.Vlaanderen@uu.nl

R.C.H.Vermeulen@uu.nl

*Correspondence address: Prof. Paolo Vineis, MRC Centre for Environment and Health, Department of Epidemiology and Public Health, Imperial College London, Room 151, St Mary's Campus, Norfolk Place W2 1PG, London, UK. E-mail: p.vineis@imperial.ac.uk ; Phone: +44 (0)20 7594 3372; Fax: +44 (0)20 7594 3196

Acknowledgments: This study was made possible by the ECNIS grant from the European Union (FOOD-CT-2005-513943). We are grateful to Ross Anderson for thoughtful comments and the permission to reprint Figure 2.

ABSTRACT

In spite of decades of epidemiological research, the etiology and causal patterns for many common diseases, such as breast and colon cancer or neurodegenerative diseases, are still largely unknown. Such chronic diseases are likely to have a largely environmental origin. For example, the observation of incidence rates in migrant populations unequivocally showed that migrants rapidly acquire the risk of disease that is typical of the populations where they move. However, “environmental” risks have been often elusive in epidemiological studies. This is a conundrum for current epidemiological research. On the other side, the relative contribution of genes to chronic diseases seems to be modest (15-50% increase in disease risk). In spite of that, genetic associations – e.g. as emerging from GWAS - are perceived as more exciting and more credible than environmental associations. Possible explanations for such divergence are (a) that genes are stable in the course of life (while exposures are changing over time), and (b) that associations are more easily replicated in independent genetic studies.

In addition to the rapid developments in genetics, a further recent revolution is the application of biomarkers in human observational studies, particularly with the introduction of “-omic” high-throughput technologies, such as transcriptomics, proteomics and metabolomics. These techniques are still largely immature. However, they are likely to provide, in the next years, powerful tools to investigate early effects of environmental exposures and understand the etiology of common diseases better. What is yet to be explored extensively is a potential model for the effects of low doses in the absence of thresholds, incorporating genetic and acquired susceptibility (“*clinical vulnerability*”) and the cumulative effects of different exposures. We propose a disease model that is compatible with the weak associations found by GWAS and the still elusive role of many (low-level) environmental exposures as measured by traditional epidemiological tools. The development of “-omics”, in spite of current limitations, could greatly contribute the elucidation of the disease model we propose.

INTRODUCTION

We are still far from understanding the etiology and causal patterns for many common diseases, such as breast and colon cancer, or neurodegenerative diseases. In spite of decades of epidemiological research, the etiology of such conditions is still largely unknown.

For a long time it has been claimed that most chronic diseases have an environmental origin (using the term “environment” in a broad sense, to indicate essentially what is not due to genetic predisposition). This claim was based on descriptive data showing the broad range of incidence rates in different parts of the world, the rapid temporal changes – such as those currently occurring in China and India – and the crucial observation of incidence rates in migrant populations. The latter unequivocally showed that migrants rapidly acquire – sometimes already in the first generation after migration - the risk of disease that is typical of the population where they move.

However, “environmental” risks have been largely elusive in epidemiological research, and there is still much debate on the real impact of the environment, whose role tends to be overshadowed in recent times by the successes of genetics, in particular Genome-wide association studies (GWAS). Clearly, we need a disease model that allows us to encompass both the mass of genetic data coming from GWAS and the still elusive role of many environmental exposures.

Genes vs. environment

The technological revolution in genetics, which in 2007-2008 led to the identification of a number of novel genes for several common diseases, is inspiring but at the same time a little disappointing. Most highly-penetrant genes already had been uncovered by linkage studies, so that “genome-wide scans” could mainly identify low-penetrant genes. It is a fact that – with a few exceptions – what has been found is genetic variants weakly associated with chronic diseases, with relative risks of 1.15-1.5. These risks are much lower than some known environmental exposures, and of the same order of magnitude of other environmental exposures that are usually rejected as

too weakly associated with disease to be meaningful. For example, Hung *et al* have found an allelic relative risk of 1.3 for a gene variant associated with lung cancer, in the putative nicotinic acetylcholine receptor subunit on chromosome 15q25 [1]. The genotype nicely combines with the habit of smoking, increasing the probability of developing lung cancer in smokers (Figure 1) [1]. However, it is clear from the figure that the effect of smoking is by far greater than the effect of the genotype. The former increases the cumulative risk in life from less than 1% (never smokers) to about 15-20% (current smokers), while the effect of the genotype is a small additional amount (~1 to 2%) that does not seem to contribute much to the Public Health problem of lung cancer. Of course genetic research is not tailored for Public Health purposes. It can be extremely worthwhile, instead, (a) for the investigation of disease mechanisms, (b) to shed light on potential therapies, or (c) to suggest clues for unidentified etiological agents in the case of diseases whose etiology is still unknown.

It is unclear, however, how goals (b) and (c) can be met if the relative contribution of genes is so small (15-50% increase in disease risk) which is in the same order of magnitude of relative risks found for unclear and controversial environmental exposures, such as electromagnetic fields and leukemia, or some pesticides and prostate cancer. It is therefore useful to reflect on the potential reasons why weak genetic associations (generally) are perceived as much more noteworthy than environmental associations.

1. One reason – perhaps the main one - is certainly the greater robustness of technical detection of gene variants compared to environmental exposures. It is estimated that the error rate with common genotyping techniques is less than 10%, while it can be as high as 50% for environmental exposure assessment, at least for dichotomous exposure variables. Moreover genes are stable in the course of life, while environmental exposures are changing over time (let us think of diet) leading to more inherent measurement error in environmental exposure than in gene variance assessment.

