

REVIEW

Environmental exposure measurement in cancer epidemiology

Christopher P. Wild

Molecular Epidemiology Unit, Centre for Epidemiology and Biostatistics,
Leeds Institute of Genetics, Health and Therapeutics, LIGHT Laboratories,
University of Leeds, Clarendon Way, Leeds LS2 9JT, UK

Environmental exposures, used in the broadest sense of lifestyle, infections, radiation, natural and man-made chemicals and occupation, are a major cause of human cancer. However, the precise contribution of specific risk factors and their interaction, both with each other and with genotype, continues to be difficult to elucidate. This is partially due to limitations in accurately measuring exposure with the subsequent risk of misclassification. One of the primary challenges of molecular cancer epidemiology therefore is to improve exposure assessment. Progress has been made with biomarkers such as carcinogens and their metabolites, DNA and protein adducts and mutations measured in various tissues and body fluids. Nevertheless, much remains to be accomplished in order to establish aetiology and provide the evidence base for public health decisions. This review considers some of the principles behind the application of exposure biomarkers in cancer epidemiology. It also demonstrates how the same biomarkers can contribute both to establishing the biological plausibility of associations between exposure and disease and be valuable endpoints in intervention studies. The potential of new technologies such as transcriptomics, proteomics and metabolomics to provide a step change in environmental exposure assessment is discussed. An increasing recognition of the role of epigenetic changes in carcinogenesis presents a fresh challenge as alterations in DNA methylation, histone modification and microRNA in response to environmental exposures demand a new generation of exposure biomarker. The overall importance of this area of research is brought into sharp relief by the large prospective cohort studies (e.g. UK Biobank) which need accurate exposure measurement in order to shed light on the complex gene:environment interactions underlying common chronic disorders including cancer. It is suggested that a concerted effort is now required, with appropriate funding, to develop and validate the required exposure assessment methodology before these cohorts come to maturity.

Introduction

The magnitude by which the environment contributes to the aetiology of human cancer is disputed, with recent estimates ranging from 1 to 19% (1–3). As these authors have discussed,

the differences in part depend on definitions of the word ‘environment’. In the current review, I will refer to the environment in the broader sense of lifestyle (including diet), infections, radiation, natural and man-made chemicals and occupational exposures. The subtleties of definition, while to some appearing esoteric, are important as the conclusions drawn may influence funding bodies and drive political priorities for research and public health, especially when espoused by the World Health Organization and its specialized cancer agency, the International Agency for Research on Cancer. Consequently, it is vitally important to have an accurate understanding of the global contribution of the environment to human cancer and to interpret those data in an informative manner for policy makers.

Necessarily with the aforementioned global estimates of cancer risk, there is an aggregation and simplification of information. Underlying the headline figures is a complexity that demands consideration in relation to cancer prevention. While the gross estimates can provide general guidance, for example towards the importance of understanding diet and cancer risk, the underpinning detail may reveal a paucity of reliable information of a more specific kind. First, the broad categories of environmental factors under discussion are themselves heterogeneous. Environmental pollution, for example, includes pesticides, wood burning fires, diesel exhaust emissions, industrial waste, etc. The prevalence and level of each exposure will vary markedly in different parts of the world and each may be relevant to cancers at specific sites. The magnitude of risk related to pesticides and Hodgkin’s lymphoma would not be the same as for diesel exhaust and lung cancer. Second, the extent and quality of epidemiological data vary markedly in each specific case. The resulting gaps in knowledge are problematic because it is not general estimates of the role of the environment in cancer causation that translate to effective cancer prevention at a national or international level, rather it is the specifics, relating particular exposures to particular risks that best inform decisions.

In a genomic era of high excitement about the influence of genes on common chronic diseases, it has never been more timely to keep in mind the seminal observations from epidemiology that point to such a strong influence of the environment on cancer occurrence, including the marked geographic variations in incidence, changes in incidence over relatively short periods of time and the altered patterns of disease in migrant populations (4). Genome-wide association studies, performed on thousands of individuals, are certainly yielding exciting findings. A recent example, pertinent to common chronic disorders, reported that common variants near the melanocortin-4 receptor gene are strongly associated with fat mass and obesity (5). However, in this specific case these variants made an estimated contribution of just ~0.14% to the overall variance in adult body:mass index. The value of such studies to public health is clearly through providing novel

To whom correspondence should be addressed. Tel: +44 113 3436601; Fax: +44 113 3436603; Email: c.p.wild@leeds.ac.uk

insights into how the environment acts on the relevant biological pathways, rather than through personalized risk assessment and medicine, even in developed countries.

The majority of cancers have a complex aetiology where one or more environmental risk factors interact with genetic background, age, sex, socio-demographic status and other factors. However, the precise contribution of individual factors and their interaction, both with each other and with genotype, continues to be difficult to elucidate. This is partially due to the challenges inherent in accurately measuring exposure. Misclassification in exposure assessment introduces uncertainty and limits the power of epidemiological studies. This is particularly damaging when the true strength of the association between exposure and disease is modest. In these cases, misclassification may blur or completely obscure underlying causal associations. Many environmental exposures of interest occur at low levels but can be ubiquitous, posing further challenges to accurate exposure assessment.

In response to the above needs, molecular cancer epidemiology promised to provide biomarkers to refine exposure assessment. Among the potential advantages of exposure, biomarkers are the provision of a more objective measure at the individual level; a relevant measure in relation to events on, or related to, the causal pathway; information relevant to the biological plausibility of an exposure-disease association and an ability to detect low levels of exposure using sensitive laboratory technology. While progress has undoubtedly been made, the accurate assessment of many environmental exposures has remained an outstanding challenge with a need to improve assessment of the 'exposome' to complement the genome (6,7). Huge investment is being made in large prospective cohort studies that include the expensive collection and banking of biological material (8–10). These cohorts require improved approaches to exposure assessment, including biomarkers. A structured approach to this requirement by major funding agencies is needed now while there is still a decade or so of grace, before the cohorts mature in terms of numbers of cases, during which time progress can be made in the development and validation of specific exposure biomarkers.

This review will briefly consider the current status of exposure biomarkers in molecular cancer epidemiology and consider where new challenges lie in order to help move the field forward. While the focus of the chapter is on biomarkers, it is important to acknowledge that refinement of exposure assessment is also being addressed by complementary approaches, including geographic information systems, personal and environmental monitoring and increasingly sophisticated questionnaires (11,12); it is eventually a combination of tools which is most likely to provide the answers required.