2. A second obvious reason is replication. In the study mentioned above on lung cancer, the association was found in five independent studies. This usually is not the case for environmental exposures, where replication is not easily achievable. For example, a cohort study from Belgium suggested that high levels of cadmium in urine are associated rather strongly with lung cancer [2]. The study is strong (though small), because exposure assessment is based on a biomarker reflecting long-term exposure (i.e. with much higher accuracy than usually available in epidemiology). However, the study is not easily reproducible, because there are not many other populations with similar levels of environmental exposure to cadmium, and there might be several confounders or effect modifiers that either obscure the association with cadmium or change the strength of the association. This is commonly observed in environmental epidemiology, i.e. a number of small-medium size associations that are difficult to reproduce in different settings, both because of methodological difficulties and because of variable study contexts.

Alternatives to epidemiology: research in animals, biomarkers and omics

For the second reason described above, Lorenzo Tomatis, former Director of the International Agency for Research on Cancer (IARC) had proposed already in the 70's that prevention of human cancer could not rely on epidemiology alone, and promoted research in laboratory animals as a surrogate for research in humans. The IARC Monographs are the most prestigious instrument for cancer prevention, thanks to their sound scientific methodology, and rely, besides human epidemiological evidence, heavily on research in laboratory animals for the categorization of carcinogens. This is illustrated in the case of 1,3-butadiene, whose carcinogenicity has been confirmed recently by a review of the Working Group of the IARC Monographs in 2007 [3]. More than 20 years ago, experiments in rodents showed that this widely used chemical induced cancers at multiple organ sites, including a very high incidence of otherwise extremely rare cancers (e.g. heart hemangiosarcomas). Also an excess of lymphopietic cancers was found in laboratory

animals (3). There was no doubt that the chemical was a potent animal carcinogen, given the consistency of the observations, the presence of a dose-response relationship, the unusual type of tumours induced, and the very high incidence. Even today, we still lack a satisfactory number of sound epidemiological studies capable of confirming these observations in humans. The results from the available studies, however, are remarkably consistent with animal studies, at least for lymphopietic tumours, given the considerable difficulties encountered in such investigations.

A few new developments occurred after the foundation of the IARC Monographs (the early seventies) of which one is the genetic revolution. This revolution is exciting, will probably contribute to the unraveling of mechanisms underlying cancer, and may contribute to the identification of some environmental carcinogenic exposures. But it is not likely to be the main avenue to prevent cancer. The second revolution is the application of biomarkers in human observational studies, which more recently has been revolutionized with the introduction of “-omic” high-throughput technologies, such as transcriptomics, proteomics and metabolomics. These techniques are still largely immature, the measurement error per analyte is unknown and large intraindividual variability is to be expected, thus blurring any association with environmental exposures and/or disease. However, they are likely to provide, in the next years, powerful tools to investigate early effects of environmental exposure and understand the etiology of common diseases better [4]. An example is the recent paper by Holmes and others on the metabonomic profiles associated with hypertension [5].

Low doses and acquired susceptibility to disease

There is also an unsung revolution which has taken place recently within epidemiology, which complements those referred to above, but which has attracted less attention. This is the increasing ability of epidemiological tools to consistently unravel the effects of low environmental exposures. The cases of air pollution and environmental tobacco smoke (ETS) are exemplar. In the seventies and eighties many researchers thought it was impossible to detect plausible causal

associations with such low-level exposures. The underlying idea was that “noise” (bias, confounding) was larger than the signal. However, a large number of well-conducted studies have been published since: in the case of ETS more than 60 studies show (with few exceptions) increased risks of lung cancer in the order of 1.25, i.e. the same magnitude as the gene for the nicotine acetylcholine receptor (1, 6, 7). For air pollution, an association with lung cancer has been reported in six cohort studies. One of these studies is sufficiently large as to show the association in non-smokers [6,7]. Again, the relative risk is around 1.25. In both cases biomarkers, such as cotinine or NNK for ETS and DNA adducts for air pollution, have contributed to make the association more plausibly causal.

We have thus to acknowledge that ETS and air pollution (mainly due to traffic exhaust) are able to induce, after long-term exposure, chronic diseases such as cancer and also coronary artery disease. The interesting observation is that this happens at dose levels that are much lower than those of tobacco smoke or “classical” carcinogens. For example, exposure to ETS occurs at levels that are 1/100 compared to active smoking [8].

One potential explanation for the effects of extremely low doses and the absence of a threshold is based on acquired susceptibility and the cumulative effects of different exposures. As defined by Rothman and Greenland, “the cause of a disease event is an antecedent event, condition or characteristic that was necessary (given that all other conditions are fixed) for the occurrence of the disease at the moment it occurred” [9]. Said in other words, a cause can also be viewed as something that “completes an incomplete causal chain” [10] or precipitates a chain of events which creates a state of vulnerability. This concept of “clinical vulnerability” is expressed in Figure 2 [11]. Exposure to low levels of e.g. ETS or air pollution is not a “cause” of cancer in itself (like an accident is the cause of a death), but because it occurs on top of pre-existing vulnerability. This could well explain why small changes in environmental exposures can have big effects, if they occur in a population of vulnerable subjects.

Vulnerability can be acquired or genetically-based. The concept of acquired “clinical vulnerability” is related to previous insults/ pathophysiological changes that predispose to disease, as expressed in Figure 2. An example is the finding of a greater effect of ETS among ex-smokers compared to never smokers in a large prospective investigation [12]. It is plausible that ex-smokers have a greater vulnerability because of already existing mutations or epigenetic changes, so that further exposure to ETS leads to selection and clonal expansion of mutated cells.

Another type of vulnerability (more often called hypersusceptibility) is genetically determined. Many years ago we showed that subjects with the genetically based NAT2 slow acetylator genotype could have greater susceptibility to being damaged by tobacco smoke related arylamines at lower levels of exposure rather than at higher levels [13]. Our reasoning was that on very rare occasions, e.g. among people exposed to extremely high doses of potent carcinogens, the whole population or a vast majority develops cancer. This is what happened among British chemical workers exposed to 2-naphthylamine in the 1950s and before. For example, all 15 workers exposed to 2-naphthylamine in a plant developed bladder cancer, probably the only example of a “sufficient” exposure in the history of carcinogenesis [14]. It is clear that in that case genetic susceptibility was totally irrelevant.

Mendelian randomization tests

If we believe in the hypothesis of a cumulative effect of exposures that increase vulnerability, so that even low doses of exposure to carcinogens can lead to the effect (like in Figure 2), how can we design better studies to test the hypothesis? One possibility is to include in questionnaires sections about the past history of exposure, so to identify vulnerabilities. However, this is a relatively weak approach that has not given clear results. For example, it is still not clear what the joint effect of exposures such as smoking and asbestos is on lung cancer, after many years of research, mainly because of exposure misclassification that leads to underestimates particularly when addressing interaction.

An elegant way to address such a sophisticated hypothesis is through Mendelian Randomization, i.e. exploiting the random assortment of genes from parents to offspring. Since gene variants are randomly assorted from parents to the offspring during conception (and following meiotic recombination), then the gene variants one has do not depend on her exposures or social class (with some caveats and exceptions). If one can show that a gene involved in a process that increases vulnerability to disease is also associated with cancer or CVD, then this is additional support to the theory that exposures increasing vulnerability are really causally involved. An example might be exposure to air pollutants or ETS, inflammatory processes, cardiovascular disease and cancer. A Mendelian Randomization test would be looking at variants of genes in the NF-kB pathway (IL1, IL6) or COX-2. If these variants are associated with CVD or lung cancer, then it is more likely that vulnerability induced by ETS and air pollution is causally involved in disease etiology.

The contribution of intermediate biomarkers and '-omics'

The concept of acquired “clinical vulnerability” is related to previous insults/pathophysiological changes that predispose to disease. Intermediate markers and specifically ‘-omics’ could be extremely useful in tracing the “history” of such insults and in reflecting the cumulative effect of different exposures.

Intermediate Biomarkers

Intermediate biomarkers directly or indirectly represent events on the continuum between exposure and disease. Intermediate biomarkers can provide important mechanistic insight into the pathogenesis of environmental diseases. As such, they complement classic epidemiological studies that use disease endpoints. Examples of early biologic effect biomarkers include measures of cellular toxicity, chromosomal alterations, DNA, RNA and protein expression, and early non-neoplastic alterations in cell function (e.g., altered DNA repair, altered immune function). For

maximum utility, an intermediate biomarker must be shown to be predictive of developing disease. The criteria for validating intermediate biomarkers have been discussed by Schatzkin and colleagues [15,16] and focus on the calculation of the etiologic fraction of the intermediate endpoint, which varies from 0 to 1. The closer the etiologic fraction is to 1, the greater the biologic marker reflects events, either directly or indirectly, on the causal pathway to disease. For instance, bone-marrow toxicity has been linked to future risk of leukemia. Therefore, environmental exposures that lead to bone-marrow toxicity are potentially leukemogens. Bone-marrow toxicity can be measured by studying peripheral blood cell counts including circulating progenitor cells. Studies based on such endpoints have the advantage that the biological marker can be measured quantitatively and therefore increase the power as compared to dichotomous (rare) disease endpoints. Furthermore, these studies can be carried out using cross-sectional or semi-longitudinal study designs in which exposure can be accurately assessed. An example of the use of intermediate markers is a recent study on hematological effects of benzene at low levels of exposure [17]. In this study peripheral blood cells and progenitor cell colony formation were significantly decreased among low exposed individuals in a dose-dependent manner. Moreover, these effects occurred at exposure levels lower than previously thought safe [17].

-Omics

“-Omics” tools can be directly applied to samples from an epidemiologic case-control or cohort study to better characterise intermediate pathways, providing the ‘missing links’ among exposures, genes, and diseases [18]. The term ‘-omics’ generally refers to the rigorous study of a complete set of biological molecules with high-throughput techniques [19]. One of the main hallmarks is the departure from traditional molecular biology focusing on a single biological structure (e.g. a single gene or protein) to a more comprehensive analysis of biological systems.

Cells and tissues alter their metabolism or gene expression in response to exposure as indicated by for example, the induction of specific groups of genes in response to heat shock,

hypoxia, or osmotic stress [20-22]. These responses can produce patterns of specific changes in gene expression, proteins, or metabolic profiles that reflect exposure to a particular agent or class of agents which can be interpreted as a “history” of such exposures and might reflect the cumulative effect of different exposures. Some of the promising –omic technologies in environment health, which offer new opportunities to detect and quantify these changes, are discussed below (also see table 1)

Genomics, one of the more mature technologies of “–omics”, is the study of all of the genes in an organism. The field of genomics is divided into three areas, namely, genotyping, gene expression profiling (or transcriptomics), and epigenomics. These technologies are applicable to detect constitutive variations as well as somatic mutations in tumours. It is expected that genomic techniques with well-planned laboratory research may help identify the causal gene variants that underlie susceptibility or resistance to common diseases in the next few years [23]. Other types of genomics with still uncertain applications to environmental diseases include functional and evolutionary genomics. Proteomics studies large numbers of proteins and protein fragments in substrates like blood, urine, and tissue samples collected at a specific point in time. The major application in epidemiology is likely to be in the detection and early diagnosis of diseases. Although proven controversial, the most prominent application has been in reporting a distinctive pattern of low-molecular weight proteins that characterises the sera of women with early-stage ovarian cancer, that is very rarely found in control sera [23].