Exposure biomarkers and mechanisms of carcinogenesis

A notable feature of much of the literature to date on exposure biomarkers in cancer epidemiology is the emphasis on compounds that damage DNA, with the associated measurement of DNA adducts and mutations. One of the most exciting future challenges for molecular epidemiology is the development of an analogous set of biomarkers for exposures acting through other mechanisms of carcinogenesis. This follows the recognition that environmental exposures can alter gene expression not only by mutation but also by epigenetic mechanisms (13). It is critical that the rapid advance in

understanding of epigenetic mechanisms (14,15) is matched by translation of this knowledge into biomarkers applicable to population-based studies of cancer aetiology. This field is particularly compelling because of the recognition that a number of epigenetic events are reversible. Consequently, this represents an opportunity not only for therapy, currently the primary focus of discussion (15), but also for cancer prevention perhaps through the targeting of pre-cancerous lesions.

Some of the major epigenetic mechanisms of interest are listed in Table I. Methylation of CpG promoter sequences and the modification of histones are common findings in human cancers (16,17). The measurement of CpG hypermethylation (18,19) and global methylation (20) in biological fluids and the observation that environmental exposures such as tobacco smoking are associated with altered methylation patterns (21) offer an opportunity to molecular epidemiology in a similar way to the detection of mutated circulating DNA sequences in peripheral blood (22). The close mechanistic link between DNA methylation and histone modification, for example through methyl-CpG-binding domain proteins, opens up a further important area of research where molecular epidemiologists have opportunities to consider the role of the environment in modifying these processes (16). In addition, the role of microRNAs, which inhibit expression of specific target genes (23), also merits consideration in terms of environmental exposures.

Molecular epidemiology must begin to address the question as to how exposures, such as diet, obesity, physical exercise, environmental chemicals, etc., influence methylation, histone modification or other epigenetic processes to alter cell function and cancer risk (24). In an even broader context, the influence of exposures on other relevant mechanisms, for example receptor binding (25) should be considered. It is possible that the discoveries from 'omics' technologies (see below) will be the catalyst for identification of further relevant non-genotoxic carcinogenic pathways and will thus stimulate the development of associated biomarkers in these areas.

Categories of exposure biomarker

A number of different categories of exposure biomarker have been identified and some of these are described in this section and summarized in Table II.

Environmental chemicals or their metabolites in human tissues, body fluids or exhaled air have been used to measure exposure and are termed markers of 'internal dose'. Examples include polycyclic aromatic hydrocarbons (PAH) (26), heterocyclic amines (27), organochlorines (28), hormones or nutrients in body fluids (29,30) and organic chemicals in exhaled air (31). In terms of biomarkers of diet, a distinction has been made between recovery biomarkers, based on an

Table I. Epigenetic mechanisms relevant for biomarker development

DNA methylation
Global hypomethylation
Hypermethylation of promoter island CpG sites
Histone modification
Chromatin remodelling
Small non-coding RNAs (microRNAs)
Receptor binding

Table II. Categories of exposure biomarker [adapted from (96)]

Biomarker type	Biomarker category	Examples
Carcinogens and their metabolites	Internal dose	1-Hydroxypyrene, aflatoxins and tobacco-specific nitrosamines in urine; faecal <i>N</i> -nitrosocompounds; organochlorines in serum and adipose tissue; metals in nails or hair
Nutrients	Internal dose	Urinary sugars and flavonoids; serum vitamins
Circulating antibodies	Internal dose	Antibodies to viruses (e.g. HBV), parasites (e.g. <i>Opisthorchis viverrini</i>) or bacteria (e.g. <i>Helicobacter pylori</i>)
DNA adducts	Biologically effective dose	Urinary AFB1-N7-gua, bulky DNA adducts, DNA strand breaks (comet assay)
Protein adducts	Biologically effective dose	Aminobiphenyl-hemoglobin, aflatoxin–albumin
Genetic alterations	Biologically effective dose or early effect	Chromosomal aberrations, micronuclei, somatic cell mutations in reporter genes (HPRT, glycophorin A) or tumor suppressor genes (TP53 mutation spectra)
Circulating proteins	Biologically effective dose or early effect	Altered growth factors (e.g. IGF-1, IGF1Bs) and cytokines (e.g. IL-6)
Altered gene expression	Biologically effective dose or early effect	Lymphocyte CYP1A1 expression; transcriptomics

HBV, hepatitis B virus; HPRT, hypoxanthine-guanine phosphoribosyltransferase; IGF, insulin-like growth factor; IGF1Bs, insulin-like growth factor binding proteins.

understanding of the balance between intake and excretion where the biomarker level translates to an estimate of intake over a given time period e.g. 24-h urinary nitrogen, and concentration biomarkers which are not time related and do not permit estimates of absolute intakes (32). The influence of metabolism and other factors such as absorption on the level of circulating biomarker means there may not always be a simple relationship between external exposure and internal dose. Indeed, this property has been exploited to use the measurement of metabolite ratios successfully as an approach to metabolic phenotyping, for example in the case of PAH and urinary phenanthrene metabolites (33). Some of the most informative internal dose biomarkers come from the area of infections and cancer (e.g. hepatitis viruses, human papilloma viruses, *Helicobacter pylori*) where the presence of pathogen proteins, or antibodies to these proteins, indicate that exposure has occurred.

DNA and protein adducts have been widely measured in human biological samples (34,35) and these have been referred to as biomarkers of ‘biologically effective dose’ i.e. a measure of the amount of the carcinogen reaching the critical cellular target. Adduct levels provide an integration of exposure, absorption, distribution, metabolism and, in the case of DNA adducts, DNA repair. Some assays are chemical specific while others, such as the ³²P post-labelling or the Comet assay, reflect general levels of DNA damage although chromatographic purification techniques or DNA repair enzymes can provide more specificity. Other markers of general DNA damage include chromosomal aberrations, micronuclei and sister chromatid exchanges, which can be induced by a wide range of exposures, reflecting cumulative exposure to a variety of environmental factors. It could be argued that the absence of an association with a specific exposure means these latter types of biomarker should not be considered as indicators of biologically effective dose. However, this is more a question of limited specificity than of categorization. For example, DNA strand breaks undoubtedly would be a critical lesion resulting from exposure to gamma irradiation and hence represent the biologically effective dose, even if the total burden of strand breaks could not be assigned exclusively to that exposure.