A further promising application of the “–omics” technologies is the assessment of normal and abnormal protein patterns of metabolites not directly connected to a disease diagnosis, referred to as ‘metabolomics’ [23]. Metabolomics has the potential to move the study of disease development to preclinical stage by defining patterns of protein abundance that have correlations with risk of future diseases and in extension to specific environmental factors. Metabolomics has recently been adapted to high throughput, highly sensitive technologies, like SELDI-TOF/MS, LC-

MS/MS, and NMR, enabling the screening of large numbers of samples for a large number of potential markers. Although there are still many limitations in the current techniques and in their applications in epidemiological studies which are characterized by small sample sizes and generally poor study designs, it is probably a matter of time before these techniques produce new leads in the discovery of novel intermediate markers. This will be largely attributable to a better understanding and consideration of methodological issues [4].

Application of “-omics” to investigate human toxicological effects from environmental insults

Genotyping

Genotyping involves genome-wide analysis of single nucleotide polymorphisms (SNPs) or high-throughput sequencing of entire genes. SNPs, the most common and best studied genetic variation, are highly abundant in the human genome and they are used as markers for genetic variation in disease-gene association studies. In environmental health, genotyping is used to associate variations to phenotypes of interest, using direct association studies and indirect association studies. Direct association studies are based on a list of candidate genes for which a plausible hypothesis can be stated to relate gene to phenotype, while indirect association studies are based on the concept that recombination seldom occurs between multiple SNPs that are in close proximity on a chromosome.

In a case-control study with 880 cases of basal cell carcinoma, 666 cases of squamous cell carcinoma and 780 controls, Applebaum *et al* investigated whether polymorphisms in nucleotide excision repair genes XPA and XPD modify the association between exposure to arsenic (As) and non-melanoma skin cancer [24]. They observed an increased risk of As exposure induced basal cell carcinoma among subjects with homozygous variant for XPA. For individuals varying for both XPD polymorphisms, there was a 2-fold increase in risk of squamous cell carcinoma associated with elevated arsenic exposure.

Gene expression profiling

Gene expression profiling can be used to determine which genes are differently expressed as a result of changes in environmental conditions. A gene expression profiling study typically involves a group of individuals with similar exposure level or phenotype (e.g. disease status) and compares the gene expression profile of this group to the profile of a reference group. To avoid the problems of large interindividual variations compared with the small changes caused by environmental exposure, Wang et al conducted a self-controlled study design, with measurements made before and after individual exposure to welding fumes. The group was stratified according to smoking status (which profoundly affected the whole blood expression profiles). Nonsmokers after exposure to welding fumes exhibited altered gene expression in 35 genes from eight functional pathways, including processes related to oxidative stress, proinflammatory responses, phosphate metabolism, cell proliferation, and apoptosis [25]. A study by Lu *et al* used a human cancer cDNA expression array to detect differential gene expression patterns determined in liver tissue samples from arsenic exposed individuals and healthy liver tissues[26]. Approximately 60 genes (10%) were differentially expressed in arsenic exposed human liver tissues compared to controls. The differentially expressed genes included those involved in cell-cycle regulation, apoptosis, DNA damage response and intermediate filaments. The observed gene alterations appeared to be reflective of hepatic degenerative lesions seen in arsenic-exposed patients.

Epigenomics

Epigenomics involves the study of epigenetic processes, which are independent of DNA sequence and are involved in the inhibition of gene expression (gene silencing), on a large scale [27-29]. Altered gene silencing plays a causal role in human diseases [30-33]. Epigenetic modifications may provide a plausible link between environmental and alterations in gene expression that might lead to disease phenotypes [30]. A large amount of laboratory animal studies already provide evidence that supports the role of environmental epigenetics in disease

susceptibility [30]. Epigenomics is mainly based on a few comprehensively studied mechanisms: (a) DNA methylation, which takes place at the carbon-5 position of cytosine in CpG dinucleotides, (b) changes to the chromatin packaging of DNA by post-translational ‘histone modification’, and more recently (c) changes in RNA interference by non-coding RNAs such as microRNA and siRNA. In mice, maternal dietary methyl-donor supplementation with folic acid, vitamin B, choline and betaine was shown to result in an increase in DNA methylation at CpG sites, thus demonstrating that the effect of a mother’s diet during pregnancy on the adult phenotype of her offspring was directly linked to DNA-methylation changes in the epigenome [30]. A few more studies on maternal dietary exposure, e.g. to phytoestrogen genistein during gestation given at a level comparable to that of human consumption with high soy diets, are particularly interesting as these results suggest the possibility that hypermethylating dietary supplements could reduce the effects of environmental toxicants, which cause DNA hypomethylation [30]. Recent experiments have shown that epigenetic changes are associated with chromatin remodelling and regulation of gene expression that underlie the development of metabolic syndrome, and to other common diseases like atherosclerosis and type 2 diabetes [34]. Deciphering epigenetic processes should allow us to target the development of new diets and drugs to prevent aberrant gene silencing, which may be involved in resistance to treatment. The importance of DNA methylation in cancer has also been established, and research now includes the mechanisms by which other chromatin modifications play a role in cancer development [35]. For example, Zhang et al demonstrated that DNA methylation binding protein (MBD2) interacts with the nucleosomal remodelling complex (NuRD) and directs the complex to methylate DNA [35].

Proteomics

Proteins are generally the functional “business end” of genomic expression affecting cellular metabolism and regulation [36]. Proteins are active agents in cells that execute the biological functions encoded by genes. The proteome consists of all proteins present in specific cell types or

tissues and, in contrast to the genome, it is highly dynamic over time, between cell types and in response to the environment.

This is perhaps the least mature of the “-omic” technologies. Whether agent-specific changes can be identified is a key question to be answered about proteomic approaches (note, the same issue applies to other “-omics”, such as metabolomics) and the issue will be resolved only after a much larger body of proteomic data on exposed cells, tissues, and biofluids is collected and analyzed [36]. Proteomics approaches are twofold: the first is the standardisation of high-throughput methods, and the second is to combine extensive protein fractionation, removal of abundant proteins and tandem MS/MS analysis of peptide sequence data to allow protein identification.

Joo *et al* performed proteomic analysis of plasma proteins, among workers exposed to benzene and matched unexposed controls and associated up-regulation of specific proteins with lymphocyte DNA damage [37]. Using MS and Western blot, they found statistically significant differences in protein profiles between the exposed and unexposed subjects. The specific proteins that were found to be up-regulated in benzene exposed workers were TCR beta, FKBP51 and MMP13. Vermeulen *et al* examined the impact of benzene on the human serum proteome in a study of benzene- exposed workers and unexposed controls, where protein-expression patterns were analysed by SELDI-TOF MS. They found that three proteins were consistently down-regulated in the exposed group compared to controls and all proteins were highly inversely correlated with individual estimates of benzene exposure ($r > 0.75$) [38].

Metabolomics

Changes in gene and protein expression can alter metabolism in particular ways that can provide distinct signatures. The metabolome is the biochemical network consisting of small molecules (e.g. lipids, vitamins) also known as metabolites. Metabolites are involved in energy

transmission in cells by interacting with other biological molecules following metabolic pathways. The metabolome is highly variable and time dependent and consists of a wide range of different chemical structures. The most important challenge in this field is to acquire qualitative and quantitative information on metabolites that occur under normal circumstances in order to be able to detect perturbations in the complement of metabolites due to changes in environmental factors.

Global metabolomic profiling may be more amenable than transcriptomic or proteomic profiling to high-throughput screening, as suggested for example by studies showing that the main yeast metabolome consists of fewer than 600 low-molecular-weight compounds (it should be noted that the yeast has only 4,000 genes and is a unicellular organism, representing thus an oversimplified model to study humans). Substantial metabolomic analysis of exposure and effects will be needed to address this possibilities offered by this approach [36,39]. Global metabolomic analysis in exposure studies in laboratory animals has just begun, but the low-level exposures of interest in large populations have not yet been addressed.

Metabolomic studies are performed with noninvasive samples such as biofluids and breath condensate as well as on tissues in vivo. Profiles within a tissue or cell are compared with profiles in biologic fluids or with cell secretion products to understand the metabolic consequences of xenobiotic-induced toxicity. Examples of this approach are the studies by Waters and colleagues, where NMR and pattern recognition analysis were used to investigate time-related metabolic effects of α -naphthylisothiocyanate on liver, urine, and plasma in laboratory animals [40,41]. Waters and colleagues found an association between hepatic lipidosis and increased urinary excretion of taurine and creatine. In addition, there was reduced urinary excretion of intermediates in the tricarboxylic acid cycle and increased excretion of plasma ketone bodies. These studies enabled a clearer understanding of key metabolic effects during the development of and recovery from toxic lesions. In the past, there was a failure in specificity for cancer, which hampered biomarker discovery by metabolomics, but technologies like NMR and MS are being reapplied to study a range of cancers

[42,43]. Recently a metabolomic study has been successfully applied to epithelial ovarian cancer and was able to discriminate women with epithelial ovarian cancer from healthy controls, based on ¹H NMR spectroscopic profiles of preoperative serum specimens [42].

The ultimate goal of using “-omics” technologies to identify environmental causes of disease is to derive an integrated view of the biological processes involved in the continuum from exposure to disease. Although high-sensitivity and data-rich toxicogenomic approaches may already be feasible for analyzing human responses to environmental stressors in some settings, applying these technologies to human populations exposed to low-level environmental contaminants will require considerably more development. In this context, the integration of information from multiple toxicogenomic approaches may provide significantly more analytical power than any one approach alone.

Current limitations of “-omics” include low reproducibility across laboratories, high intra-individual variability that hampers inter-individual comparisons, high costs, difficulties in data analysis and uncertainties in biological interpretation (see box 1). Type I, II, and III errors need to be addressed. Type I error refers to false positive results deriving from the large number of comparisons. This can be tackled with in-built replication with repeat measures in independent studies. Type II error derives from the relative lack of sensitivity of some techniques and the ensuing false negatives issue. Type III error refers to the lack of relevance of some biomarkers, with uncertain biological meaning, for the interpretation of exposure disease relationships.

“-Omics” has already played an important role in generation of new insights into the etiology of disease and gene-environment interactions, but in applying “-omics” in research, it is important to learn “what technologies can be usefully applied to which questions” [23]. With the development of validated technologies, appropriate study designs, adequate sample size, inclusion

of quality control and advanced statistical methods for data interpretation, “-omics” could potentially contribute significantly to the identification of environmental causes of disease and will help the field progress towards an integrated view of the interaction between environment and human health.