As mentioned above, many carcinogens act other than by damaging DNA and in these cases the biologically effective dose may be the amount of the chemical exposure bound to

a cellular receptor, an altered enzyme activity or its consequences. For example, the mycotoxin, fumonisin B1 classified by the IARC as ‘possibly carcinogenic to humans’ (36), appears to act through its inhibition of ceramide synthase and sphingolipid biosynthesis (37). In this case, the biomarker of biologically effective dose might be the alteration in the ratio of sphingolipid precursors.

Somatic mutations, either in reporter genes (38,39) or in cancer-related genes (40), have also been used to investigate human exposure to environmental agents. Again, here the classification by category of biomarker is less clear and these may also be referred to as biomarkers of ‘early biological effect’ together with the genetic alterations mentioned above. Mutation analyses may be limited to the mutation frequency or may also investigate the pattern of mutations in order to infer something more specific about their environmental origin. Altered gene, protein or metabolite expression mentioned in the previous section may prove to be informative about both the exposure and the biological effects of those exposures.

Sensitivity and specificity of exposure biomarkers

Many environmental exposures of interest occur at low levels and the exquisite sensitivity of some of the laboratory assays applied to exposure assessment can enable detection and quantification at these low levels. A good example of the success in this area is with environmental tobacco smoke where biomarkers clearly demonstrated that exposure occurs and that it exerts biological effects at environmental levels (41,42).

Specificity is another important consideration in selecting an exposure biomarker. For example, an assay of a DNA adduct will be specific for the target chemical. However, while some chemicals will equate to the environmental exposure that is of interest to the epidemiologist, some may not. For example, urinary aflatoxin–DNA adducts are a specific measure for dietary exposure to aflatoxins. However, with benzo(a)pyrene [B(a)P]–DNA adducts, the B(a)P may originate from any one of a number of environmental exposures e.g. tobacco smoke, diet, air pollution or certain occupational exposures (43). This is also the case for general markers of DNA damage, such as chromosomal aberrations, where the biomarker cannot be simply related to a specific exposure.

Validation of exposure biomarkers

Validity has been defined as the (relative) lack of systematic measurement error when comparing the actual observation with a standard (reference) method, which represents the 'truth' (44). The lack of biomarker validity may therefore result not only from analytical error in the laboratory but also from other underlying factors that distort the relationship between external exposure and biomarker including lack of dose-response relationship between exposure and biomarker, failure of the biomarker to integrate exposure over the time period of interest or inter-individual variability in the exposure-biomarker relationship. Examples include urinary aflatoxin P1, which does not quantitatively reflect aflatoxin exposure (45) and plasma ochratoxin A levels which did not correlate with ochratoxin A intake (46). Empirical data are therefore required to establish the validity of a biomarker in exposed people, but unfortunately such evidence is not always obtained prior to application in epidemiological studies.

Exposure biomarkers and surrogate tissues

The interpretation of an exposure biomarker measurement will depend not only on the inherent nature of the biomarker but also in what biological medium it is measured. For example, a DNA adduct in peripheral blood cells will provide information that may differ qualitatively and quantitatively from that of the same DNA adduct measured in urine. While the critical level of a DNA adduct may occur at the tissue, cell, gene and even DNA sequence level, it is rarely possible to make measurements even in the target organ in relation to cancer risk. Instead, the measures are made in accessible samples, typically plasma or serum, white blood cells or urine. The potential disconnection between what would be desirable and what is possible should be considered when interpreting the information gained from such surrogate molecules or tissues.

In addition to where the biomarker is measured, the timing of measurement is important to interpretation. Environmental exposures will vary qualitatively and quantitatively over time due to changes in lifestyle, place of residence, income, occupation, etc., and the impact of a given exposure on cancer risk may not be consistent across the lifespan of the individual (47,48). This recognition coupled with the knowledge that biomarkers are inherently transient in nature poses further challenges to the appropriate design of epidemiological studies incorporating exposure biomarkers.

Past exposure assessment to infectious agents has been successful due to the immunological fingerprint left by circulating antibodies to pathogen proteins. However, for chemical exposures, with some notable exceptions such as chemicals in adipose tissue or nail clippings for example (49,50), this has not been the case. In the majority of instances, the biomarker half-life for a chemical exposure is expected to be relatively short (days to months), although in reality the instances of this being established empirically in humans are few. Potential solutions include the search for longer lived adducts, for example on histone proteins (51) or mutational fingerprints, particularly when measured in plasma (52,53).

The short-term nature of the majority of the currently available exposure biomarkers makes them largely inapplicable to the case:control study design, not only because they may not reflect past exposure but because concurrent disease could

influence the biomarker level. Despite this risk of reverse causation, some adducts have been applied in case:control studies with positive findings (54-56).

Current exposure biomarkers are generally better suited to application in prospective cohort studies. A nested case:control study within the cohort or a case:cohort design (57) limits the resources needed for biomarker analysis. Recruitment and biological sampling from healthy individuals at entry to the study help avoid the problems of reverse causation. The design also provides an opportunity for repeat exposure measures and evaluation of the intra-individual variation, although often in practice the periods of follow-up tend to be short (a few years rather than decades) and biological sampling is only performed at a single time point (most often recruitment) due to logistic and financial constraints. Repeat measures may be possible in the newer and better-funded cohort studies. It is within the prospective cohort design that the most successful examples of biomarkers and disease outcome are found (45,58-60).

Biomarkers of exposure can provide far more to the field of cancer epidemiology than improvement in exposure assessment. In the following sections, examples of contributions to biological plausibility and intervention strategies are briefly discussed. Other valuable applications not considered here include the potential to permit extrapolation of data from animals to humans (61) and the application to human biomonitoring or biosurveillance (62,63).

Biological plausibility

One of the criteria for establishing a causal association between an exposure and disease is biological plausibility. In this context, biomarkers may contribute by illuminating some of the carcinogenic steps linked to a particular risk factor. This is possibly an undervalued area where biomarkers can make significant contributions to cancer epidemiology.

If a particular chemical exposure from ambient air is associated with increased risk, the additional information that exposed individuals have higher levels of DNA damage would add support to the exposure-disease association (42). If genetic polymorphisms in carcinogen metabolizing or DNA repair enzymes are associated with both an increased cancer risk and higher levels of a biomarker on the presumed causal pathway e.g. DNA adducts, this would provide support for the original association (64,65). Some questions of biological plausibility can be addressed in human experimental studies. For example, subjects fed diets high in red meat had higher levels of *N*-nitroso compounds (NOC), and NOC-related DNA adducts in exfoliated cells, in faeces (66) providing a potential mechanism for the observed association between red meat consumption and colorectal cancer risk. These types of mechanistic data are increasingly being considered in the processes of hazard identification and cancer risk assessment and will be of increasing relevance to these processes as epidemiology struggles to link exposure to disease where increased risks in the population are modest.