Conclusions - a model of disease

The prevailing model of “environmental disease” for decades has been influenced by toxicology, with single agents evaluated at a time, and dose as a key issue. A more recent model, which was not overall very successful, was that of “gene-environment” interactions, in which environmental agents at low doses interact with individual genetic susceptibility. What has not been explored extensively yet is another potential model for the effects of low doses and the absence of thresholds, represented by acquired susceptibility and the cumulative effects of different exposures. The concept of “*clinical vulnerability*” could explain why small changes in exposure can have big effects, if they occur in a population of vulnerable subjects. The concept of acquired “clinical vulnerability” is related to previous insults/pathophysiological changes that predispose to disease. –

The new “-omics” can be extremely useful in tracing the “history” of such insults and to reflect the cumulative effect of different exposures. However, two caveats are necessary. First, the success of “-omics” applications to epidemiology will depend on a coherent and effective model of disease, that addresses issues such the role of environmental low-level exposures. Second, “-omics” are still largely limited by low sensitivity and specificity, low reproducibility, uncertainties about the biological meaning of the findings, poor study designs and the large number of statistical comparisons. Good study designs, with in-built replication phases on the model of GWAS, are warranted before sound conclusions can be drawn.

References

- [1] Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; **452**(7187): 633-637.
- [2] Nawrot T, Plusquin M, Hogervorst J, Roels HA, Celis H, Thijs L *et al.* Environmental exposure to cadmium and risk of cancer: a prospective population-based study. *Lancet Oncol* 2006; **7**(2): 119-126.
- [3] Grosse Y, Baan R, Straif K, Secretan B, El GF, Bouvard V *et al.* Carcinogenicity of 1,3-butadiene, ethylene oxide, vinyl chloride, vinyl fluoride, and vinyl bromide. *Lancet Oncol* 2007; **8**(8): 679-680.
- [4] Vlaanderen *et al.* Application of new biomarker technologies in occupational and environmental health research; current status and projections. *OEH* 2008; (In press).
- [5] Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 2008; **453**(7193): 396-400.
- [6] Vineis P, Hoek G, Krzyzanowski M, Vigna-Taglianti F, Veglia F, Airoidi L *et al.* Lung cancers attributable to environmental tobacco smoke and air pollution in non-smokers in different European countries: a prospective study. *Environ Health* 2007; **6**: 7.
- [7] Beelen R, Hoek G, van den Brandt PA, Goldbohm RA, Fischer P, Schouten LJ *et al.* Long-term exposure to traffic-related air pollution and lung cancer risk. *Epidemiology* 2008; **19**(5): 702-710.
- [8] Vineis P, Alavanja M, Garte S. Dose-response relationship in tobacco-related cancers of bladder and lung: a biochemical interpretation. *Int J Cancer* 2004; **108**(1): 2-7.

- [9] Rothman KJ, Greenland S. Modern Epidemiology. Lippincott Williams & Wilkins; 3 edition (January 15, 1998), 1998.
- [10] Vineis P, Kriebel D. Causal models in epidemiology: past inheritance and genetic future. *Environ Health* 2006; **5**: 21.
- [11] Anderson HR, Atkinson RW, Bremner SA, Marston L. Particulate air pollution and hospital admissions for cardiorespiratory diseases: are the elderly at greater risk? *Eur Respir J Suppl* 2003; **40**: 39s-46s.
- [12] Vineis P, Airolidi L, Veglia F, Olgiati L, Pastorelli R, Autrup H *et al*. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ* 2005; **330**(7486): 277.
- [13] Vineis P, Bartsch H, Caporaso N, Harrington AM, Kadlubar FF, Landi MT *et al*. Genetically based N-acetyltransferase metabolic polymorphism and low-level environmental exposure to carcinogens. *Nature* 1994; **369**(6476): 154-156.
- [14] Case RA, Hosker ME, McDonald DB, Pearson JT. Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. I. The role of aniline, benzidine, alpha-naphthylamine, and beta-naphthylamine. *Br J Ind Med* 1954; **11**(2): 75-104.
- [15] Schatzkin A, Freedman LS, Schiffman MH, Dawsey SM. Validation of intermediate end points in cancer research. *J Natl Cancer Inst* 1990; **82**(22): 1746-1752.
- [16] Schatzkin A, Gail M. The promise and peril of surrogate end points in cancer research. *Nat Rev Cancer* 2002; **2**(1): 19-27.

- [17] Lan Q, Zhang L, Li G, Vermeulen R, Weinberg RS, Dosemeci M *et al.* Hematotoxicity in workers exposed to low levels of benzene. *Science* 2004; **306**(5702): 1774-1776.
- [18] Thomas DC. High-volume "-omics" technologies and the future of molecular epidemiology. *Epidemiology* 2006; **17**(5): 490-491.
- [19] Smith MT, Vermeulen R, Li G, Zhang L, Lan Q, Hubbard AE *et al.* Use of 'Omic' technologies to study humans exposed to benzene. *Chem Biol Interact* 2005; **153-154**: 123-127.
- [20] Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. *Nature* 2000; **408**(6809): 239-247.
- [21] Berra E, Ginouves A, Pouyssegur J. The hypoxia-inducible-factor hydroxylases bring fresh air into hypoxia signalling. *EMBO Rep* 2006; **7**(1): 41-45.
- [22] Chellappan SP. HOG on the promoter: regulation of the osmotic stress response. *Sci STKE* 2001; **93**, PE1.
- [23] Hunter DJ. Genomics and proteomics in epidemiology: treasure trove or "high-tech stamp collecting"? *Epidemiology* 2006; **17**(5): 487-489.
- [24] Applebaum KM, Karagas MR, Hunter DJ, Catalano PJ, Byler SH, Morris S *et al.* Polymorphisms in nucleotide excision repair genes, arsenic exposure, and non-melanoma skin cancer in New Hampshire. *Environ Health Perspect* 2007; **115**(8): 1231-1236.
- [25] Wang Z, Neuburg D, Li C, Su L, Kim JY, Chen JC *et al.* Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ Health Perspect* 2005; **113**(2): 233-241.