Intervention studies

Another valuable application of exposure biomarkers is in evaluating the potential of intervention strategies (67). This may be primary prevention to reduce exposure or more mechanism-based approaches such as chemoprevention. In either case, biomarkers can be used as endpoints, permitting

a proof of principle to be established in advance of long-term interventions where pre-cancerous lesions or cancer itself might be the outcome. The proof of principle stage may also provide valuable information pertinent to the design of interventions, for example in establishing appropriate doses of chemopreventive agents.

One type of intervention study aims to modulate a particular biochemical pathway, perhaps using a micro-nutrient or pharmaceutical agent. The outcome is a better understanding *in vivo* of mechanisms of carcinogenesis and a stronger scientific rationale for interventions. For example, Collins *et al.* (68) reported that kiwi fruit consumption reduced both endogenous oxidative DNA damage and damage induced by an *ex vivo* challenge in peripheral lymphocytes, as well as enhancing DNA repair. An elegant series of chemoprevention studies in China using oltipraz and chlorophyllin have also demonstrated the modulation of aflatoxin metabolism in exposed individuals (69–72). Alternatively, exposure biomarkers may be used as endpoints in primary prevention studies. In a community-based post-harvest primary prevention trial, targeted at the groundnut crop in Guinea, West Africa (73), aflatoxin–albumin adduct levels were >50% lower in subjects in the intervention villages. Other examples come from quit smoking studies where a panel of biomarkers can be used to monitor the success of the intervention (74).

Transcriptomics, proteomics, metabonomics and the future opportunities for exposure biomarkers

The application of transcriptomics, proteomics and metabonomics to cancer research may contribute not only to understanding mechanisms of carcinogenesis but also potentially to the development of a new generation of biomarkers of exposure and early effect (75). It may prove to be that the technologies themselves are not applied routinely to large numbers of subjects, although metabonomics may be an exception (see below), but these approaches may highlight novel responses to exposure within particular biological pathways, which would then provide more specific targets for assessment in relation to exposure.

The value of these technologies will primarily depend on whether specific environmental exposures are indeed reflected in the human body by altered levels of specific mRNA, proteins or metabolites. Will distinct signatures or fingerprints of environmental exposures be found across a broad spectrum of mechanisms of action? If so these new technologies may be in a position to permit a step change in the development of biomarkers of both exposure and effect.

There are some early indications that this is a fruitful area of research. A number of environmental and occupational exposures have been explored in preliminary studies in a population setting (see Table III) (50,76–84). To date, these are characterized by relatively small numbers of subjects, partially reflecting the initial high cost of the microarrays and frequently a lack of consideration of the full range of potential confounding factors. In addition, the fold changes in gene expression have tended to be modest (<2-fold). The differences in scope of genes represented on the different types of array, their reproducibility and confirmation by independent techniques (e.g. reverse transcriptase-polymerase chain reaction) differ across the reports. However, the increasingly reproducible performance of commercial arrays and the

sophisticated data analysis in terms of identification of alterations in pathway-related genes is starting to prove valuable.

One of the most informative examples to date has been the study of arsenic exposure in pregnant women and the effect on gene expression in cord blood samples from their children (50). In this study, a small number of only 11 genes permitted correct prediction of the arsenic exposure status of the mother, suggesting these not only as promising candidates for the prenatal response to arsenic but also as biomarkers of exposure. Of the total 447 arsenic-modulated gene transcripts, 105 encoded proteins were identified as being part of an interacting network, with three significant sub-networks. These findings start to provide exciting functional information about, in this case a prenatal, environmental exposure to a known carcinogen and altered gene expression.

An additional study of great interest, because it sampled from the target cells, was that by Spira *et al.* (84) where the gene expression in bronchial epithelium cells was compared in smokers, ex-smokers and non-smokers. Genes in a number of categories, notably oxidant stress and glutathione metabolism, xenobiotic metabolism and secretion, differed between the groups.

Overall, these preliminary transcriptomic data indicate that environmental exposures elicit changes in gene expression and that the nature of the changes varies depending on the type of exposure. This encourages further exploration of the sensitivity, specificity and stability of these changes. The analytical reproducibility of transcriptomic approaches also needs to be considered. How distinct the eventual patterns of gene expression will prove to be across exposures remains to be determined, with some similarities already emerging, perhaps unsurprisingly, in terms of inflammation, oxidative stress, cell proliferation and apoptosis with several different exposures. This will be where greater understanding of the epigenetic effects of environmental exposures discussed earlier in this paper will prove valuable.

To date, proteomics has been less applied to exposure assessment than to the clinical detection of cancer or of pre-cancerous lesions. An exception, however, was the application of plasma proteomics to the study of print workers exposed to benzene (85). Examination of the 3,000 or so major metabolites that constitute the metabolome also offers opportunities to address exposure assessment. Urinary metabolite fingerprints were obtained in mice infected with *Schistosoma mansoni* (86) and in a nutritional study in people, a change from non-soy to soy-containing diet was associated with changes in the plasma metabolome (87). More recently, a major metabonomic study of 4630 subjects by ¹H nuclear magnetic resonance (88) demonstrated marked differences in urinary metabolite profiles by population (China, Japan, UK and USA) with an interesting observation being that Japanese resident in Japan had different profiles than those living in America. Different diets were also associated with different phenotypes and furthermore some specific metabolites, notably formate, alanine (positively) and hippurate (negatively) were associated with blood pressure. Finally, the observed similarity between two 24-h urine samples from the same individual collected on average 3 weeks apart suggests that intra-individual variations are less significant than some of the demographic and dietary effects mentioned above. The potential application of omics technologies to characterize dietary exposures and to understand the biological effects of diet at the cellular level (89,90) therefore receives some support from these early investigations.