- [26] Lu T, Liu J, LeCluyse EL, Zhou YS, Cheng ML, Waalkes MP. Application of cDNA microarray to the study of arsenic-induced liver diseases in the population of Guizhou, China. *Toxicol Sci* 2001; **59**(1): 185-192.
- [27] Callinan PA, Feinberg AP. The emerging science of epigenomics. *Hum Mol Genet* 2006; **15 Spec No 1**: R95-101.
- [28] Bird A. Perceptions of epigenetics. *Nature* 2007; **447**(7143): 396-398.
- [29] Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004; **429**(6990): 457-463.
- [30] Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 2007; **8**(4): 253-262.
- [31] Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007; **8**(4): 286-298.
- [32] Liu L, Wylie RC, Andrews LG, Tollefsbol TO. Aging, cancer and nutrition: the DNA methylation connection. *Mech Ageing Dev* 2003; **124**(10-12): 989-998.
- [33] Moss TJ, Wallrath LL. Connections between epigenetic gene silencing and human disease. *Mutat Res* 2007; **618**(1-2): 163-174.
- [34] Gallou-Kabani C, Junien C. Nutritional epigenomics of metabolic syndrome: new perspective against the epidemic. *Diabetes* 2005; **54**(7): 1899-1906.
- [35] Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007; **128**(4): 683-692.
- [36] Committee on Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment NRC. Application to Exposure Assessment. In Board on Environmental

Studies and Toxicology, Board on Life Sciences eds. *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*. The National Academies Press 2007: 59-72.

- [37] Joo WA, Sul D, Lee DY, Lee E, Kim CW. Proteomic analysis of plasma proteins of workers exposed to benzene. *Mutat Res* 2004; **558**(1-2): 35-44.
- [38] Vermeulen R, Lan Q, Zhang L, Gunn L, McCarthy D, Woodbury RL *et al*. Decreased levels of CXC-chemokines in serum of benzene-exposed workers identified by array-based proteomics. *Proc Natl Acad Sci U S A* 2005; **102**(47): 17041-17046.
- [39] Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998; **16**(9): 373-378.
- [40] Waters NJ, Holmes E, Williams A, Waterfield CJ, Farrant RD, Nicholson JK. NMR and pattern recognition studies on the time-related metabolic effects of alpha-naphthylisothiocyanate on liver, urine, and plasma in the rat: an integrative metabonomic approach. *Chem Res Toxicol* 2001; **14**(10): 1401-1412.
- [41] Waters NJ, Holmes E, Waterfield CJ, Farrant RD, Nicholson JK. NMR and pattern recognition studies on liver extracts and intact livers from rats treated with alpha-naphthylisothiocyanate. *Biochem Pharmacol* 2002; **64**(1): 67-77.
- [42] Odunsi K, Wollman RM, Ambrosone CB, Hutson A, McCann SE, Tammela J *et al*. Detection of epithelial ovarian cancer using ¹H-NMR-based metabonomics. *Int J Cancer* 2005; **113**(5): 782-788.
- [43] Beckonert O, Monnerjahn J, Bonk U, Leibfritz D. Visualizing metabolic changes in breast-cancer tissue using ¹H-NMR spectroscopy and self-organizing maps. *NMR Biomed* 2003; **16**(1): 1-11.

Figure 1 Cumulative risk of lung cancer by rs8034191 genotype. Relevance of smoking and of rs8034191 genotype to lung cancer mortality in men aged 45-75 years. Cumulative risk (in the absence of other causes of death) based on national lung cancer death rates for men in Poland in the year 2000, assuming that the prevalence of current smoking, former smoking, and never smoking are as in this study and that the relative risks for lung cancer incidence and mortality are similar (Hung *et al* 2008) [1].