Table III. Examples of environmental and occupational exposures and altered global gene expression

Exposure	Location	Subjects	Samples	Altered gene expression data	Gene networks or categories affected	Reference
Smoking	Washington State, USA	Cigarette smokers ^a (32 EX, 33 UNEX)	PBMC	861 probe sets differentially expressed, 36 top candidates: sensitivity 90%, specificity 100% to identify smokers	Various—including immune function and inflammation	(81)
Smoking	East Flanders, The Netherlands	Cigarette smokers ^b (9EX, 9UNEX)	Whole blood	34 differentially expressed genes between twins discordant for smoking status and 76 between smokers and non-smokers	Various—including ATF4, MAPK14, SOD2 confirmed by RT-PCR	(82)
Benzene	Tianjin, China	Shoe factory workers ^c (6EX, 6 UNEX)	PBMC	2129 probe sets differentially expressed, 19/508 cytokine genes differentially expressed	Various—CXCL16, ZNF331, JUN and PF4 confirmed by RT-PCR in larger subject group	(77)
Arsenic	Lanyang Basin, Taiwan	Subjects consuming arsenic-tainted well water ($N = 24$) ^d	PBL	62 cDNA clones with altered expression in high/ intermediate versus low	Cytokine and growth related, signal transduction, transcription machinery, cell cycle control	(78)
Arsenic	Ron Pibul and Bangkok, Thailand	Newborns (23 EX, 11 UNEX) ^e	Whole cord blood	11 genes predicted mother's exposure status with 83% accuracy	Stress response, cell cycle regulation, enriched for genes with NF-kB and serum response factor binding sites	(50)
Metal fumes	Massachusetts, USA	Welders (15 EX, 7 UNEX)—compared pre- and post-shift ^f	Whole blood	139 genes altered post-shift in at least 50% of arrays among welders	Pro-inflammatory and immune responses, oxidative stress, phosphate metabolism, cell proliferation, apoptosis	(79)
Air pollution	Teplice and Prachatice, Czech Republic	Children (23 EX, 24 UNEX) and parents (12 EX, 12 UNEX) ^g	Whole blood	471 differentially expressed genes in children and 140 in parents ($P < 0.01$)	In children only: nucleosome, immune response and IFN-associated motif groups	(83)
Cigarette smoke	Massachusetts, USA	Adults (34 EX, 41UNEX) ^h	Bronchial epithelium	97 genes differentially expressed between never and current smokers	Various—oxidative stress and glutathione metabolism, xenobiotic metabolism and secretion	(84)

PBMC, peripheral blood mononuclear cells; PBL, peripheral blood lymphocytes; EX, exposed; UNEX, unexposed; ATF4, activating transcription factor 4; MAPK14, mitogen-activated protein kinase 14; SOD, superoxide dismutase; IFN, interferon.

^aExposed group: >10 cigarettes per day, plasma cotinine >30 ng/ml, hu25k arrays.

^bNine pairs of monozygotic twins discordant for smoking, cigarette smoking (mean ~15 cigarettes per day) confirmed by plasma cotinine, Human Tox 66 cDNA, Phase-1 Molecular Toxicology.

^cExposed group: mean benzene 47.3 ± 24.3 ppm, personal air monitoring, U133A/B Affymetrix.

^dTwenty-four subjects tested for blood arsenic in three groups of eight at low (0–4.32 $\mu\text{g/l}$), intermediate (4.64–9.0 $\mu\text{g/l}$) and high (9.6–46.5 $\mu\text{g/l}$) but RNA from all eight subjects per group pooled for analysis, custom array with 708 cDNA probes.

^eExposed group: mean arsenic >0.5 $\mu\text{g/g}$ in toenails, HGU133 plus 2.0 Affymetrix.

^fExposed group: median PM_{2.5} 2.44 mg/m^3 , U133A Affymetrix.

^gExposed group: various parameters of air pollution using stationary air quality monitoring stations, Human 1A (V2) 22k Agilent.

^hExposed group: self-reported smokers, ex-smokers and never smokers, HGU133A, Affymetrix.

This new generation of technologies requires further research before the potential to yield a new generation of exposure biomarkers is evaluated. It remains to be seen if mRNA, protein or metabolite expression can be specific and sensitive enough to define exposures at low levels in human populations. It will be important to understand whether complex mixtures or families of chemicals acting on the same pathways affect common targets. As with many other exposure biomarkers, the analysis is being conducted in peripheral blood rather than target organs and the relevance of these changes to the disease pathways will need to be elucidated. In addition, the dynamic nature of each of these systems may militate against long-term exposure assessment, unless some of the changes prove stable over time.

As well as the above biological considerations, the omics technology will need to be tailored in terms of sensitivity, sample requirement, throughput and cost. Purification procedures in the case of metabolomics and proteomics will be important to permit analysis of less abundant but possibly more informative, proteins or metabolites among the background of quantitatively more dominant species. Although not explicitly discussed here, the need for sophisticated statistical analysis emerges as crucial to any eventual application. As with the earlier generation of exposure biomarkers, a carefully planned strategy, starting with model systems and small-scale human studies, is likely to be most successful (45).

Conclusions

This review has considered some general principles in relation to biomarkers of exposure as well as presenting examples of existing and currently explored approaches. There are a number of future priorities which merit highlighting. One is the need for appropriate biomarkers of physical activity, important in a number of major cancers (91,92). As potential mechanisms of carcinogenesis are investigated, biomarkers are needed that reveal what type and frequency of physical exercise have most impact in affecting those mechanisms and cancer risk, such that prevention strategies can be informed by a strong scientific rationale. Similarly, biomarkers have much to contribute to the study of diet, obesity and cancer, including assessment of specific nutrients, dietary patterns and energy balance. Application of biomarkers to European Prospective Investigation into Cancer and Nutrition, for example, continues to hold much promise (93) whilst mechanism-based biomarkers need to emerge from understanding the way in which obesity alters cell function and subsequently cancer risk. Other environmental chemical exposures would also merit from better biomarkers, as would non-ionizing radiation.

The new generation of cohort studies (8,10,94,95) provides the framework to investigate genetic variation, environment, lifestyle and chronic disease for the next two to three decades. These studies represent substantial investment, with UK Biobank for example recruiting half a million adults at a cost of ~£60 million in the initial phase. A considerable part of this cost is driven by the collection and banking of biological material. This investment is at least partially justified on the assumption that biochemical and molecular measures on this material will help resolve important aetiological questions. It should be self-evident that unravelling complex environmental and genetic aetiologies in order to plan effective public health interventions demand that both environmental exposures and genetic variation are reliably measured. In fact, even the effects of genetic variation may be undetected if the amplification of

that effect due to interaction with an environmental risk factor is itself unmeasured. Advances in statistical methods and in bioinformatics in relation to large data sets are also of critical importance in addressing these challenges. The further development, validation and application of biomarkers of exposure in this context are manifestly a critical part of the future of cancer epidemiology in the 21st century.

Funding

National Institute of Environmental Health Sciences, USA (no. ES06052) to C.P.W.