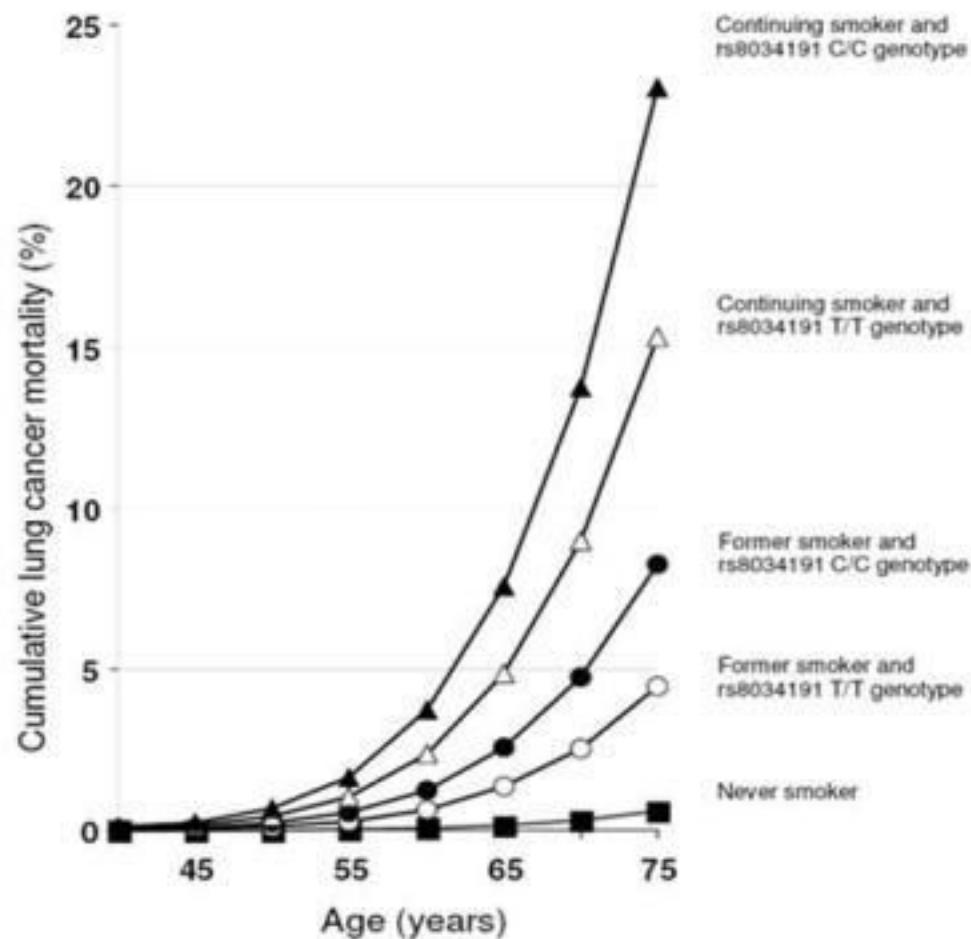
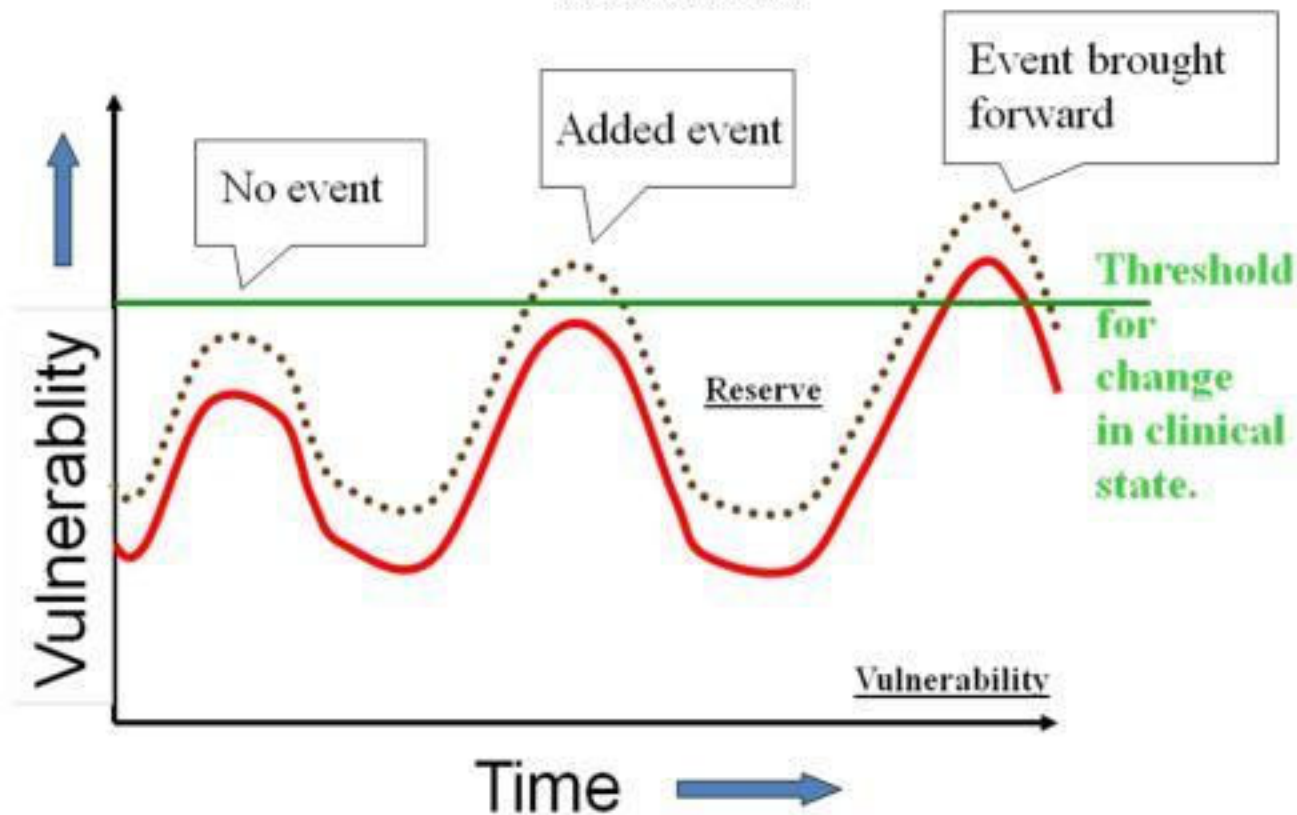


Figure 2. Vulnerability plus exposure events change the physiological state, but a reserve is present. When the reserve is overtaken, clinical manifestations appear. From http://www.sahsu.org/jubilee_presentations/Anderson_ppt#392_21, courtesy of R Anderson.



Interaction between air pollution and clinical vulnerability

Additional files provided with this submission:

Additional file 1: table 1.doc, 38K

<http://www.ehjournal.net/imedia/3804458526611907/supp1.doc>

Additional file 2: box1.doc, 25K

<http://www.ehjournal.net/imedia/1164901766266118/supp2.doc>