Acknowledgements

Thanks are due to Dr Rodolfo Saracci for his insightful comments and to Margaret Jones for help in preparing the manuscript. This was an invited paper following the award of the UK Molecular Epidemiology Group 10th anniversary medal to C.P.W. The author would like to thank all his colleagues in the Molecular Epidemiology Unit at the University of Leeds and his collaborators worldwide for their scientific interaction and friendship over the period relevant to this award.

Conflict of interest statement: None declared.

References

- Boffetta, P., McLaughlin, J. K., La Vecchia, C., Autier, P. and Boyle, P. (2007) 'Environment' in cancer causation and etiological fraction: limitations and ambiguities. *Carcinogenesis*, **28**, 913–915.
- Saracci, R. and Vineis, P. (2007) Disease proportions attributable to environment. *Environ. Health*, **6**, article no. 38.
- Pruss-Ustun, A. and Corvalan, C. (2006) *Preventing Disease through Health Environments. Towards an Estimate of the Environmental Burden of Disease*. World Health Organisation, Geneva, Switzerland.
- Colditz, G. A., Sellers, T. A. and Trapido, E. (2006) Epidemiology—identifying the causes and preventability of cancer? *Nat. Rev. Cancer*, **6**, 75–83.
- Loos, R. J. F., Lindgren, C. M., Li, S. X. *et al.* (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.*, **40**, 768–775.
- Wild, C. P. (2005) Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 1847–1850.
- Vineis, P. (2004) A self-fulfilling prophecy: are we underestimating the role of the environment in gene-environment interaction research? *Int. J. Epidemiol.*, **33**, 945–946.
- Manolio, T. A., Bailey-Wilson, J. E. and Collins, F. S. (2006) Opinion—genes, environment and the value of prospective cohort studies. *Nat. Rev. Genet.*, **7**, 812–820.
- Palmer, L. J. (2007) UK Biobank: bank on it. *Lancet*, **369**, 1980–1982.
- Potter, J. D. (2004) Toward the last cohort. *Cancer Epidemiol. Biomarkers Prev.*, **13**, 895–897.
- Weis, B. K., Balshaw, D., Barr, J. R. *et al.* (2005) Personalized exposure assessment: promising approaches for human environmental health research. *Environ. Health Perspect.*, **113**, 840–848.
- Nuckols, J. R., Ward, M. H. and Jarup, L. (2004) Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ. Health Perspect.*, **112**, 1007–1015.
- Fraga, M. F., Ballestar, E., Paz, M. F. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA*, **102**, 10604–10609.
- Esteller, M. (2006) The necessity of a human epigenome project. *Carcinogenesis*, **27**, 1121–1125.
- Esteller, M. (2008) Molecular origins of cancer: epigenetics in cancer. *N. Engl. J. Med.*, **358**, 1148–1159.
- Lopez-Serra, L. and Esteller, M. (2008) Proteins that bind methylated DNA and human cancer: reading the wrong words. *Br. J. Cancer*, **98**, 1881–1885.
- Fraga, M. F., Ballestar, E., Villar-Garea, A. *et al.* (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.*, **37**, 391–400.
- Carvalho, A. L., Jeronimo, C., Kim, M. M. *et al.* (2008) Evaluation of promoter hypermethylation detection in body fluids as a screening/

- diagnosis tool for head and neck squamous cell carcinoma. *Clin. Cancer Res.*, **14**, 97–107.
19. Hsu, H. S., Chen, T. P., Hung, C. H. *et al.* (2007) Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma. *Cancer*, **110**, 2019–2026.
 20. Hsiung, D. T., Marsit, C. J., Houseman, E. A. *et al.* (2007) Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 108–114.
 21. Russo, A. L., Thiagalingam, A., Pan, H. J. *et al.* (2005) Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. *Clin. Cancer Res.*, **11**, 2466–2470.
 22. Gormally, E., Caboux, E., Vineis, P. and Hainaut, P. (2007) Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. *Mutat. Res.*, **635**, 105–117.
 23. Calin, G. A. and Croce, C. M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
 24. Feil, R. (2006) Environmental and nutritional effects on the epigenetic regulation of genes. *Mutat. Res.*, **600**, 46–57.
 25. Denison, M. S. and Nagy, S. R. (2003) Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.*, **43**, 309–334.
 26. Kang, D. H., Rothman, N., Poirier, M. C. *et al.* (1995) Interindividual differences in the concentration of 1-hydroxypyrene-glucuronide in urine and polycyclic aromatic hydrocarbon-DNA adducts in peripheral white blood-cells after charbroiled beef consumption. *Carcinogenesis*, **16**, 1079–1085.
 27. Strickland, P. T., Qian, Z., Friesen, M. D., Rothman, N. and Sinha, R. (2002) Metabolites of 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP) in human urine after consumption of charbroiled or fried beef. *Mutat. Res.*, **506**, 163–173.
 28. Helzlsouer, K. J., Alberg, A. J., Huang, H. Y. *et al.* (1999) Serum concentrations of organochlorine compounds and the subsequent development of breast cancer. *Cancer Epidemiol. Biomarkers Prev.*, **8**, 525–532.
 29. Tasevska, N., Runswick, S. A., McTaggart, A. and Bingham, S. A. (2005) Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 1287–1294.
 30. Nielsen, S. E., Freese, R., Kleemola, P. and Mutanen, M. (2002) Flavonoids in human urine as biomarkers for intake of fruits and vegetables. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 459–466.
 31. Amorim, L. C. A. and Cardeal, Z. D. L. (2007) Breath air analysis and its use as a biomarker in biological monitoring of occupational and environmental exposure to chemical agents. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, **853**, 1–9.
 32. Kaaks, R., Riboli, E. and Sinha, R. (1997) Biochemical markers of dietary intake. In Toniolo, P., Boffetta, P., Shuker, D. E. G., Rothman, N., Hulka, B. and Pearce, N. (eds), *Application of Biomarkers in Cancer Epidemiology*, IARC Scientific Publication No. 142. IARC, Lyon, France, pp. 103–126.
 33. Hecht, S. S., Carmella, S. G., Yoder, A. *et al.* (2006) Comparison of polymorphisms in genes involved in polycyclic aromatic hydrocarbon metabolism with urinary phenanthrene metabolite ratios in smokers. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 1805–1811.
 34. Poirier, M. C. (2004) Chemical-induced DNA damage and human cancer risk. *Nat. Rev. Cancer*, **4**, 630–637.
 35. Phillips, D. H. (2005) DNA adducts as markers of exposure and risk. *Mutat. Res.*, **577**, 284–292.
 36. IARC. (2002) Some traditional herbal medicines, some mycotoxins, naphthalene and styrene. In *Monographs on the Evaluation of Carcinogenic Risks to Humans* (Vol. 82). IARC Press, Lyon, France, pp. 301–366.
 37. Merrill, A. H., Sullards, M. C., Wang, E., Voss, K. A. and Riley, R. T. (2001) Sphingolipid metabolism: roles in signal transduction and disruption by fumonisins. *Environ. Health Perspect.*, **109**, 283–289.
 38. Albertini, R. J. (2001) Hprt mutations in humans: biomarkers for mechanistic studies. *Mutat. Res.*, **489**, 1–16.
 39. Lindholm, C., Murphy, B. P., Bigbee, W. L. *et al.* (2004) Glycophorin A somatic cell mutations in a population living in the proximity of the Semipalatinsk nuclear test site. *Radiat. Res.*, **162**, 164–170.
 40. Hussain, S. P. and Harris, C. C. (2000) Molecular epidemiology and carcinogenesis: endogenous and exogenous carcinogens. *Mutat. Res.*, **462**, 311–322.
 41. Tang, D., Warburton, D., Tannenbaum, S. R. *et al.* (1999) Molecular and genetic damage from environmental tobacco smoke in young children. *Cancer Epidemiol. Biomarkers Prev.*, **8**, 427–431.
 42. Vineis, P. and Husgafvel-Pursiainen, K. (2005) Air pollution and cancer: biomarker studies in human populations. *Carcinogenesis*, **26**, 1846–1855.
 43. Rothman, N., Correavillasenor, A., Ford, D. P. *et al.* (1993) Contribution of occupation and diet to white blood cell polycyclic aromatic hydrocarbon DNA adducts in wildland firefighters. *Cancer Epidemiol. Biomarkers Prev.*, **2**, 341–347.
 44. Vineis, P. and Garte, S. (2008) Biomarker validation. In Wild, C. P., Garte, S. and Vineis, P. (eds), *Molecular Epidemiology of Chronic Disease*. Wiley Press, Chichester, pp. 71–82.
 45. Qian, G. S., Ross, R. K., Yu, M. C. *et al.* (1994) A follow-up study of urinary markers of aflatoxin exposure and liver cancer risk in Shanghai, People's Republic of China. *Cancer Epidemiol. Biomarkers Prev.*, **3**, 3–10.
 46. Gilbert, J., Brereton, P. and MacDonald, S. (2001) Assessment of dietary exposure to ochratoxin A in the UK using a duplicate diet approach and analysis of urine and plasma samples. *Food Addit. Contam.*, **18**, 1088–1093.
 47. Anderson, L. M., Diwan, B. A., Fear, N. T. and Roman, E. (2000) Critical windows of exposure for children's health: cancer in human epidemiological studies and neoplasms in experimental animal models. *Environ. Health Perspect.*, **108**, 573–594.
 48. Wild, C. P. and Kleinjans, J. (2003) Children and increased susceptibility to environmental carcinogens: evidence or empathy? *Cancer Epidemiol. Biomarkers Prev.*, **12**, 1389–1394.
 49. Stellman, S. D., Djordjevic, M. V., Muscat, J. E. *et al.* (1998) Relative abundance of organochlorine pesticides and polychlorinated biphenyls in adipose tissue and serum of women in Long Island, New York. *Cancer Epidemiol. Biomarkers Prev.*, **7**, 489–496.
 50. Fry, R. C., Navasumrit, P., Valiathan, C. *et al.* (2007) Activation of inflammation/NF-kappa B signaling in infants born to arsenic-exposed mothers. *PLoS Genet.*, **3**, 2180–2189.
 51. Ozbal, C. C., Dasari, R. R. and Tannenbaum, S. R. (1998) Stability of histone adducts in murine models: implications for long-term molecular dosimetry. *Abstr. Pap. Am. Chem. Soc.*, **216**, U560–U561.
 52. Kirk, G. D., Lesi, O. A., Mendy, M. *et al.* (2005) TP53 mutation in plasma DNA, hepatitis B viral infection, and risk of hepatocellular carcinoma. *Oncogene*, **24**, 5858–5867.
 53. Greenblatt, M. S., Bennett, W. P., Hollstein, M. and Harris, C. C. (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.*, **54**, 4855–4878.
 54. Jackson, P. E., Kuang, S. Y., Wang, J. B. *et al.* (2003) Prospective detection of codon 249 mutations in plasma of hepatocellular carcinoma patients. *Carcinogenesis*, **24**, 1657–1663.
 55. Skipper, P. L., Tannenbaum, S. R., Ross, R. K. and Yu, M. C. (2003) Nonsmoking-related arylamine exposure and bladder cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 503–507.
 56. Gan, J. P., Skipper, P. L., Gago-Dominguez, M. *et al.* (2004) Alkyl-aniline-hemoglobin adducts and risk of non-smoking-related bladder cancer. *J. Natl. Cancer Inst.*, **96**, 1425–1431.
 57. Rundle, A. G., Vineis, P. and Ahsan, H. (2005) Design options for molecular epidemiology research within cohort studies. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 1899–1907.
 58. Wang, L. Y., Hatch, M., Chen, C. J. *et al.* (1996) Aflatoxin exposure and risk of hepatocellular carcinoma in Taiwan. *Int. J. Cancer*, **67**, 620–625.
 59. Tang, D. L., Phillips, D. H., Stampfer, M. *et al.* (2001) Association between carcinogen-DNA adducts in white blood cells and lung cancer risk in the physicians health study. *Cancer Res.*, **61**, 6708–6712.
 60. Peluso, M., Munnia, A., Hoek, G. *et al.* (2005) DNA adducts and lung cancer risk: a prospective study. *Cancer Res.*, **65**, 8042–8048.
 61. Walker, V. E., Wu, K. Y., Upton, P. B. *et al.* (2000) Biomarkers of exposure and effect as indicators of potential carcinogenic risk arising from in vivo metabolism of ethylene to ethylene oxide. *Carcinogenesis*, **21**, 1661–1669.
 62. Imbriani, M. and Ghittori, S. (2005) Gases and organic solvents in urine as biomarkers of occupational exposure: a review. *Int. Arch. Occup. Environ. Health*, **78**, 1–19.
 63. Watson, W. P. and Mutti, A. (2004) Role of biomarkers in monitoring exposures to chemicals: present position, future prospects. *Biomarkers*, **9**, 211–242.
 64. Wojnowski, L., Turner, P. C., Pedersen, B. *et al.* (2004) Increased levels of aflatoxin-albumin adducts are associated with CYP3A5 polymorphisms in The Gambia, West Africa. *Pharmacogenetics*, **14**, 691–700.
 65. Matullo, G., Peluso, M., Polidoro, S. *et al.* (2003) Combination of DNA repair gene single nucleotide polymorphisms and increased levels of DNA adducts in a population-based study. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 674–677.
 66. Lewin, M. H., Bailey, N., Bandaletova, T. *et al.* (2006) Red meat enhances the colonic formation of the DNA adduct O-6-carboxymethyl guanine: implications for colorectal cancer. *Cancer Res.*, **66**, 1859–1865.

67. Sharma, R. A. and Farmer, P. B. (2004) Biological relevance of adduct detection to the chemoprevention of cancer. *Clin. Cancer Res.*, **10**, 4901–4912.
68. Collins, A. R., Harrington, V., Drew, J. and Melvin, R. (2003) Nutritional modulation of DNA repair in a human intervention study. *Carcinogenesis*, **24**, 511–515.
69. Wang, J. S., Shen, X. N., He, X. *et al.* (1999) Protective alterations in phase 1 and 2 metabolism of aflatoxin B-1 by oltipraz in residents of Qidong, People's Republic of China. *J. Nat. Cancer Inst.*, **91**, 347–354.
70. Jacobson, L. P., Zhang, B. C., Zhu, Y. R. *et al.* (1997) Oltipraz chemoprevention trial in Qidong, People's Republic of China: study design and clinical outcomes. *Cancer Epidemiol. Biomarkers Prev.*, **6**, 257–265.
71. Kensler, T. W., He, X., Otieno, M. *et al.* (1998) Oltipraz chemoprevention trial in Qidong, People's Republic of China: modulation of serum aflatoxin albumin adduct biomarkers. *Cancer Epidemiol. Biomarkers Prev.*, **7**, 127–134.
72. Egner, P. A., Wang, J. B., Zhu, Y. R. *et al.* (2001) Chlorophyllin intervention reduces aflatoxin-DNA adducts in individuals at high risk for liver cancer. *Proc. Natl Acad. Sci. USA*, **98**, 14601–14606.
73. Turner, P. C., Sylla, A., Gong, Y. Y. *et al.* (2005) Reduction in exposure to carcinogenic aflatoxins by postharvest intervention measures in west Africa: a community-based intervention study. *Lancet*, **365**, 1950–1956.
74. Mooney, L. A., Santella, R. M., Covey, L. *et al.* (1995) Decline of DNA damage and other biomarkers in peripheral blood following smoking cessation. *Cancer Epidemiol. Biomarkers Prev.*, **4**, 627–634.
75. Albertini, R. J. (2001) Developing sustainable studies on environmental health. *Mutat. Res.*, **480**, 317–331.
76. Amundson, S. A., Do, K. T., Shahab, S., Bittner, M., Meltzer, P., Trent, J. and Fornace, A. J. (2000) Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiat. Res.*, **154**, 342–346.
77. Forrest, M. S., Lan, O., Hubbard, A. E. *et al.* (2005) Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. *Environ. Health Perspect.*, **113**, 801–807.
78. Wu, M. M., Chiou, H. Y., Ho, I. C., Chen, C. J. and Lee, T. C. (2003) Gene expression of inflammatory molecules in circulating lymphocytes from arsenic-exposed human subjects. *Environ. Health Perspect.*, **111**, 1429–1438.
79. Wang, Z. X., Neuburg, D., Li, C. *et al.* (2005) Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. *Environ. Health Perspect.*, **113**, 233–241.
80. van Leeuwen, D. M., van Herwijnen, M. H. M., Pedersen, M. *et al.* (2006) Genome-wide differential gene expression in children exposed to air pollution in the Czech Republic. *Mutat. Res.*, **600**, 12–22.
81. Lampe, J. W., Stepaniants, S. B., Mao, M. *et al.* (2004) Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol. Biomarkers Prev.*, **13**, 445–453.
82. van Leeuwen, D. M., van Agen, E., Gottschalk, R. W. H. *et al.* (2007) Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs. *Carcinogenesis*, **28**, 691–697.
83. van Leeuwen, D. M., Pedersen, M., Hendriksen, P. J. M. *et al.* (2008) Genomic analysis suggests higher susceptibility of children to air pollution. *Carcinogenesis*, **29**, 977–983.
84. Spira, A., Beane, J., Shah, V. *et al.* (2004) Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc. Natl Acad. Sci. USA*, **101**, 10143–10148.
85. Joo, W. A., Suli, D., Lee, D. Y., Lee, E. N. and Kim, C. W. (2004) Proteomic analysis of plasma proteins of workers exposed to benzene. *Mutat. Res.*, **558**, 35–44.
86. Wang, Y. L., Holmes, E., Nicholson, J. K. *et al.* (2004) Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification. *Proc. Natl Acad. Sci. USA*, **101**, 12676–12681.
87. Solanky, K. S., Bailey, N. J. C., Beckwith-Hall, B. M. *et al.* (2003) Application of biofluid H-1 nuclear magnetic resonance-based metabonomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Anal. Biochem.*, **323**, 197–204.
88. Holmes, E., Loo, R. L., Stamler, J. *et al.* (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, **453**, 396–400.
89. Davis, C. D. and Milner, J. (2004) Frontiers in nutrigenomics, proteomics, metabolomics and cancer prevention. *Mutat. Res.*, **551**, 51–64.
90. Elliott, R. and Ong, T. J. (2002) Science, medicine, and the future—nutritional genomics. *Br. Med. J.*, **324**, 1438–1442.
91. McTiernan, A. (2008) Mechanisms linking physical activity with cancer. *Nat. Rev. Cancer*, **8**, 205–211.
92. Rundle, A. (2005) Molecular epidemiology of physical activity and cancer. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 227–236.
93. Bingham, S. and Riboli, E. (2004) Diet and cancer—the European prospective investigation into cancer and nutrition. *Nat. Rev. Cancer*, **4**, 206–215.
94. Collins, F. S. (2004) The case for a US prospective cohort study of genes and environment. *Nature*, **429**, 475–477.
95. Cyranoski, D. and Williams, R. (2005) Health study sets sights on a million people. *Nature*, **434**, 812.
96. Wild, C. P. (2008) Biomarkers of exposure and effect. In Rebbeck, T., Ambrosone, C. and Shields, P. G. (eds), *Molecular Epidemiology Applications in Cancer and Other Human Diseases*. Informa Healthcare, New York, pp. 81–98.

Received on September 9, 2008; revised on October 2, 2008;
accepted on October 2, 2008